

REVIEW ARTICLE

Open Access

Implications of publicly available genomic data resources in searching for therapeutic targets of obesity and type 2 diabetes

Sungwon Jung^{1,2}

Abstract

Obesity and type 2 diabetes (T2D) are two major conditions that are related to metabolic disorders and affect a large population. Although there have been significant efforts to identify their therapeutic targets, few benefits have come from comprehensive molecular profiling. This limited availability of comprehensive molecular profiling of obesity and T2D may be due to multiple challenges, as these conditions involve multiple organs and collecting tissue samples from subjects is more difficult in obesity and T2D than in other diseases, where surgical treatments are popular choices. While there is no repository of comprehensive molecular profiling data for obesity and T2D, multiple existing data resources can be utilized to cover various aspects of these conditions. This review presents studies with available genomic data resources for obesity and T2D and discusses genome-wide association studies (GWAS), a knockout (KO)-based phenotyping study, and gene expression profiles. These studies, based on their assessed coverage and characteristics, can provide insights into how such data can be utilized to identify therapeutic targets for obesity and T2D.

Introduction

Obesity and T2D are major public health problems, and their rates are increasing. It has been reported that 40% of adults in the UK will have obesity by 2025¹, and the worldwide population with T2D will approach 600 million in the next 20 years². Understanding the molecular mechanisms of these conditions is important to identify their therapeutic targets, but there has been limited success in identifying target genes because they are not genetic disorders in general aside from rare cases of clear genetic abnormalities, such as maturity onset diabetes of the young, Donohue syndrome, or Rabson-Mendenhall syndrome³. Another challenge is that they are generally not initiated from a single organ, unlike cancer. For

example, a major mechanism of obtaining T2D is acquiring insulin resistance, which may involve the accumulation of various environmental factors and multiple organs such as adipose, liver, and muscle are involved in that process. These characteristics imply that obesity and T2D result from abnormal dynamic states of relevant biological functions rather than aberrations of certain driver genes, which has created challenges in searching for simple therapeutic targets. For this reason, approaches to medically treat obesity or T2D are more about controlling the phenotypes of subjects, such as reducing caloric intake or appetite for obesity and decreasing blood glucose levels, increasing sensitivity to insulin, increasing insulin secretion, or using insulin therapy for T2D, rather than curing the disease by eliminating its drivers or altering the metabolic status back to a normal state.

Considering that obesity and T2D are due to abnormal dynamic states of relevant biological functions, it can be challenging to find therapeutic targets that can be applied

Correspondence: Sungwon Jung (sjung@gachon.ac.kr)

¹Department of Genome Medicine and Science, Gachon University School of Medicine, Incheon, Republic of Korea

²Gachon Institute of Genome Medicine and Science, Gachon University Gil Medical Center, Incheon, Republic of Korea

© The Author(s) 2018



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

to all subjects, and it may be necessary to identify different points of intervention for different subjects as an abnormality of the same biological function can be achieved from multiple points of aberration of molecular activities. For this reason, understanding the overall mechanisms and identifying the therapeutic candidates of obesity and T2D in the general population requires studying cohorts of sufficient size that are large enough to include variances in metabolic phenotypes and potentially diverse driving mechanisms, along with comprehensive data that can represent the exact status of individual subjects, such as detailed phenotypes and multi-omic profiles (such as genomic, epigenetic, metabolic,

proteomic profiles). However, research communities studying obesity and T2D lack such comprehensive data resources, which is unlike other diseases, such as cancer, where many comprehensive multi-omic data resources are publicly available.

Even though there are no comprehensive data resources for obesity and T2D, individual studies can constitute certain aspects of comprehensive data collections. This review will discuss currently available genomic data resources that can be utilized to identify therapeutic candidates for obesity and T2D, including GWAS, a KO-based phenotyping study, and gene expression studies that have observed expression changes in subjects with

Table 1 Selected popular GWAS related to obesity or T2D

Study	URL	Studied phenotypes related to obesity or T2D	Data availability
DIAGRAM consortium	http://www.diagram-consortium.org	T2D	Summary statistics (open to public)
EPIC-Norfolk study	http://www.srl.cam.ac.uk/epic/	Blood lipid profiles, BMI, body fat compositions, body measurements, glycated hemoglobin	Application required
Fenland study	http://www.mrc-epid.cam.ac.uk/research/studies/fenland	Blood insulin level, BMI, body fat compositions, body measurements, fasting blood glucose level, fasting lipid profiles, oral glucose tolerance, resting energy expenditure level, various blood metabolic profiles	Application required
GENESIS consortium ⁷	Not available	Blood lipid profiles, BMI, fasting glucose level, fasting insulin level, insulin sensitivity	PI contact required
GIANT consortium	http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium	BMI, body measurements	Summary statistics (open to public)
Global Lipids Genetics Consortium ⁸	http://lipidgenetics.org	Blood lipid profiles	Summary statistics (open to public)
GoT2D	http://www.type2diabetesgenetics.org/projects/got2d	T2D	Genetic data (application required), summary statistics (open to public)
InterAct project	http://www.inter-act.eu	Blood lipid profiles, T2D, various blood metabolic profiles	Application required
MAGIC consortium	https://www.magicinvestigators.org	Fasting glucose level, fasting insulin level, fasting proinsulin level, glucose-stimulated insulin secretion, HbA1c, insulin resistance/sensitivity, oral glucose tolerance	Summary statistics (open to public)
T2D-GENES	http://www.type2diabetesgenetics.org/projects/t2dGenes	T2D	Genetic data (application required), summary statistics (open to public)
UK Biobank	http://www.ukbiobank.ac.uk	Blood lipid profiles, body fat compositions, various blood metabolic profiles	Application required
UKHLS study	https://www.understandingsociety.ac.uk	Blood lipid profiles, BMI, body measurements, various blood metabolic profiles	Application required

Table 2 Selected major GWAS publications related to obesity or T2D

Authors	Year	Studied phenotype of association	Related GWAS from Table 1
Kathiresan et al. ³⁵	2008	Blood lipid profiles	GLGC
Sandhu et al. ³⁶	2008	LDL	EPIC-Norfolk
Willer et al. ³⁷	2008	Blood lipid profiles	GLGC
Zeggini et al. ³⁸	2008	T2D	DIAGRAM
Kathiresan et al. ³⁹	2009	Blood lipid profiles	GLGC
Prokopenko et al. ⁴⁰	2009	Fasting glucose levels	EPIC-Norfolk, MAGIC
Ingelsson et al. ⁴¹	2010	Fasting proinsulin, insulinogenic index, insulin sensitivity by intravenous methods, insulin sensitivity by oral glucose tolerance	MAGIC
Li et al. ⁴²	2010	BMI	EPIC-Norfolk
Saxena et al. ⁴³	2010	Oral glucose tolerance	Fenland, MAGIC
Speliotes et al. ⁴⁴	2010	BMI	Fenland, GIANT
Teslovich et al. ⁴⁵	2010	Blood lipid profiles	Fenland, GLGC
Voight et al. ⁴⁶	2010	T2D	DIAGRAM, GIANT, MAGIC
Dupuis et al. ⁴⁷	2010	Fasting glucose level, fasting insulin level, HOMA-B, HOMA-IR	Fenland, MAGIC
Strawbridge et al. ⁴⁸	2011	Fasting proinsulin level	DIAGRAM, Fenland, GIANT, MAGIC
Wang et al. ⁴⁹	2011	Obesity	
Manning et al. ⁵⁰	2012	Fasting glucose and insulin (while considering BMI)	DIAGRAM, Fenland, MAGIC
Morris et al. ⁵	2012	T2D	DIAGRAM
Scott et al. ⁵¹	2012	Fasting glucose, fasting insulin, oral glucose tolerance	DIAGRAM, Fenland, MAGIC
Xue et al. ⁵²	2012	BMI, waist and hip circumferences	EPIC-Norfolk
Yang et al. ⁵³	2012	BMI	EPIC-Norfolk, Fenland, GIANT
Berndt et al. ⁵⁴	2013	BMI, clinical classes of obesity, WHR	EPIC-Norfolk, Fenland, GIANT
den Hoed et al. ⁵⁵	2013	BMI, body fat percentage, overweight and obesity, waist circumference	EPIC-Norfolk, Fenland
Randall et al. ⁵⁶	2013	BMI, hip circumference, waist circumference, weight, WHR (while considering sex)	DIAGRAM, GIANT, MAGIC
van Vliet-Ostaptchouk ⁵⁷	2013	Influence of BMI, WHR associated variants on blood lipids and glycemic traits	EPIC-Norfolk, Fenland
Willer et al. ⁸	2013	Blood lipid profiles	EPIC-Norfolk, Fenland, GLGC
Langenberg et al. ⁵⁸	2014	Combination of genetic factors and other risk factors on T2D	InterAct
DIAGRAM Consortium et al. ⁵⁹	2014	T2D	DIAGRAM, T2D-GENES
Scott et al. ⁶⁰	2014	Insulin resistance/secretion, metabolic profiles along with T2D	InterAct
Gaulton et al. ⁶¹	2015	T2D	DIAGRAM
Locke et al. ⁶²	2015	BMI	DIAGRAM, GIANT, GLGC, MAGIC
Shungin et al. ⁶³	2015	Waist and hip circumferences	GIANT
Fuchsberger et al. ⁶⁴	2016	T2D	DIAGRAM, GoT2D, T2D-GENES
Knowles et al. ⁷	2016	Insulin sensitivity	GENESIS
Yaghootkar et al. ⁶⁵	2016	Low disease risk adiposity	EPIC-Norfolk, UK Biobank
Lotta et al. ⁹	2017		

Table 2 continued

Authors	Year	Studied phenotype of association	Related GWAS from Table 1
Graff et al. ⁶⁶	2017	Fasting insulin and lipid profiles along with adipose mass in peripheral compartments BMI, BMI-adjusted waist circumference, WHR (while considering physical activity)	DIAGRAM, EPIC-Norfolk, Fenland, GENESIS, GLGC, InterAct, MAGIC, UK Biobank, UKHLS EPIC-Norfolk, Fenland, GIANT, InterAct
Liu et al. ⁶⁷	2017	Blood lipid profiles	GLGC, InterAct
Prins et al. ⁶⁸	2017	Serum biomarker levels including lipid profiles	Fenland, UKHLS
Scott et al. ⁶⁹	2017	T2D	DIAGRAM

LDL low-density lipoprotein, *BMI* body mass index, *HOMA-B* homeostatic model assessment-beta, *HOMA-IR* homeostatic model assessment-insulin resistance, *WHR* waist-hip ratio

obesity and type 2 diabetes across relevant organs. The included data sets range from individual studies to large data sets curated by many international consortia. Utilizing these data sets and considering their characteristics can be an alternative approach that mimics comprehensive molecular profiling and provides a useful reference by curating customized genomic data sets to study therapeutic candidates for specific phenotypic conditions.

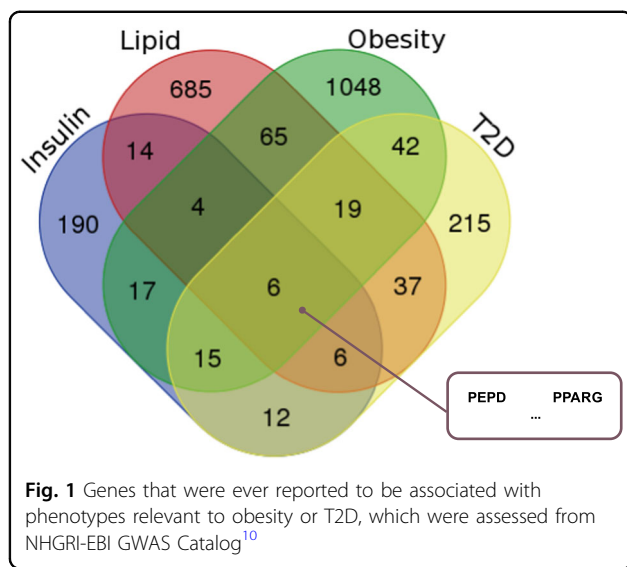
DNA-level susceptibility to obesity and T2D

Many early approaches to identify genetic effects on obesity and T2D were GWAS. GWAS observes known or candidate single-nucleotide polymorphisms (SNPs) and phenotypes that are related to obesity and T2D, where the statistical association between each SNP and phenotype is evaluated. Based on GWAS, it is possible to identify genes that have or are close to loci that are associated with susceptibility to the studied phenotypes. Unlike rare cases of diabetes with clear genetic drivers, variants at these susceptibility loci can have subtle effects on the function of relevant genes, as previous studies reported rather modest effect sizes of genetic variants on T2D that range from 10 to 35%^{4, 5}. Nevertheless, T2D is known to have a notable genetic basis, as the co-occurrence of T2D in monozygotic twins is significantly higher at ~70% frequency, whereas dizygotic twins showed a frequency of only 20–30%⁶. In normal populations with susceptibility loci, these subtle effects can generate long-term phenotypic differences in conjunction with other non-genetic, often environmental factors.

Table 1 lists selected popular GWAS that assessed phenotypes related to obesity or T2D. Most consortia or studies are based on a collection of cohorts, and it should be noted that occasionally, some cohorts are included in multiple consortia or studies. Phenotype information available from these individual cohorts may not be completely coherent with each other within a consortium or study. Thus, the phenotypes listed in Table 1 are those that each consortium or study made an effort to generate

via coherent collections and analyses. Some consortia or studies directly analyzed the association with disease outcomes (obesity or T2D; DIAGRAM, InterAct, GoT2D, and T2D-GENES), whereas others studied associations with more detailed phenotypes, such as body measurements or fat compositions (EPIC-Norfolk, Fenland, GENESIS⁷, GIANT, UK Biobank, and UKHLS), lipid profiles (EPIC-Norfolk, Fenland, GENESIS⁷, GLGC⁸, InterAct, UK Biobank, and UKHLS), and insulin resistance/sensitivity (Fenland, GENESIS⁷, and MAGIC). Individual-level genetic data are rarely available except for a few that accept applications; thus, it is difficult to collect individual-level raw genetic data from multiple cohorts together with phenotypic information to conduct an association analysis. However, the analyzed summary statistics of statistical associations between SNPs and phenotypes are often publicly available, where *p*-values of statistical significance, frequencies in cohorts, and effect sizes are available in general, and this information is useful for designing and conducting a meta-analysis of interest.

Table 2 lists selected major GWAS publications that assessed genetic associations with phenotypes relevant to obesity or T2D. Studied phenotypes are listed for each work, but it should be noted that most studies consider additional phenotypes for the adjustment of statistical associations or prioritization of associated variants. Most studies are meta-analyses that utilize multiple cohorts from several consortia or studies. A general approach of these meta-analyses is to identify novel loci with susceptibility by increasing the size of population with multiple cohorts or by providing independent evidential support for the identified novel loci by using extra cohorts as independent validation data. Another approach of meta-analysis is systematically integrating the results of multiple GWAS of various phenotypes to model certain types of conditions or diseases. A good example of this type of meta-analysis is the work by Lotta et al.⁹, where they identified candidate loci that are associated with lipodystrophy-like phenotypes by integrating the results of



several GWAS consortia. Most studies provide a list of identified loci, and some studies also provide more detailed summary statistics through their related consortia. A novel meta-analysis of GWAS can be designed to study genetic loci susceptible to specific combinations of phenotypes by integrating GWAS summary statistics that were derived from analyzing associations with individual phenotypes.

In addition to GWAS-derived data sources from individual consortia or studies, there are online data resources in which previous GWAS results are curated and can be accessed with user-friendly interfaces. NHGRI-EBI GWAS Catalog¹⁰ provides searches and visualization of published SNP-trait associations and bulk download of its contents for systematic analysis. It currently contains 63,205 unique SNP-trait associations from >3200 publications, and it contains GWAS on phenotypes other than obesity and T2D. Type 2 Diabetes Knowledge Portal¹¹ is a T2D-focused online data portal in which 22 GWAS/exome chip/whole genome sequencing/exome sequencing data sets are curated with association information for 47 traits. It provides user interfaces that can simulate the systematic integration of multiple GWAS with various phenotypes, where users can search for variants of interests in individual GWAS data sets from participating consortia and form combinations. However, it does not provide bulk download of entire integrated data sets. These data portals provide the functionality of various searches on diseases, genes, phenotypes, or variants.

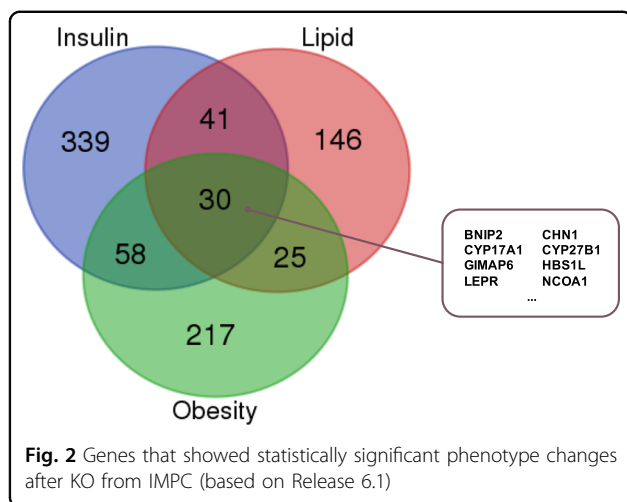
Available GWAS results cover associations with various phenotypes that are related to obesity or T2D, most of which belong to one of four categories: insulin resistance/sensitivity-related phenotypes, lipid profile-related phenotypes, outcome of obesity, and outcome of T2D. For a

better understanding of gene coverages that are associated with these phenotypes, genes that have been associated with any of the four phenotype categories were collected from the NHGRI-EBI GWAS Catalog¹⁰. Specifically, the bulk GWAS result data of all 63,205 SNPs that have ever been reported to be associated with phenotypes were obtained, and SNPs that were associated with phenotypes of at least one of the four categories were collected. For each SNP with such an association, a gene that includes the SNP was determined to be associated with the corresponding phenotype, or a gene that is closest to the SNP was determined to be associated if the SNP was in an intergenic location. Fig. 1 shows a Venn diagram of the 2375 genes that are associated with at least one of the four obesity/T2D-related phenotype categories. A certain degree of common genes is shown, but each phenotype category has its own genes of exclusive associations. The six genes that show associations with all four categories of phenotypes include the well-known peroxisome proliferator activated receptor gamma (PPARG), where PPARG is a regulator of adipocyte differentiation¹² and has been implicated in numerous diseases, including obesity¹³ and T2D¹⁴. Another gene is peptidase D (PEPD), and it is known to play an important role in collagen metabolism¹⁵.

As already mentioned, the direct effect size of GWAS-identified loci to obesity/T2D-related phenotypes is relatively small. It should be noted that the genes related to GWAS-identified loci imply the biological functions of certain roles in developing metabolic disorders rather than these genes being decisive disease drivers. For this reason, considering the genes from GWAS generally requires further direct validation of the mechanisms that drive these metabolic disorders.

Causal gene identification with gene KO mouse models

GWAS takes a passive observational approach that searches for associations between the phenotypes of interest and genetic variants in real populations. For this reason, it is challenging to uncover specific mechanisms of action from the identified susceptible loci as they can explain marginal effect sizes in general. In comparison, understanding the function of genes by knocking them out in model species and observing the resulting phenotypes is an extreme interventional approach. In this approach, knocking-out each gene is done for model species and the resulting phenotypes are observed based on predefined protocols. A good example of this approach is the International Mouse Phenotyping Consortium (IMPC)¹⁶, where the objective is producing KO mouse lines for >20,000 known genes and observing various resulting phenotypes with standardized protocols. It is an international consortium of multiple institutions, and



these institutions produce germ line transmissions of targeted KO mutations in embryonic stem cells for known/predicted mouse genes. Each mutant mouse line is tested through a standardized primary phenotyping pipeline (see the website of the consortium for a complete list of studied phenotypes) in all major adult organ systems and most areas of major human disease. Briefly, phenotypes are observed from embryonic status until the 16th week and include fatality, body measurements and compositions, metabolic profiles, insulin-related phenotypes, pathological, physical, and physiological phenotypes. It is an ongoing project, and the current release (Release 6.1) includes phenotype information from knock outs of 3371 mouse genes. IMPC provides online search functionality for genes, diseases, and phenotypes, and detailed phenotype information is provided if available for queried KO models.

Among the studied phenotypes from IMPC, phenotypes relevant to obesity or T2D can also be grouped into the following three categories: insulin resistance/sensitivity-related phenotypes, lipid profile-related phenotypes, and obesity-related phenotypes, such as weight changes. Among the 3371 studied IMPC genes, genes that showed statistically significant changes in phenotypes that belong to any of the three categories were assessed from IMPC Release 6.1. Fig. 2 shows the Venn diagram of 856 genes that caused these statistically significant phenotypic changes for each phenotype category. Like the case of GWAS-identified genes, genes from KO-based phenotyping studies also show a certain degree of overlap and unique genes in each phenotype category. There are 30 genes that show changes in all three phenotype categories, and they include previously known genes involved in energy transfer and metabolism. CHN1 is a GTPase-activating protein¹⁷, BNIP2 is related to myogenesis¹⁸ and GTPase activator activity¹⁹, and HBS1L and GIMAP6²⁰ are related to GTP binding. NCOA1 is involved in

controlling the energy balance between white and brown adipose tissues²¹. CYP17A1 and CYP27B1 are members of the cytochrome P450 superfamily of enzymes²², and they are monooxygenases that catalyze many reactions involved in drug metabolism and the synthesis of cholesterol, steroids and other lipids. LEPR is a receptor for leptin and is involved in the regulation of fat metabolism²³.

The advantage of this KO-based phenotyping approach is its direct observation of resulting phenotypes from individual gene KO, which minimizes the undesirable effects of other factors in analyzing the biological function of the target gene. However, there are a few challenges with this approach. Establishing KO mouse models itself is a challenging task, often requiring significant time and effort. Controlling the quality of the standardized phenotyping protocol can also be a technical obstacle, especially when multiple independent organizations collaborate internationally. There is also an inherent limitation that lethal genes are hard to study with this approach, as KO of these genes will disable producing adult mouse lines and the following phenotyping processes. In addition to such challenges in a KO-based phenotyping approach, a few characteristics should be noted before utilizing the phenotyping results of gene KO. Current phenotyping protocols are focused on identifying phenotypes in normal environments (for example, feeding normal chow); thus, these studies do not represent possible phenotypic changes under certain environmental stresses of interest (for example, a high fat diet) that were not considered in the phenotyping protocols. As this approach is conducted based on model species, potential discrepancies between the model species and humans should be considered. Another issue is that this approach performs KO of genes in the whole body rather than tissue-specific silencing, whereas in realistic situations, several relevant organs can have individual roles via specific biological functions in developing metabolic disorders. Thus, consideration of the genes from KO-based phenotyping studies requires an understanding on these pros and cons and their relationships with human disease mechanisms.

Human gene expression profiling of obesity and T2D

A metabolic disorder is a condition in which the dynamic status of in vivo metabolism falls into disorder throughout the body (for example, insulin-resistant state of T2D). Thus, developing effective therapeutic approaches can require an understanding of the exact dynamic states of metabolic systems within the body of individual patients. This understanding of exact dynamical states of in vivo metabolic systems can require the following considerations. First, comprehensive molecular profiling is

necessary to form broad multi-omic observations, including gene expression, protein expression, and metabolic profiles. Second, this comprehensive molecular profiling needs to be conducted on various relevant organs, such as adipose, liver, and muscle to study insulin resistance. However, gene expression profiling is the only relatively popular approach for high-throughput molecular profiling due to its advantages of higher reliability and lower costs than the other techniques. There are also certain challenges in acquiring the human tissue samples needed for molecular profiling as surgical treatment is not a general treatment for obesity or T2D. For these reasons, few studies are currently available that have conducted comprehensive molecular profiling in various relevant organs, even when only gene expression is considered.

Nevertheless, some studies have conducted gene expression profiling in specific organs in certain conditions of interest. Like the case of GWAS with various phenotypes, appropriate integration of these data sets can enable data set assessment in a way that mimics comprehensive multi-organ profiling. To integrate multiple gene expression profiles from independent studies, normalization of data sets between data sets is required to achieve data-level coherency. The most desirable normalization of data sets requires all data sets to be generated from the same platform; however, gene expression profiling has been performed with various microarray and next-generation sequencing platforms. There are many different platforms for gene expression profiling, but the most popular platform with the largest number of studies is the Affymetrix GeneChip Human Genome U133 Plus 2.0 microarray, despite recent advancements in next-generation sequencing platforms. Table 3 lists the studies on obesity or T2D with available gene expression profiles based on the Affymetrix GeneChip Human Genome U133 Plus 2.0 microarray. Most studies profiled samples of only one tissue, except for two data sets (GSE13070 and GSE41168). The approaches of the studies vary, such as studying gene expression profiles of disease only, comparing disease profiles with normal control profiles, comparing profiles across different stages of disease, comparing profiles before and after certain interventions, and comparing profiles from siblings or twins to reduce the effect of genetic backgrounds. From this collection of expression profiles of various conditions performed with the same profiling platform (as listed in Table 3), gene expression profiles from multiple studies can be integrated into a single normalized data set so that the subject conditions of the studies match our conditions of interest.

As a simple example of integrating gene expression profiles of several studies with subjects of interest, differentially expressed genes (DEGs) between lean healthy subjects and obese healthy or obese diabetic subjects were identified in a tissue-specific way. From the 20 data sets

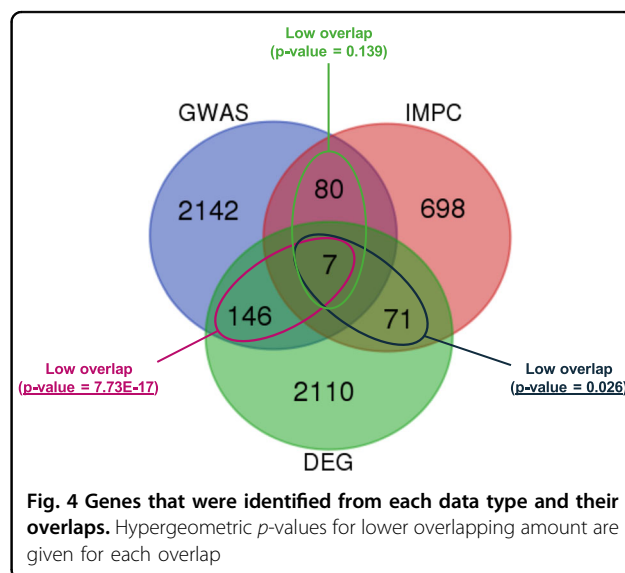
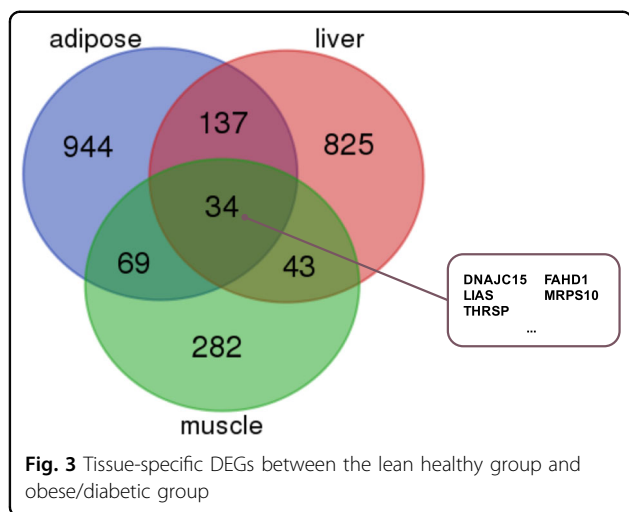
listed in Table 3, 17 studies (except for E-TABM-325, GSE27916, and E-MTAB-1895) provide BMI information and metabolic profiles or insulin resistance/sensitivity information. A total of 602 gene expression profiles of adipose, liver, and muscle samples from the 17 studies were integrated into a single data set, where lean/obese conditions of the samples were determined based on BMI and healthy/diabetic conditions of the samples were determined based on the metabolic profiles and insulin resistance/sensitivity information. For each tissue type, a gene was declared as a DEG if it showed more than a 1.5-fold change in expression with an FDR-adjusted p -value $< 1E-6$ (t -test) between lean healthy samples and obese/diabetic samples. Fig. 3 shows the Venn diagram of 2334 DEGs identified from three tissue types. Due to tissue-specific gene expression, many DEGs are differentially expressed in a tissue-specific manner. For example, PPARG is an adipose-specific DEG, which is a regulator of adipocyte differentiation. There are 34 common DEGs that show differential expressions from all three tissue types. Five of these 34 DEGs are known to be related to metabolism or mitochondria. FAHD1 is related to tyrosine metabolism and a mitochondrial enzyme²⁴, and THRSP is related to regulation of lipid metabolism and lipogenesis²⁵. DNAJC15²⁶ is a negative regulator of the mitochondrial respiratory chain, prevents mitochondrial hyperpolarization states and restricts mitochondrial generation of ATP, MRPS10 is a mitochondrial ribosomal protein, and LIAS is localized in mitochondria and known to be associated with hyperglycemia²⁷. Note that they are DEGs common to all tissue types, and the relevance to mitochondria and metabolism may not be tissue-specific. Compared to DNA-level genetic variants, which make a relatively small contribution to effect sizes, DEGs of significant expression changes from phenotypes of interest can imply more direct representation of the biological mechanisms that drive such phenotypes because these expression changes are a snapshot of the current biological dynamic status. Thus, searching therapeutic targets based on gene expression profiles may provide higher chances of identifying points of intervention compared to searching solely based on DNA-level susceptible genetic variants. However, it should be noted that gene expression profiles are based on transcription profiles; thus, they have their own limitations. First, there can be discrepancies between transcription-level activities and protein levels or metabolic activity levels, as there are many post-transcriptional regulatory mechanisms, such as small RNA activities. Second, identifying key driver events of these transcriptional changes is still a challenge. Nevertheless, publicly available gene expression profiles from relevant studies of obesity and T2D are important and beneficial resources as they provide unique information

Table 3 Gene expression data sets studying obesity or T2D generated with Affymetrix GeneChip Human Genome U133 Plus 2.0 microarray

Contributors	Year	Data set ID	Tissue	Subject	Number of samples (CEL files)	Study type
Park et al. ⁷⁰	2006	GSE5109	Muscle	Morbidly obese patients before and after gastric bypass surgery	6	Disease only
Pietilainen et al. ⁷¹	2008	E-MEXP-1425	Adipose	Monozygotic twins discordant for BMI	27	Twins
Palsgaard et al. ⁷²	2009	E-MEXP-2559	Muscle	Subjects with T2D, first degree relatives with T2D, and healthy control	35	Normal vs. disease
Sears et al. ⁷³	2009	GSE13070	Adipose, muscle	Insulin-resistant and insulin-sensitive subjects, before and after hyperinsulinemic-euglycemic clamp at baseline and after three-month TZD treatment	364	Normal vs. disease, with intervention
Misu et al. ⁷⁴	2010	GSE23343	Liver	Subjects with T2D and independent subjects with normal glucose tolerance	17	Normal vs. disease
Gallagher et al. ⁷⁵	2010	GSE18732	Muscle	Subjects with normal glucose tolerance, with glucose intolerance, and with T2D	118	Normal vs. stages of disease
Jin et al. ⁷⁶	2011	GSE25462	Muscle	Subjects with T2D, normal subjects with family history of T2D, and normal subjects with no history of T2D	50	Normal vs. disease
Keller et al. ⁷⁷	2011	GSE27949	Adipose	Subjects with normal glucose tolerance, with glucose intolerance, and with T2D	33	Normal vs. stages of disease
Hardy et al. ⁷⁸	2011	GSE15773	Adipose	Insulin-resistant obese and insulin-sensitive obese subjects undergoing gastric bypass surgery	19	Obese vs. insulin resistance
Hardy et al. ⁷⁸	2011	GSE20950	Adipose	Insulin-resistant obese and insulin-sensitive obese subjects undergoing gastric bypass surgery	39	Obese vs. insulin resistance
Soronen et al. ⁷⁹	2012	GSE26637	Adipose	Lean insulin-sensitive females and obese insulin-resistant females at fasting and after three hours hyperinsulinemia	20	Normal vs. disease, with intervention
Alligier et al. ⁸⁰	2012	GSE28005	Adipose	Healthy lean subjects and overweight subjects, with high fat diet – at day 0, 14, and 56	38	Normal vs. disease, with intervention
van Tienen et al. ⁸¹	2012	GSE19420	Muscle	Insulin-treated T2D subjects before and after 52 weeks of exercise training, normoglycemic control subjects, and pre-diabetes subjects	42	Normal vs. disease, with intervention
Min et al. ⁸²	2012	E-TABM-325	Adipose	Normoglycemic twins	208	Twins
Murphy et al. ⁸³	2013	GSE49541	Liver	Subjects with NAFLD	72	Disease
Yoshino et al.	2013	GSE41168	Adipose, muscle	Non-obese female before and after placebo, resveratrol supplementation, and calorie restriction intervention	140	Normal, before and after intervention
Nookaew et al. ⁸⁴	2013	GSE27916	Adipose	Sibling subjects with obesity	375	Sibling pair
Naukkarinen et al. ⁸⁵	2013	E-MTAB-1895	Adipose	Monozygotic twins discordant for obesity	52	Twins
Lopez-Vicario et al. ⁸⁶	2014	GSE37031	Liver	Subjects with NASH and normal control subjects	15	Normal vs. disease
Frades et al. ⁸⁷	2015	GSE63067	Liver	Subjects with steatosis, with NASH, and healthy control subjects	18	Normal vs. stages of disease

Data sets with IDs beginning with "GSE" are available at Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), and data sets with IDs beginning with "E-" are available at ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>). All data sets provide raw CEL files of microarrays.

T2D Thiazolidinedione, NAFLD non-alcoholic fatty liver disease, NASH non-alcoholic steatohepatitis



on dynamical gene-regulations that cannot be inferred from DNA-level phenotype associations.

Comparing biological coverage of GWAS, KO-based phenotyping, and gene expression profiles

To compare the coverage of obesity/T2D-related genes that can be identified from currently available data from GWAS, KO-based phenotyping, and gene expression profiles, the genes that were identified from different data types were compared to one another. Fig. 4 illustrates the Venn diagram of the obesity/T2D-related genes that were identified from each data type in the previous sections and the amount of overlap between them. The identified genes show very little overlap between different data types, where DEGs from gene expression profiles show significantly low overlap with the other two data types (*p*-value of low overlap: DEG–GWAS = 7.73E-17, DEG–IMPC = 0.026). The overlap between the genes identified from GWAS and the KO phenotyping study is also very low, but its statistical significance is not as strong as the other cases. This low commonality between the obesity/T2D-related genes from different data types suggests that their different approaches to assessing the relationships between genes and phenotypes cause biases in the coverage of identified genes. The discrepancy is clearer between the DEGs from gene expression profiles and the genes from the other two data types, suggesting that gene expression-level changes and DNA-level genetic effects may cover different biological aspects. This difference in coverage between the results of studying gene expression profiles and the results of studying DNA-level genetics becomes more evident when their enriched biological functions are compared one another. For each list of genes identified from studies of gene expression profiles, GWAS, and KO-based phenotyping, the statistical enrichment levels of known biological functions were

evaluated to identify the most strongly relevant biological functions for each list of genes. Molecular Signatures Database^{28, 29} is a collection of annotated gene sets, where 17,774 gene sets are curated with a related list of genes (Molecular Signatures Database v5.2). Among these data, each of the 6659 gene sets that represent known biological pathways (curated from pathway databases, such as KEGG³⁰ and REACTOME^{31, 32}) and Gene Ontology^{33, 34} biological processes and molecular functions was evaluated for its overlap with each list of genes identified from gene expression profiles, GWAS, and KO-based phenotyping, and the statistical significance of overlap was computed as a hypergeometric *p*-value. For the list of genes from each data type, biological functions with an FDR-adjusted *p*-value < 1E-10 were declared as the most strongly relevant functions, and Fig. 5a shows the Venn diagram of the most strongly relevant biological functions for the three data types. The biological functions that are very strongly enriched in the genes that showed obesity/T2D-related phenotypes from KO-based phenotyping (IMPC) were mostly discovered by other data types except for one function, whereas 37 biological functions were discovered by both GWAS and gene expression profile-based analysis, and three functions were also discovered by gene expression profile-based analysis. The biological functions from gene expression profile-based studies show large discrepancies with those from GWAS, which strongly implies differences in the biological coverage of gene expression profiles and DNA-level genetic susceptibility information. Fig. 5b illustrates the very strongly enriched biological functions for different data types that are relevant to obesity/T2D, and it shows different biological mechanisms that are specifically enriched in DEGs from gene expression profiles. From Fig. 5b, the list of genes from gene expression profiles, GWAS, and KO-

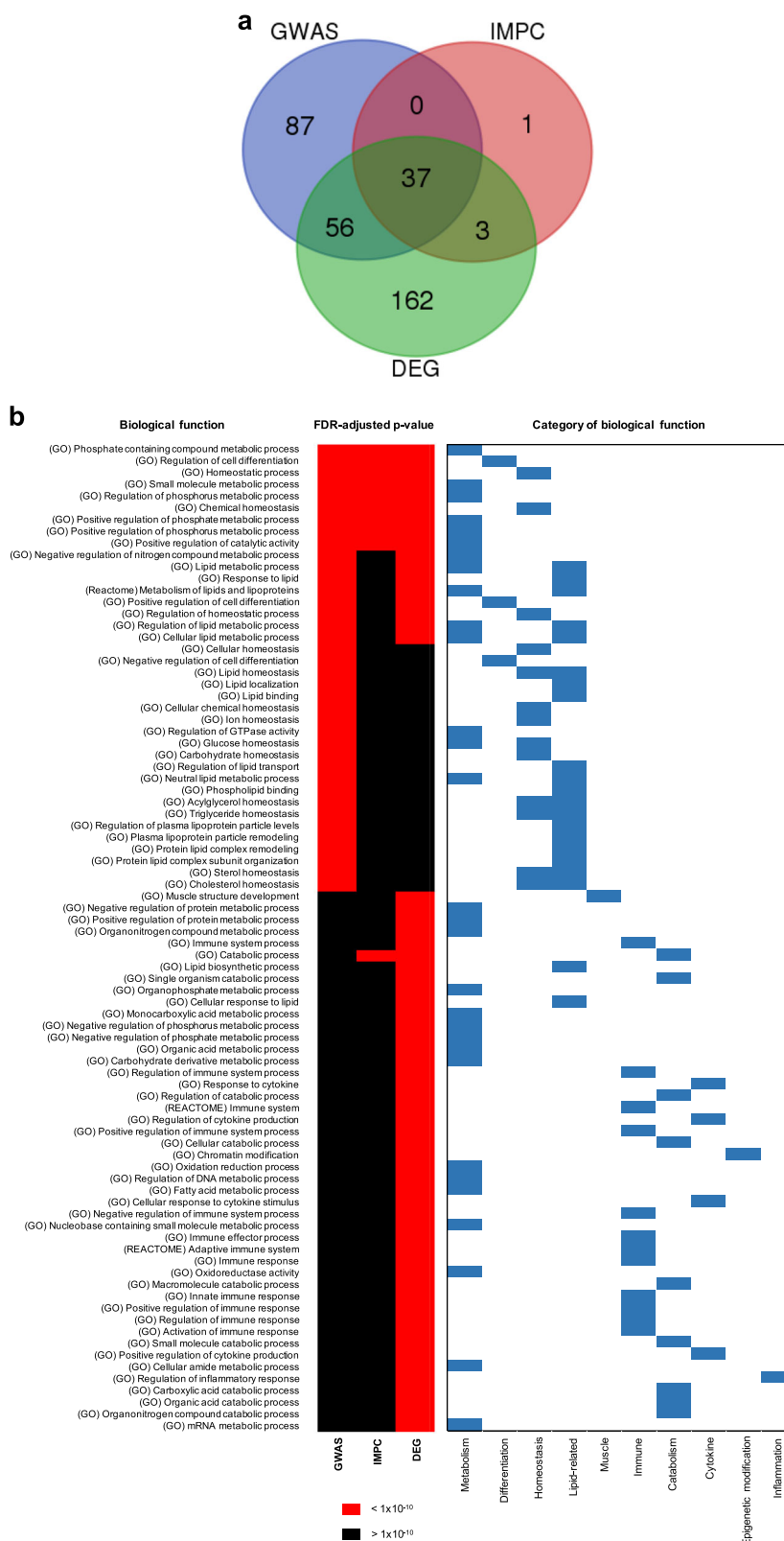


Fig. 5 Biological functions that are very strongly enriched (FDR-adjusted p -value <math>< 1 \times 10^{-10}</math>) in the list of obesity/T2D-related genes from each data type . **a The Venn diagram of the very strongly enriched biological functions. **b** Very strongly enriched biological functions for at least one data type, while being relevant to Obesity/T2D-relevant. For each biological function, its related functional categories are also presented**

based phenotyping commonly have strongly enriched biological functions that are related to metabolism, differentiation, homeostasis, and lipids. However, biological functions that are related to muscle, immune, catabolism, cytokine, epigenetic modification, and inflammation are specifically enriched in the genes from gene expression profiles in general. This finding implies that genes involved in such biological functions are more affected by dynamic gene expression changes than by static genetic backgrounds. These results emphasize that we need to consider all discrepancies in gene coverage and biological functions that can be identified with different data types in searches for therapeutic targets and strategies.

Conclusion

Many efforts to understand obesity and T2D and find their therapeutic targets have been made. However, few data resources exist with comprehensive high-throughput molecular profiles for obesity or T2D whereas such comprehensive molecular information is essential for understanding these conditions. In this review, publicly available genomic data resources of obesity and T2D are discussed, covering major GWAS, a KO-based phenotyping study, and studies with gene expression profiles based on a popular microarray platform. While no comprehensive data resource is available, systematic integrations of these individual data sources based on their associated phenotypes and experimental conditions give us a chance to mimic comprehensive collections of genomic data. GWAS and the KO-based phenotyping study provided insights into the function of individual genes, whereas gene expression profiles provided complementary opportunities to observe dynamical systematic changes of biological functions that could not be observed with DNA-level information. A comparison of obesity/T2D-associated genes that were identified from different data types showed different coverage of identifiable genes, and a comparison of their enriched biological functions provided stronger clues into the biological discrepancies that can be recognized with different data types. Thus, utilizing these data resources for own studies with specific disease models requires the consideration of such discrepancies in data characteristics and coverage.

From this point of view, a desirable approach to building a comprehensive molecular profile for obesity or T2D requires consideration of the following. First, a cohort must be broadly collected so that it can represent various ranges of metabolic conditions as metabolic conditions, such as obesity or T2D, are continuously developed with varying states of metabolic dynamics. Second, a comprehensive collection of phenotypes must be monitored to precisely model the progression status of metabolic conditions. Third, a collection of tissue samples for relevant organs must be collected from individuals in the cohort as

several organs participate in the development of metabolic conditions. Lastly, efforts should be put towards making the molecular profiles of tissue samples as comprehensive as possible by covering various levels of molecular mechanisms, including information at the DNA, transcript or gene expression, epigenetic, protein, and metabolic profile levels. Such comprehensive molecular profiling from human multiple organs (if possible) or even organs from model species will give us information on molecular activities in obesity and T2D with an unparalleled level of resolution, and this rich information will become a solid basis for searching for therapeutic targets and developing treatment strategies.

Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government Ministry of Science and Information & Communication Technology (MSIT) (2016R1C1B2016354). This work was also supported by a Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry for Health and Welfare, Korea (HI14C1135).

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 28 December 2017 Accepted: 28 January 2018.

Published online: 20 April 2018

References

1. ButlandJebb, B. S. K. P., McPherson, K., Thomas, S., Mardell, J. & Parry, V. *Tackling obesity: future choices—Project Report*. (Government Office for Science, London, 2007).
2. DiabetesUK. *Diabetes: Facts and Stats*. (2014).
3. Melmed, S., Polonsky, K. S., Larsen, P. R. & Kronenberg, H. M. *Williams Textbook of Endocrinology*. (Elsevier, Philadelphia, PA, USA, 2011).
4. McCarthy, M. I. Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.* **363**, 2339–2350 (2010).
5. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
6. Kaprio, J. et al. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* **35**, 1060–1067 (1992).
7. Knowles, J. W. et al. Identification and validation of N-acetyltransferase 2 as an insulin sensitivity gene. *J. Clin. Invest* **126**, 403 (2016).
8. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
9. Lotta, L. A. et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).
10. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
11. Type 2 Diabetes Knowledge Portal. [Internet]. Available from: <http://www.type2diabetesgenetics.org/>.
12. Rosen, E. D., Walkey, C. J., Puigserver, P. & Spiegelman, B. M. Transcriptional regulation of adipogenesis. *Genes Dev.* **14**, 1293–1307 (2000).
13. Sharma, A. M. & Stael, B. Review: peroxisome proliferator-activated receptor gamma and adipose tissue—understanding obesity-related changes in regulation of lipid and glucose metabolism. *J. Clin. Endocrinol. Metab.* **92**, 386–395 (2007).

14. Celi, F. S. & Shuldiner, A. R. The role of peroxisome proliferator-activated receptor gamma in diabetes and obesity. *Curr. Diab. Rep.* **2**, 179–185 (2002).
15. Surazynski, A., Miltyk, W., Palka, J. & Phang, J. M. Prolidase-dependent regulation of collagen biosynthesis. *Amino Acids* **35**, 731–738 (2008).
16. Koscielny, G. et al. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res.* **42**, D802–D809 (2014).
17. Miyake, N. et al. Human CHN1 mutations hyperactivate alpha2-chimaerin and cause Duane's retraction syndrome. *Science* **321**, 839–843 (2008).
18. Kang, J. S. et al. A Cdo-Bnip-2-Cdc42 signaling pathway regulates p38alpha/beta MAPK activity and myogenic differentiation. *J. Cell Biol.* **182**, 497–507 (2008).
19. Low, B. C., Lim, Y. P., Lim, J., Wong, E. S. & Guy, G. R. Tyrosine phosphorylation of the Bcl-2-associated protein BNIP-2 by fibroblast growth factor receptor-1 prevents its binding to Cdc42GAP and Cdc42. *J. Biol. Chem.* **274**, 33123–33130 (1999).
20. Krucken, J. et al. Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. *Gene* **341**, 291–304 (2004).
21. Picard, F. et al. SRC-1 and TIF2 control energy balance between white and brown adipose tissues. *Cell* **111**, 931–941 (2002).
22. Nebert, D. W., Wilkval, K. & Miller, W. L. Human cytochromes P450 in health and disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120431 (2013).
23. Harris, R. B. Direct and indirect effects of leptin on adipocyte metabolism. *Biochim. Biophys. Acta* **1842**, 414–423 (2014).
24. Pircher, H. et al. Identification of human fumarylacetoacetate hydrolase domain-containing protein 1 (FAHD1) as a novel mitochondrial acylpyruvase. *J. Biol. Chem.* **286**, 36500–36508 (2011).
25. Feng, X., Jiang, Y., Meltzer, P. & Yen, P. M. Thyroid hormone regulation of hepatic genes in vivo detected by complementary DNA microarray. *Mol. Endocrinol.* **14**, 947–955 (2000).
26. Hatle, K. M. et al. MCJ/DnajC15, an endogenous mitochondrial repressor of the respiratory chain that controls metabolic alterations. *Mol. Cell Biol.* **33**, 2302–2314 (2013).
27. Baker, P. R. 2nd et al. Variant non ketotic hyperglycinemia is caused by mutations in LIAS, BOLA3 and the novel gene GLRX5. *Brain* **137**, 366–379 (2014).
28. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
29. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
30. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
31. Croft, D. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
32. Fabregat, A. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
33. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
34. The Gene Ontology C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
35. Kathiresan, S. et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**, 189–197 (2008).
36. Sandhu, M. S. et al. LDL-cholesterol concentrations: a genome-wide association study. *Lancet* **371**, 483–491 (2008).
37. Willer, C. J. et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
38. Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
39. Kathiresan, S. et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65 (2009).
40. Prokopenko, I. et al. Variants in MTNR1B influence fasting glucose levels. *Nat. Genet.* **41**, 77–81 (2009).
41. Ingelsson, E. et al. Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. *Diabetes* **59**, 1266–1275 (2010).
42. Li, S. et al. Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *Am. J. Clin. Nutr.* **91**, 184–190 (2010).
43. Saxena, R. et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
44. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
45. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
46. Voight, B. F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
47. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
48. Strawbridge, R. J. et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–2634 (2011).
49. Wang, K. et al. A genome-wide association study on obesity and obesity-related traits. *PLoS ONE* **6**, e18939 (2011).
50. Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
51. Scott, R. A. et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
52. Xue, F. et al. A latent variable partial least squares path modeling approach to regional association and polygenic effect with applications to a human obesity study. *PLoS ONE* **7**, e31927 (2012).
53. Yang, J. et al. FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–272 (2012).
54. Berndt, S. I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
55. den Hoed, M. et al. Evaluation of common genetic variants identified by GWAS for early onset and morbid obesity in population-based samples. *Int J. Obes.* **37**, 191–196 (2013).
56. Randall, J. C. et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9**, e1003500 (2013).
57. van Vliet-Ostapchouk, J. V. et al. Pleiotropic effects of obesity-susceptibility loci on metabolic traits: a meta-analysis of up to 378,774 individuals. *Diabetologia* **56**, 2134–2146 (2013).
58. Langenberg, C. et al. Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. *PLoS Med.* **11**, e1001647 (2014).
59. Replication DIG, Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes C, South Asian Type 2 Diabetes C, Mexican American Type 2 Diabetes C, Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples C. et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
60. Scott, R. A. et al. Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. *Diabetes* **63**, 4378–4387 (2014).
61. Gaulton, K. J. et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
62. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
63. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
64. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
65. Yaghootkar, H. et al. Genetic evidence for a link between favorable adiposity and lower risk of type 2 diabetes, hypertension, and heart disease. *Diabetes* **65**, 2448–2460 (2016).
66. Graff, M. et al. Genome-wide physical activity interactions in adiposity—a meta-analysis of 200,452 adults. *PLoS Genet* **13**, e1006528 (2017).
67. Liu, D. J. et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758–1766 (2017).
68. Prins, B. P. et al. Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Sci. Rep.* **7**, 11008 (2017).

69. Scott, R. A. et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
70. Park, J. J., Berggren, J. R., Hulver, M. W., Houmard, J. A. & Hoffman, E. P. GRB14, GPD1, and GDF8 as potential network collaborators in weight loss-induced improvements in insulin action in human skeletal muscle. *Physiol. Genom.* **27**, 114–121 (2006).
71. Pietiläinen, K. H. et al. Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med.* **5**, e51 (2008).
72. Palsgaard, J. et al. Gene expression in skeletal muscle biopsies from people with type 2 diabetes and relatives: differential regulation of insulin signaling pathways. *PLoS ONE* **4**, e6575 (2009).
73. Sears, D. D. et al. Mechanisms of human insulin resistance and thiazolidinedione-mediated insulin sensitization. *Proc. Natl Acad. Sci. USA* **106**, 18745–18750 (2009).
74. Misu, H. et al. A liver-derived secretory protein, selenoprotein P, causes insulin resistance. *Cell Metab.* **12**, 483–495 (2010).
75. Gallagher, I. J. et al. Integration of microRNA changes in vivo identifies novel molecular features of muscle insulin resistance in type 2 diabetes. *Genome Med.* **2**, 9 (2010).
76. Jin, W. et al. Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. *J. Clin. Invest.* **121**, 918–929 (2011).
77. Keller, P. et al. Gene-chip studies of adipogenesis-regulated microRNAs in mouse primary adipocytes and human obesity. *BMC Endocr. Disord.* **11**, 7 (2011).
78. Hardy, O. T. et al. Body mass index-independent inflammation in omental adipose tissue associated with insulin resistance in morbid obesity. *Surg. Obes. Relat. Dis.* **7**, 60–67 (2011).
79. Soronen, J. et al. Adipose tissue gene expression analysis reveals changes in inflammatory, mitochondrial respiratory and lipid metabolic pathways in obese insulin-resistant subjects. *BMC Med. Genom.* **5**, 9 (2012).
80. Alligier, M. et al. Subcutaneous adipose tissue remodeling during the initial phase of weight gain induced by overfeeding in humans. *J. Clin. Endocrinol. Metab.* **97**, E183–E192 (2012).
81. van Tienen, F. H. et al. Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes. *J. Clin. Endocrinol. Metab.* **97**, 3261–3269 (2012).
82. Min, J. L. et al. Coexpression network analysis in abdominal and gluteal adipose tissue reveals regulatory genetic loci for metabolic syndrome and related phenotypes. *PLoS Genet.* **8**, e1002505 (2012).
83. Murphy, S. K. et al. Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. *Gastroenterology* **145**, 1076–1087 (2013).
84. Nookaew, I. et al. Adipose tissue resting energy expenditure and expression of genes involved in mitochondrial function are higher in women than in men. *J. Clin. Endocrinol. Metab.* **98**, E370–E378 (2013).
85. Naukkarinen, J. et al. Characterising metabolically healthy obesity in weight-discordant monozygotic twins. *Diabetologia* **57**, 167–176 (2014).
86. Lopez-Vicario, C. et al. Molecular interplay between Delta5/Delta6 desaturases and long-chain fatty acids in the pathogenesis of non-alcoholic steatohepatitis. *Gut* **63**, 344–355 (2014).
87. Frades, I. et al. Integrative genomic signatures of hepatocellular carcinoma derived from nonalcoholic Fatty liver disease. *PLoS ONE* **10**, e0124544 (2015).