Genome-wide gene-based analyses of weight loss interventions identify a potential role for NKX6.3 in metabolism

Valsesia, Armand; Wang, Qiao-Ping; Gheldof, Nele; Carayol, Jérôme; Ruffieux, Hélène; Clark, Teleri; Shenton, Victoria; Oyston, Lisa J; Lefebvre, Gregory; Metairon, Sylviane; Chabert, Christian; Walter, Ondine; Mironova, Polina; Lau, Paulina; Descombes, Patrick; Viguerie, Nathalie; Langin, Dominique; Harper, Mary-Ellen; Astrup, Arne; Saris, Wim H; Dent, Robert; Neely, Greg G; Hager, Jörg

Published in:
Nature Communications

DOI:
10.1038/s41467-019-08492-8

Publication date:
2019

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

Document license:
CC BY

Citation for published version (APA):
Valsesia, A., Wang, Q-P., Gheldof, N., Carayol, J., Ruffieux, H., Clark, T., ... Hager, J. (2019). Genome-wide gene-based analyses of weight loss interventions identify a potential role for NKX6.3 in metabolism. Nature Communications, 10, [540]. https://doi.org/10.1038/s41467-019-08492-8

Download date: 06. Mar. 2020
Genome-wide gene-based analyses of weight loss interventions identify a potential role for NKX6.3 in metabolism

Armand Valsesia1, Qiao-Ping Wang2,3, Nele Gheldof1, Jérôme Carayol1, Hélène Ruffieux1,4, Teleri Clark2, Victoria Shenton2, Lisa J. Oyston2, Gregory Lefebvre1, Sylviane Metairon1, Christian Chabert1, Ondine Walter1, Polina Mironova1, Paulina Lau5, Patrick Descombes1, Nathalie Viguerie6, Dominique Langin6,7, Mary-Ellen Harper8, Arne Astrup9, Wim H. Saris10, Robert Dent4, Greg G. Neely2 & Jörg Hager1

Hundreds of genetic variants have been associated with Body Mass Index (BMI) through genome-wide association studies (GWAS) using observational cohorts. However, the genetic contribution to efficient weight loss in response to dietary intervention remains unknown. We perform a GWAS in two large low-caloric diet intervention cohorts of obese participants. Two loci close to NKX6.3/MIR486 and RBSG4 are identified in the Canadian discovery cohort (n = 1166) and replicated in the DiOGenes cohort (n = 789). Modulation of HGTX (NKX6.3 ortholog) levels in Drosophila melanogaster leads to significantly altered triglyceride levels. Additional tissue-specific experiments demonstrate an action through the oenocytes, fly hepatocyte-like cells that regulate lipid metabolism. Our results identify genetic variants associated with the efficacy of weight loss in obese subjects and identify a role for NKX6.3 in lipid metabolism, and thereby possibly weight control.
Obesity is a world-wide issue and a major risk factor for cardiovascular disease, dyslipidemia, hypertension, insulin resistance and type 2 diabetes as well as cancer^{1-3}. A recent report from the NCD-RisC network has shown the increasing prevalence of obesity and estimated that with current post-2000 trends, the global obesity frequency would surpass 18% in men and 21% in women by 2025^4.

Multiple studies have shown that weight loss through energy restricted dietary interventions improves metabolic dysfunction^5-8. Nevertheless, a large inter-individual variability is observed regarding the capacity to lose weight and to maintain the lost weight^9-11. Genome-wide association studies (GWAS) from the GIANT consortium have identified about 100 loci associated with body mass index (BMI) variability in the general population^9. Those candidate obesity loci were investigated in two lifestyle interventions: the Diabetes Prevention Program (DPP)^12 and Look AHEAD^12,13. In these candidate analyses, only one marker (MTHFR rs1885988) was associated with degree of weight loss and none to weight regain tendency^14. To date, no genome-wide approach for weight loss success has been undertaken^15.

Here we present results from a genome-wide association (GWA) study for weight loss using two large low-caloric diet interventions: the Canadian Optifast900® meal replacement program^16 and the DIOGenes clinical trial^17,18. Our analyses implement a gene-based GWA to maximize statistical power^19,20. One cohort is used for discovery and the second, for replication. Next, we perform Bayesian risk variant inference based on joint modeling of the GWA signals and large-scale epigenome annotation data to restrict the association signals to the most likely associated SNPs. Finally, we perform functional RNAi knockdown in Drosophila melanogaster to study the potential in vivo metabolic impact of the regional candidate genes.

Our study provides evidence for a weight loss locus on chromosome 8p11 and knock out experiments in Drosophila melanogaster suggest the NKX6.3 gene in the region as a potential functional candidate.

### Results

#### Cohort descriptions

The Optifast900 cohort included both obese and severely obese subjects (mean BMI = 43.2 kg/m^2 ± 0.3 standard error of the mean) and the DIOGenes cohort included overweight and obese participants (mean BMI = 34.5 kg/m^2 ± 0.2). Clinical characteristics of the participants are available in Table 1. Upon a 5-week low calorie diet (LCD), participants lost on average 9.3% (11.3 kg) and 7.5% (7.5 kg) of initial body weight, respectively for the Optifast900 and DIOGenes participants. At baseline, Optifast900 participants were considered more insulin-resistant than DIOGenes subjects (HOMA-IR = 4.16 ± 0.14 vs. 3.15 ± 0.10), as expected given the more severe obesity.

#### Gene-based association studies

We searched for SNPs associated with degree of weight loss using the largest cohort (Optifast900, n = 1166) as a discovery dataset. The smaller cohort (DIOGenes, n = 789) was then used for replication. To unravel associations using single-SNP approaches at genome-wide scale, very large cohorts are needed, but such sample size cannot be obtained in randomized clinical trials. To better extract association signals, we used a gene-based approach that enables to integrate individual SNP association signals into a locus-wise signal (see Methods). Upon gene-based analyses of the Optifast900 cohort, we identified 12 genes, corresponding to 6 distinct loci with nearby SNPs significantly associated with weight loss (genome-wide FDR < 0.05 see Table 2). A Manhattan plot of the gene-based GWA results is available in Fig. 1.

Next, replication of those loci was attempted using the DIOGenes cohort. Two out of the six loci were successfully replicated in the DIOGenes cohort (two-stage FDR < 0.05, Table 2): the RBSG4 locus on chromosome 4q24 and the MIR486/NKX6.3 locus on chromosome 8p11. Meta-analysis using random-effect modeling of the two cohorts also showed significant association for these loci (both at genome-wide levels and with a two-stage approach) with effect sizes that were consistent between the two cohorts. Regional plots for those two loci are shown in Fig. 2. The MIR486 gene has two isoforms with similar coordinates, thus essentially the same SNPs were included in the gene-based analyses leading to very similar p-values for MIR486-1 and MIR486-2 (as seen in Table 2).

#### Bayesian framework for risk variant inference

We took advantage of the development of a recent Bayesian framework, RIVIERA-beta^21 to infer posterior probabilities of disease association (PPA) that were then used to rank the associated SNPs. Upon Bayesian modeling of SNPs with nominal p-values less than 1e-3 for the two replicated loci (RBSG4 and MIR486/ NKX6.3), we were able to restrict the list of candidate SNPs to four markers, with a possible regulatory impact. Figure 3 summarizes the results of the Bayesian modeling and presents the overlap between variants and epigenome annotations. Table 3 provides the effect size (from single-SNP GWAS) for each identified marker. For the RBSG4 locus (Fig. 3a), we identified three markers (rs873822, rs870879, rs1027493) significantly enriched in epigenome annotations. Those markers were in strong LD with each other (r^2 > 90%) and thus any of those would tag the others. Those three markers are common variants (with Minor Allele Frequency (MAF) = 27%). For the MIR486/NKX6.3 locus, the rs6981587 SNP (MAF = 34%) emerged as the most likely risk marker.

### Table 1 Descriptive statistics for the two studies used in the analysis

|                        | OPTIFAST900 (n = 1166) | DIOGENES (n = 789) | p-value |
|------------------------|------------------------|--------------------|---------|
| Number of males (%)    | 237 (26.58%)           | 310 (33.95%)       | -       |
| Age at baseline (years)| 46.50 ± 0.32           | 41.37 ± 0.23       | p < 0.001 |
| BMI at baseline (kg/m^2)| 43.17 ± 0.23           | 43.53 ± 0.19       | p < 0.001 |
| Weight at baseline (kg) | 121.66 ± 0.76          | 99.74 ± 0.67       | p < 0.001 |
| Weight after 5-week LCD (kg) | 110.31 ± 0.68   | 92.20 ± 0.61       | p < 0.001 |
| Change in weight during LCD (kg) | −11.35 ± 0.11  | −7.55 ± 0.11       | p < 0.001 |
| Change in weight during LCD (%) | −9.28 ± 0.06   | −7.51 ± 0.09       | p < 0.001 |
| Fasting glucose levels (mmol/L) at baseline | 5.71 ± 0.05    | 5.12 ± 0.03        | p < 0.001 |
| HOMA-IR at baseline      | 4.16 ± 0.14            | 3.15 ± 0.10        | p < 0.001 |

The p-value was obtained from a two-sided t-test and assesses differences between the two cohorts. LCD: low caloric diet, HOMA-IR: homeostasis model assessment of insulin resistance.
focused our analysis on Ank/ANK1 and RBSG4. Because marks (Fig. 3b). For the four SNPs in the two genomic loci, yet none showed such enrichment in epigenome annotation consistent effect sizes were observed between the two cohorts, as functional assessment in follow-up.

Indeed, within the top regulatory variant and that would also deserve functional follow-up.

Table 2 Top hits from the gene-based GWA

| Gene   | Chr | Start | Stop   | OPTIFAST900 (n = 1166) | DIOGENES (n = 789) | Meta-analysis |
|--------|-----|-------|--------|------------------------|---------------------|--------------|
| MIR486-2 | 8   | 41497961 | 41538025 | 1.7e-05 (0.043)         | 0.000618 (0.004)    | 1e-06 (1.2e-05) |
| MIR486  | 8   | 41497958 | 41538026 | 2.2e-05 (0.043)         | 0.000661 (0.004)    | 2e-06 (1.2e-05) |
| NKX6-3  | 8   | 41483828 | 41524878 | 2.4e-05 (0.043)         | 0.00185 (0.0074)    | 2.6e-05 (0.0001) |
| RBSG4   | 1   | 167124598 | 167185042 | 2e-05 (0.043)           | 0.011 (0.033)       | 3.8e-05 (0.0001) |
| DBND1   | 16  | 90051272 | 90106539 | 7e-06 (0.035)           | 0.0305 (0.073)      | 0.0116 (0.02) |
| PTPRT   | 20  | 40681391 | 41838557 | 6e-06 (0.035)           | 0.1728 (0.28)       | 0.0062 (0.012) |
| GAS8    | 16  | 90066036 | 90131379 | 7e-06 (0.035)           | 0.1958 (0.28)       | 0.06909 (0.1) |
| CBE1    | 18  | 57078170 | 57384644 | 8e-06 (0.035)           | 0.2008 (0.28)       | 0.000346 (0.00083) |
| C16orf3 | 16  | 90075315 | 90163209 | 2.1e-05 (0.043)         | 0.2116 (0.28)       | 0.1049 (0.14) |
| URAHP   | 16  | 90086168 | 90134191 | 1.1e-05 (0.04)          | 0.2747 (0.33)       | 0.1339 (0.16) |
| ORSK2   | 3   | 98196524 | 98237475 | 7e-06 (0.035)           | 0.5784 (0.63)       | 0.1578 (0.17) |
| ORSK1   | 3   | 98168420 | 98209372 | 1.3e-05 (0.04)          | 0.8541 (0.85)       | 0.3117 (0.31) |

This table presents gene-based association p-values for the two cohorts and their meta-analysis. False-discovery rate (FDR) adjusted p-value is indicated within parenthesis. For the discovery cohort (Optifast900), the FDR is a genome-wide FDR. For the replication cohort (DIOGENES) and the meta-analysis results, the FDR is adjusted for a two-stage analysis. FDRs less than 5% are shown in bold. Source data are provided as a Source Data file.

Functional assessment in Drosophila melanogaster. To investigate a potential in vivo metabolic function for the genes around the risk variants, we used the fruit fly Drosophila melanogaster. Because RBSG4 and MIR486 are not conserved in the fly, we focused our analysis on Ank/ANK1, HGTX/NKX6-3 and CG3209/AGPAT6 and each gene was targeted using whole body RNAi knockdown (Actin-Gal4). There were no major developmental effects for Actin-Gal4 > UAS-Ank and Actin-Gal4 > UAS-CG3209 RNAi flies and we did not observe significant changes in TAG levels compared to their wild-type controls (Supplementary Fig. 2a). We also performed overexpression (OE) of ANK1 using a whole-body driver (Actin-Gal4). There was no impact on triglyceride levels in the Actin-Gal4 > UAS-ANK1 OE animals compared to controls (Supplementary Fig. 2b). The majority of Actin-Gal4 > UAS-HGTX RNAi flies were developmentally lethal (>95% pupal lethality), however some animals did survive. From the viable HGTX knockdown flies (Actin-Gal4 > UAS-HGTX RNAi), we observed a significant reduction in triglyceride (TAG) level compared to controls (Fig. 4a). This finding was confirmed using a second RNAi hairpin (Supplementary Fig. 3a). Complete lethality of F1 progenies was observed in HGTX RNAi or HGTX overexpression flies using additional ubiquitous drivers Tub-Gal4 and DA-Gal4. To bypass developmental effects of HGTX knockdown, we next performed an adult-specific inducible knockdown using the TubGal80 system. Actin-Gal4; Gal80b >
UAS-HGTX RNAi animals were raised at 18 °C during developmental stage, which suppresses RNAi and then hatched flies were shifted to 29 °C for 6 days at which time RNAi is activated. Induced knock-down animals displayed a similar level of TAG reduction as constitutive HGTX knockdown animals compared to the parental controls (Fig. 4b). To confirm inducible RNAi knockdown efficiency, we used qPCR and observed approximate 60% reduction in HGTX mRNA levels (Fig. 4c). Further metabolic characterization of these inducible knockdown animals showed no significant difference in levels of glycogen (Supplementary Fig. 3b) or trehalose (Supplementary Fig. 3c), and body weight (Supplementary Fig. 3d), food intake (Supplementary Fig. 3e), and starvation response (Supplementary Fig. 3f) were also similar to the controls. Of note, HGTX inducible knockdown did not affect fly insulin-like peptide Ilp2 or Ilp5 expression but resulted in a decrease in Ilp3 expression. However, we did not observe any difference in dilp3 expression at the protein level (Supplementary Fig. 4).

To further confirm the role of HGTX in regulation of TAG, we used inducible over-expression of HGTX in adults with mRNA expression ~9 times higher in Actin-Gal4; Gal80ts > UAS-HGTX OE animals (Fig. 4d) compared to the parental controls. HGTX over-expression led to a mild reduction in TAG (Fig. 4e). No significant impact was observed for Ilp2, 3 and 5 mRNA expression or dilp3 protein levels (Supplementary Fig. 5).

To find the specific tissue in which HGTX acts, we carried out tissue specific HGTX RNAi targeting expression in the fat body (Ppl-Gal4), muscle (Mef2-Gal4), brain (nSyb-Gal4) or oenocytes (Oeno-Gal4). We found that only oenocyte-specific knock down of HGTX resulted in a significant reduction in TAG compared to parental controls (Fig. 4f). Conversely, oenocyte-specific over-expression of HGTX resulted in a significant increase in TAG (Fig. 4g). Together, our data supports a role for HGTX/NKX6.3 acting in the fly oenocyte to regulate triglyceride content in vivo.

**Discussion**

Here we describe a genome-wide association for weight loss. We used two independent weight loss cohorts: a Canadian Optifast900 cohort (n = 1166) for discovery and the pan-European DiOGenes cohort (n = 789) for replication.

Recent analysis14 of the Look AHEAD and DPP cohorts only focused on 91 established obesity loci8,22 and found association of only one of the loci with weight loss and regain after three years,
namely the MTIF3-rs1885988 SNP. Analysis of this SNP in our two cohorts did not confirm this association (meta-analysis p-value = 0.52). This could be due to differences between the behavior-based and low-caloric dietary interventions as well as the length/type of the intervention in the different studies. Analyses of the remaining 90 BMI loci in our data confirmed the lack of association of these loci with weight loss (all loci had FDR > 10%); further highlighting the need to investigate the genetic contribution to weight loss in well-controlled weight loss studies (i.e. with similar intervention design).

GWAs of complex, polygenic traits (such as metabolic diseases) require very large sample sizes to unravel common variants with low/moderate effects. Such large sample sizes are typically achieved with meta-analyses of numerous observational studies. For illustration, the latest GWAs from the GIANT consortium used a sample size greater than 700,000 and identified 716 variants representing about 5% of BMI variation. Such sample sizes are not achievable for clinical intervention studies. However, over the past few years, the development of new statistical methods has enabled to increase statistical power using multi-marker analyses. We thus used a well-established gene-based strategy and took advantage of recent improvements for single-SNP analyses by using a Bayesian linear-mixed effect model. Gene-based approaches enable to combine association p-values from individual SNPs into a single locus (e.g. gene-level) p-value. This leads to several benefits. First, statistical evidence is strengthened by integrating association p-values from markers located within a same region. Some of those markers may already display association signals close to genome-wide significance thresholds, yielding a more extreme combined p-value. These approaches account for LD relationships between markers by using resampling approaches (e.g. Monte Carlo simulations). The second benefit from gene-based approaches pertains to a significant reduction of the multiple testing burden (as the number of tests is about 20,000 regions instead of millions of SNPs). Finally, it has been previously discussed that gene-based approaches are less prone to spurious associations caused by population stratification compared to single-SNP or haplotype-based analyses.

By using such a combination of tools, we identified six different loci with genome-wide FDR < 5% in the discovery phase. Two of those loci (RBGS4 and MIR486/NKX6.3) replicated in a two-stage analysis (two-stage FDR < 5%) using the DiOGenes study and yielded significant and consistent results using a meta-analysis approach. Bayesian modeling of epigenomic annotation was able to highlight markers in these loci with a possible regulatory impact. Minor alleles from markers nearby RBGS4 were found associated with increased weight loss. Interestingly, the major allele (C) for rs6981587, near MIR486/NKX6.3, was associated with decreased weight loss. Our two cohorts were of European ancestry and the C allele frequency (77%) was consistent with other European populations as well as Asian populations (>70%). However, these allele frequencies were much lower in several African populations (with frequency ranging from 35–46%). This observation would deserve additional follow-up in weight loss studies with participants from different ancestries; as it may have implication for weight loss intervention in these populations, including in admixed populations (e.g. African Americans).

Our analyses highlighted several genes in the vicinity of these association signals. Within the first locus, RBGS4, also known as LINC01363, encodes a long non-coding RNA (lncRNA) and has not been previously reported to associate with obesity and weight loss interventions. Our current understanding of lncRNA remains limited though this class of RNAs can play an important role in gene regulation. The contribution of lncRNAs as regulators of the endocrine system is widely accepted. Many GWAs have reported association signals in the vicinity of lncRNA loci and gene expression studies reported that those GWA signals had a cis-eQTL effect specific to lncRNA and not to other neighboring
protein-coding genes\textsuperscript{29}. In addition, the contribution of lncRNA genes to local gene regulation involves complex processes that are not necessarily linked to the lncRNA transcripts themselves but instead includes processes associated with their production (such as enhancer activities, transcription processes and splicing)\textsuperscript{30}. It is also worth noting that lncRNA have been reported as regulators of adipogenesis\textsuperscript{31}. Thus, the contribution of lncRNA, such as \textit{RBGS4}, in response to clinical intervention for metabolic diseases may be under-appreciated.

The second locus associating with weight loss encompasses several genes: \textit{MIR486}, \textit{ANK1}, \textit{AGPAT6} and \textit{NKKX6.3}.

The \textit{MIR486} locus includes two mir genes encoding pre-mir isoforms of the same microRNA (miRNA). MiRNAs emerged as regulators of important biological processes and have been involved in many complex diseases including obesity, insulin resistance, T2D and cancer\textsuperscript{32-38}. \textit{MIR486} has been shown to regulate \textit{SIRT1} and its expression in human adipose tissue-derived mesenchymal stem cells is controlled by high glucose\textsuperscript{39}. Additional studies have shown that \textit{mir486} could impact NFKB signaling by inhibition of \textit{NKFB} inhibitors\textsuperscript{40} and that \textit{mir486} could inhibit \textit{FOXO1}, a key mediator of insulin signaling\textsuperscript{41} and triglyceride metabolism\textsuperscript{42}.

\textit{ANK1} belongs to the ankyrin family that links membrane proteins to the cytoskeleton and play a large role in cell motility, proliferation and activation. Genetic variants located within \textit{ANK1} have been found associated with glycemic traits, impaired insulin release\textsuperscript{43} and T2D onset both in European and Asian populations\textsuperscript{44-45}. \textit{ANK1} variants were also shown to have cis-eQTL effect on the expression of small \textit{ANK1} in skeletal muscle\textsuperscript{46,47}. In our gene-based analyses, the \textit{ANK1} eQTL effect on the expression of small \textit{ANK1} was replicated with independent RNAi hairpins and confirmed using adult-inducible knockdown. \textit{NKKX6.3}, \textit{ANK1} and \textit{AGPAT6} genes, collectively, these cells regulate whole-body adipose tissue metabolism\textsuperscript{59}.

The third gene located in the vicinity of the identified variants is \textit{AGPAT6} (also known as \textit{GPAT4}, Glycerol-3-Phosphate Acyltransferase 4). The enzyme encodes the conversion of lysophosphatidic acid to phosphatidic acid and thus is an important contributor in TAG biosynthesis. \textit{AGPAT6}-deficient mice were shown to have alteration in lipid metabolism in tissue such as adipose and liver\textsuperscript{48}. \textit{AGPAT6}-deficient mice were also shown to exhibit a 25% reduction in body weight and resistance to both diet-induced and genetically induced obesity\textsuperscript{49}.

The fourth gene in this locus is \textit{NKKX6.3}, member of the NKK family that contribute to numerous developmental processes. In particular, the NKK homeobox 3 gene is involved in the development of the central nervous system, gastro-intestinal tract and pancreas\textsuperscript{50}. Interestingly, \textit{NKKX6.3} is located close to variants which have been associated with T2D\textsuperscript{43}. The recent GWAS by Mahajan and colleagues\textsuperscript{51} highlighted the rs13262861 SNP (8 kb away from our identified variants) as being associated with T2D. This SNP was not tested in our data as it did not achieve a good imputation metrics ($r^2 < 80\%$). This SNP is known as an \textit{NKKX6-3} cis-eQTL in human islets\textsuperscript{52}. Additionally, \textit{NKKX6-3} transcripts encode transcription factors required for the development of alpha and beta cells in the pancreas\textsuperscript{53} and have shown to influence insulin secretion\textsuperscript{54,55}. Glycemic improvements following LCD were only measured in the DioGenes cohort. Association between our three top \textit{NKKX6.3} variants and rs13262861 did not reveal any significant association with insulin sensitivity\textsuperscript{54,55}. Our prioritization analyses based on epigenomic annotation highlighted rs6981587 as the top regulatory variant. Since these annotation marks are derived from cell lines, deciphering the exact underlying mechanism may be premature and would require access to specific tissues for a subset of our GWA participants (e.g. with liver and fat biopsies). Yet, investigation of the BIOS QTL data\textsuperscript{56} found that rs6981587 was an eQTL, in whole blood, of the \textit{NKKX6.3}, \textit{ANK1} and \textit{AGPAT6} genes. Interestingly, only the \textit{NKKX6.3} eQTL reached genome-wide significance (FDR 5%), with the rs6981587-T allele associating with decreased expression levels.

To investigate a role for the genes \textit{NKKX6.3}, \textit{ANK1} and \textit{AGPAT6} in metabolic regulation in vivo, we used a knockdown strategy in \textit{Drosophila melanogaster}, as molecular mechanisms controlling fat mass are largely conserved between flies and humans\textsuperscript{57}. We used triglyceride content as the main metabolic readout. Triglycerides constitute the major components of lipids\textsuperscript{58}, and total triglyceride levels are used as a direct measure of fly adiposity\textsuperscript{59}. Whole-body knockdown of \textit{NKKX6.3}/\textit{HGTX} led to significant reduction of whole-body triglyceride content. This observation was replicated with independent RNAi hairpins and confirmed using adult-inducible knockdown. \textit{NKKX6.3}/\textit{HGTX} overexpression also led to a reduction of whole-body triglyceride (TAG) content. While this was surprising, this observation is not uncommon in functional screens\textsuperscript{60,61} and it suggests that a tight \textit{NKKX6.3}/\textit{HGTX} gene dosage is important to maintain TAG levels.

Tissue-specific inducible knockdown showed that \textit{NKKX6.3}/\textit{HGTX} acts in oenocytes to maintain TAG levels. RNAi knockdown in oenocytes led to decreased TAG levels while overexpression of \textit{NKKX6.3}/\textit{HGTX} led to increased TAG levels. Oenocytes are hepatocyte-like cells and are important to regulate the fly lipid metabolism\textsuperscript{59}. Specifically, these cells regulate whole-body TAG level and have a bidirectional metabolic role. Under starvation conditions, oenocytes accumulate lipid droplets; when food is abundant, they regulate growth, development and feeding behavior. This two-way coupling between body fat and oenocytes is analogous to the liver—adipose axis in mammals. While the exact mechanism of how \textit{NKKX6.3}/\textit{HGTX} can control lipid metabolism remains unclear, the effect could be oenocyte-specific, or via inter-organ communications between the fat body, brain and oenocytes.

Our fly results show phenotypic evidence that \textit{NKKX6.3} may be involved in lipid metabolism and may thereby contribute to weight loss variation. On the other hand, the published mouse data on \textit{AGPAT6} makes this gene also a plausible candidate. Another possibility is that the associated genetic variant influences all of these genes and that they jointly contribute to the

| SNP          | Chr | Position | Effect allele | Effect allele frequency | Optifast900 | Meta-analysis |
|--------------|-----|----------|---------------|-------------------------|-------------|---------------|
| rs873822     | 1   | 167126910| C             | 66%                     | 0.13 ± 0.04 (p = 0.00053) | 0.14 ± 0.03 (p = 3.43e-6) |
| rs870879     | 1   | 167126987| G             | 66%                     | 0.13 ± 0.04 (p = 0.00068) | 0.13 ± 0.03 (p = 4.27e-6) |
| rs1027493    | 1   | 167132882| C             | 67%                     | 0.13 ± 0.04 (p = 0.00056) | 0.13 ± 0.03 (p = 5.06e-6) |
| rs6981587    | 8   | 41516915 | C             | 77%                     | 0.19 ± 0.05 (p = 0.000043) | −0.17 ± 0.03 (p = 1.54e-6) |

This table presents single-SNP association $p$-values for the two cohorts and their meta-analysis for the top risk variant SNPs. Beta coefficients, with their standard error and $p$-value are provided, as estimated by the linear mixed effect model. Positive betas indicate that the effect allele associates with greater weight loss. Source data are provided as a Source Data file.

\textbf{ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-08492-8}
variability of weight loss in humans. The biological importance of the genes near the weight-loss associated variants and the proximity or overlap with annotated regulatory marks, provides evidence for a functional role in metabolism within this locus. However, given the complex relationship between different organs (brain, adipose, pancreas, liver and muscle) as well as the interplay between metabolic pathways, further investigation would require access to different tissues and under different conditions (e.g. weight loss, induced obesity) to elucidate the potential contribution of each gene.

In conclusion, we performed a weight loss GWA using data from a large clinical practice (the Canadian Optifast900 meal replacement program). Two loci (RBSG4 and MIR486/NKX6.3) were successfully replicated with data from a controlled trial (DiOGenes). Several independent studies provided evidence for a biological link between the NKX6-3/MIR486 locus and metabolic disorders including T2D and obesity. This lends additional confidence in our results. Our work opens opportunities for additional functional and preclinical studies to fully elucidate the link between the identified markers and the underlying metabolic and molecular mechanisms.

Methods

Ethics. Local Human Research Ethics Committees, from the Ottawa Hospital and DiOGenes studies, approved the study and all procedures were conducted in accordance with the Declaration of Helsinki. All participants gave informed written consent prior to any testing.

Study samples. The Canadian samples consisted of patients enrolled in the Weight Management Clinic who had completed 6 to 12 weeks meal-replacement regimen consisting of a product uniquely available in Canada, Optifast900 (Nestlé...
Genetic structure (Supplementary Fig. 1). Upon all genetic QCs, 1166 Ottawa and from both cohorts were all of European ancestry and the two cohorts had similar Principal component analyses (PCA) were performed independently on each GenomeStudio Software (Illumina). Quality control excluded SNPs with call rate

Clinical data. For the Canadian Optifast900 study, weight was measured weekly during the LCD intervention. For the DiOGenes study, weight was measured at baseline and after 1, 3, 5, 7, 8 weeks of LCD. The time point 5 week provided the largest sample size (i.e. smallest percentage of missing values) for both cohorts. Analyses investigated BMI after 5-weeks of LCD, adjusted for baseline BMI, age and gender.

Genotype data. Genotype data were generated using HumanCoreExome-12 v1.1 with 200 k tag SNPs from Genome-wide focused markers. The samples were processed with the Illumina TM platform following Illumina® HD Assay. Manual, according to manufacturer’s instructions. Genotypes were called with the GenomeStudio Software (Illumina). Quality control excluded SNPs with call rate <95%, violating Hardy-Weinberg equilibrium (FDR < 20%), low minor allele frequency <1%. Subjects were excluded if they had low call rate (<95%), abnormal high autosomal heterozygosity (FDR <1%), an XXY karyotype, or gender inconsistencies between genotype data and clinical records. For subjects with high identity-by-state (IBS > 95%), only the one having the highest call rate was kept. Principal component analyses (PCA) were performed independently on each cohort to discard subjects that were outliers in terms of genetic structure. Subjects from both cohorts were all of European ancestry and the two cohorts had similar genetic structure (Supplementary Fig. 1). Upon all genetic QCs, 1166 Ottawa and 789 DiOGenes subjects were kept for subsequent analyses.

Genotype imputation was then performed using SHAPEIT2 and IMPUTE2 based on the European reference panel from the 1000 Genome project as reference population for LD patterns, set-based test filtering, potentially the same SNPs could be assign to different genes, leading to lack of statistical power.

Statistical analyses. Single-SNP analyses were performed using BOLT-LiMM, a Bayesian linear mixed effect model to adjust for population structure and cryptic relatedness between individuals. The rate of weight loss was adjusted for sex, age and starting BMI. Results from the two cohorts, were subsequently meta-analyzed using Genome-Wide Association Meta Analysis (GWAMA) software with random-effect modeling and a double genomic-control (GC) correction. Genome-wide significant threshold was set to FDR < 5%.

Prioritization of GWA signals was performed using a Bayesian framework to model the joint likelihood of association p-values with large-scale epigenomic annotations. Such risk variance inference was performed using the RIVIERA-beta framework and with 450 epigenomic annotations (including histone marks, DNase I hypersensitivity, transcription factor binding, and localization within exons). The goal of this framework is to infer for each input SNP the posterior probability of disease given p-value. This is not an issue as the goal of such analyses is only to highlight loci associated with the trait of interest. Adjustment for multiple-testing was performed using the Benjamini-Hochberg correction. Genome-wide significance threshold was set to FDR < 5%.

Analysis of fly phenotypes was performed using one-way ANOVA with Bonferroni adjustment for multiple comparisons.
Reporting Summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

All summary statistics are freely available (Supplementary Data 1–6). All data that support the findings of this study are available from the corresponding author (Armand.Valsesia@rd.nestle.com) upon reasonable request.

Received: 15 April 2017 Accepted: 7 January 2019
Published online: 01 February 2019

References

1. Haslam, D. W. & James, W. P. T. Obesity. Lancet 366, 1197–1209 (2005).
2. Dixon, J. B. The effect of obesity on health outcomes. Mol. Cell. Endocrinol. 316, 104–108 (2010).
3. Lean, M. E. Pathophysiology of obesity. Proc. Nutr. Soc. 59, 331–336 (2000).
4. Risk, N. C. D. Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet Lond. Engl. 387, 1377–1396 (2016).
5. Soare, A., Weiss, E. P. & Pozzilli, P. Benefits of caloric restriction for cardiometabolic health, including type 2 diabetes mellitus risk. Diabetes Metab. Res. Rev. 36, 41–47 (2014).
6. Alves, N. E. G. et al. Meal replacement based on human ration modulates metabolic risk factors during body weight loss: a randomized controlled trial. Eur. J. Nutr. 53, 939–950 (2014).
7. Neiberg, R. H. et al. Patterns of weight change associated with long-term weight change and cardiovascular disease risk factors in the Look AHEAD Study. Obes. Silver Spring Md 20, 2048–2056 (2012).
8. Valsesia, A., Saris, W. H., Astrup, A., Hager, J. & Masoodi, M. Distinct lipid profiles predict improved glycemic control in obese, nondiabetic patients after a low-caloric diet intervention: the diet, obesity and genes randomized trial. Am. J. Clin. Nutr. 104, 566–575 (2011).
9. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197–206 (2015).
10. Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N. Engl. J. Med. 346, 393–403 (2002).
11. Delahanty, L. M. et al. Genetic predictors of weight loss and weight regain after intensive lifestyle modification, metformin treatment, or standard care in the diabetes prevention program. Diabetes Care 35, 363 (2012).
12. McCaffery, J. et al. FTO predicts weight regain in the Look AHEAD Clinical study. Proc. Nutr. Soc. 75, 335–362 (2004).
13. Knoll, N. et al. The fatty acid amide hydrolase (FAAH) gene variant rs324420 AA/AC is not associated with weight loss in a 1-year lifestyle intervention for obese children and adolescents. Horm. Metab. Res. 44, 75–77 (2012).
14. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10, 155–159 (2009).
15. Kornfeld, J.-W. & Brüning, J. C. Regulation of metabolism by long, non-coding RNAs. Front. Genet. 5, 57 (2014).
16. Neale, B. M. & Sham, P. C. The future of association studies: gene-based analysis and replication. Am. J. Hum. Genet. 75, 353–362 (2004).
17. Attie, M. et al. The role of genetic variation in insulin resistance, type 2 diabetes and obesity. Annu. Rev. Med. 64, 387–403 (2013).
18. Chen, X. et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis and other diseases. Cell Res. 18, 997–1006 (2008).
19. Prats-Pigué, A. et al. Changes in circulating microRNAs are associated with childhood obesity. J. Clin. Endocrinol. Metab. 98, E1655–E1660 (2013).
20. Scott, L. J. et al. The genetic regulatory signature of type 2 diabetes in human adipose tissue and obesity. Nat. Rev. Endocrinol. 11, 276–288 (2015).
21. McGregor, R. & Choi, M. microRNAs in the regulation of adipogenesis and obesity. Curr. Mol. Med. 11, 304–316 (2011).
22. Flowers, E. et al. Circulating MicroRNA-320a and MicroRNA-486 predict thiazolidinedione response: moving towards precision health for diabetes prevention. Metabolism 64, 1051–1059 (2015).
23. Wang, X. et al. Determination of 14 Circulating microRNAs in Swedes and Iraquis with and without diabetes mellitus type 2. PLoS ONE 9, e86792 (2014).
24. Chen, X. et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 18, 997–1006 (2008).
25. Prats-Pigué, A. et al. Changes in circulating microRNAs are associated with childhood obesity. J. Clin. Endocrinol. Metab. 98, E1655–E1660 (2013).
26. Arner, P. & Kylute, A. MicroRNA regulatory networks in human adipose tissue and obesity. Nat. Rev. Endocrinol. 11, 276–288 (2015).
27. Kim, Y. J. et al. miR-486-5p induces replicative senescence of human adipose tissue-derived mesenchymal stem cells and its expression is controlled by high glucose. Stem. Cells Dev. 21, 1749–1760 (2011).
28. Song, L. et al. miR-486 sustains NF-κB activity by disrupting multiple NF-κB-negative feedback loops. Cell Res. 23, 274–289 (2013).
29. Xu, J. et al. Transcription factor FoxO1, the dominant mediator of muscle wasting in chronic kidney disease, is inhibited by microRNA-486. Kidney Int. 82, 401–411 (2012).
30. Altomonte, J. et al. FoxO1 mediates insulin action on apocIII and triglyceride metabolism. J. Clin. Invest. 114, 1493–1503 (2004).
31. Lim, B. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 44, 981–990 (2012).
32. Imamura, M. et al. A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. Hum. Mol. Genet. 21, 3042–3049 (2012).
33. Chen, P. et al. Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. Diabetes 63, 2551–2562 (2014).
34. Yan, R. et al. A novel type 2 diabetes risk allele increases the promoter activity of the muscle-specific small ankyrin 1 gene. Sci. Rep. 6, 25105 (2016).
35. Scott, L. J. et al. The genetic regulatory signature of type 2 diabetes in human skeletal muscle. Nat. Commun. 7, 11764 (2016).
36. Nagle, C. A. et al. Identification of a novel sn-glycerol-3-phosphate acyltransferase isoform, GPAT4, as the enzyme deficient in Agpat6−/− mice. J. Lipid Res. 49, 823–831 (2008).
37. Verghes, L. et al. Agpat6 deficiency causes subdermal lipodystrophy and resistance to obesity. J. Lipid Res. 47, 745–754 (2006).
38. Alam, T. et al. Cloning and analysis of Nkx6.3 during CNS and gastrointestinal development. Gene. Expr. Patterns 6, 162–170 (2006).
39. Mihailov, A. et al. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Preprint at https://doi.org/10.1101/245506 (2018).
40. Varshney, A. et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc. Natl Acad. Sci. USA 114, 2301–2306 (2017).
41. Henseke, K. D. et al. NKKX6 transcription factor activity is required for α- and β-cell development in the pancreas. Development 132, 3139 (2005).
42. van de Bunt, M. et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. PLoS Genet. 11, e1005694 (2015).
43. Prokopenko, I. et al. A Central Role for GRB10 in regulation of islet function in man. PLoS Genet. 10, e1004235 (2014).
56. Zhemakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
57. Smith, W. W., Thomas, J., Liu, J., Li, T. & Moran, T. H. From fat fruitley to human obesity. *Physiol. Behav.* **6**, 15–21 (2014).
58. Arrese, E. L. & Soulages, J. L. Insect fat body: energy, metabolism, and regulation. *Annu. Rev. Entomol.* **55**, 207–225 (2010).
59. Pospisilik, J. A. et al. *Drosophila* genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* **140**, 148–160 (2010).
60. Freemuth, P. S. & Fitzsimons, H. L. The ERM protein Eesin is essential for neuronal morphogenesis and long-term memory in *Drosophila*. *Mol. Brain* **10**, 41 (2017).
61. Fitzsimons, H. L. & Scott, M. J. Genetic modulation of Rpd3 expression impairs long-term courtship memory in *Drosophila*. *PLoS ONE* **6**, e29171 (2011).
62. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Meth.* **9**, 179–181 (2012).
63. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
64. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
65. Kim, J. & Neufeld, T. P. Dietary sugar promotes systemic TOR activation in *Drosophila* through AKH-dependent selective secretion of Dilp3. *Nat. Commun.* **6**, 6846 (2015).
66. Veenstra, J. A., Agricola, H.-J. & Sellami, A. Regulatory peptides in fruit fly midgut. *Cell Tissue Res.* **334**, 499–516 (2008).
67. Mägi, R. & Morris, A. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinforma.* **11**, 288 (2010).
68. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
69. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
70. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).

Acknowledgements

We are grateful to Virginie Alexandre for useful scientific discussions and help with project management. We are also grateful to Olivier Simon, Radovan Chytracek, Richard Cote, and Ktsa Paluku for development of a clinical and omics data integration system. We thank Dr. Jan Veenstra for generous gifts of antibodies. The DiOGenes project was supported by the European Commission (Food Quality and Safety Priority of the Sixth Framework Program: FP6-2005- 513946). Local sponsors made financial contributions to the shop centers, which also received a number of foods free of charge from food manufacturers. A full list of these sponsors can be seen at [www.DiOGenes-eu.org/sponsors/](http://www.DiOGenes-eu.org/sponsors/). Functional analyses in fly were supported by the Australian government NHMRC (National Health and Medical Research Council) project grants APP1028887 and APP1124723. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Author contributions

A.V. and J.H. designed, supervised the study and interpreted results. A.V. designed and led statistical analyses. A.V. performed analyses with contributions from J.C., R.H. and G.L. Q.P.W., V.S., L.S. and T.C. performed analyses in flies, under supervision from N.G. and G.N. S.M. and C.C. performed genotyping under supervision from P.D. O.W., P.M. and P.L. managed biological samples. N.V. and D.L. contributed data and biological expertise. W.S. and A.A. designed the DiOGenes clinical study; R.D. and M.E.H. designed the Canadian program. A.V. wrote the manuscript with input from all authors. A.V. had primary responsibility for final content.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08492-8.

Competing interests: A.V., J.C., N.G., H.R., G.L., S.M., C.C., O.W., P.M., P.D. and J.H. are full-time employees at Nestlé Institute of Health Sciences SA. D.L. is a member of Institut Universitaire de France. W.H.S. reports having received research support from several food companies such as Nestlé, DSM, Unilever, Nutrition et Sante and Danone as well as Pharmaceutical companies such as GSK, Novartis and Novo Nordisk. He is an unpaid scientific advisor for the International Life Science Institute, ILSI Europe. A.A. reports grants or personal fees from McCain Foods, USA, personal fees from McDonald’s, USA, personal fees from Basic Research, USA, personal fees from Nestlé, Lausanne, personal fees from Dutch Beer Knowlidge Institute, NL, personal fees from Gelesis, USA, personal fees from Novo Nordisk, DK, personal fees from S-Biotek, DK, grants from Arla Foods, DK, grants from Danish Dairy Research Council, grants from Nordea Foundation, DK, outside the submitted work; and Royalties received for the book first published in Danish as Verdens Bedste Kur (Politiken, Copenhagen) and subsequently published in Dutch as Het beste dieet ter wereld (Kosmos Uitgevers, Utrecht/Antwerpen), and in English as World’s Best Diet (Penguin, Australia). G.N. is supported by an NHMRC career development fellowship II CDF1111940. The remaining authors declare no competing interests.

Reprints and permission information is available online at [http://npg.nature.com/reprintsandpermissions/](http://npg.nature.com/reprintsandpermissions/)

Journal peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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