

Topological analysis of metabolic networks integrating co-segregating transcriptomes and metabolomes in type 2 diabetic rat congenic series

Marc-Emmanuel Dumas, Céline Domange, Sophie Calderari, Andrea Rodríguez Martínez, Rafael Ayala, Steven P. Wilder, Nicolas Suárez-Zamorano, Stephan C. Collins, Robert H. Wallis, Quan Gu, et al.

▶ **To cite this version:**

Marc-Emmanuel Dumas, Céline Domange, Sophie Calderari, Andrea Rodríguez Martínez, Rafael Ayala, et al.. Topological analysis of metabolic networks integrating co-segregating transcriptomes and metabolomes in type 2 diabetic rat congenic series. Genome Medicine, 2016, 8 (1), pp.101. 10.1186/s13073-016-0352-6 . hal-01587798

HAL Id: hal-01587798

https://agroparistech.hal.science/hal-01587798

Submitted on 26 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dumas et al. Genome Medicine (2016) 8:101
DOI 10.1186/s13073-016-0352-6

Genome Medicine

RESEARCH Open Access

Topological analysis of metabolic networks integrating co-segregating transcriptomes and metabolomes in type 2 diabetic rat congenic series

Marc-Emmanuel Dumas[1,2,3*†], Céline Domange[2,4†], Sophie Calderari[5], Andrea Rodríguez Martínez[1], Rafael Ayala[1], Steven P. Wilder[6], Nicolas Suárez-Zamorano[5], Stephan C. Collins[6], Robert H. Wallis[6], Quan Gu[1,7], Yulan Wang[1,8], Christophe Hue[4], Georg W. Otto[6], Karène Argoud[6], Vincent Navratil[2], Steve C. Mitchell[3], John C. Lindon[1], Elaine Holmes[1], Jean-Baptiste Cazier[6,9], Jeremy K. Nicholson[1] and Dominique Gauguier[1,5,6*]

Abstract

Background: The genetic regulation of metabolic phenotypes (i.e., metabotypes) in type 2 diabetes mellitus occurs through complex organ-specific cellular mechanisms and networks contributing to impaired insulin secretion and insulin resistance. Genome-wide gene expression profiling systems can dissect the genetic contributions to metabolome and transcriptome regulations. The integrative analysis of multiple gene expression traits and metabolic phenotypes (i.e., metabotypes) together with their underlying genetic regulation remains a challenge. Here, we introduce a systems genetics approach based on the topological analysis of a combined molecular network made of genes and metabolites identified through expression and metabotype quantitative trait locus mapping (i.e., eQTL and mQTL) to prioritise biological characterisation of candidate genes and traits.

Methods: We used systematic metabotyping by ^{1}H NMR spectroscopy and genome-wide gene expression in white adipose tissue to map molecular phenotypes to genomic blocks associated with obesity and insulin secretion in a series of rat congenic strains derived from spontaneously diabetic Goto-Kakizaki (GK) and normoglycemic Brown-Norway (BN) rats. We implemented a network biology strategy approach to visualize the shortest paths between metabolites and genes significantly associated with each genomic block.

Results: Despite strong genomic similarities (95–99 %) among congenics, each strain exhibited specific patterns of gene expression and metabotypes, reflecting the metabolic consequences of series of linked genetic polymorphisms in the congenic intervals. We subsequently used the congenic panel to map quantitative trait loci underlying specific mQTLs and genome-wide eQTLs. Variation in key metabolites like glucose, succinate, lactate, or 3-hydroxybutyrate and second messenger precursors like inositol was associated with several independent genomic intervals, indicating functional redundancy in these regions. To navigate through the complexity of these association networks we mapped candidate genes and metabolites onto metabolic pathways and implemented a shortest path strategy to highlight potential mechanistic links between metabolites and transcripts at colocalized mQTLs and eQTLs. Minimizing the shortest path length drove prioritization of biological validations by gene silencing.

(Continued on next page)

* Correspondence: m.dumas@imperial.ac.uk; dominique.gauguier@crc.jussieu.fr
†Equal contributors
[1]Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Sir Alexander Fleming Building, Imperial College, London SW7 2AZ, UK
Full list of author information is available at the end of the article

BioMed Central

© 2016 The Author(s). **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

(Continued from previous page)

Conclusions: These results underline the importance of network-based integration of multilevel systems genetics datasets to improve understanding of the genetic architecture of metabotype and transcriptomic regulation and to characterize novel functional roles for genes determining tissue-specific metabolism.

Keywords: ^{1}H NMR, Metabolomics, Transcriptomics, Genome Mapping, mQTL, eQTL, Metabolic networks

Background

Type 2 diabetes mellitus is a prime example of a multifactorial disease, combining genetic risk factors and environmental influences, including the gut microbiome [1]. Complexity in diabetes etiology and pathogenesis relates to the existence of numerous risk genes, which often lack clear biological roles and have small effects on relevant disease traits [2–8], and the contribution of organ-specific cellular mechanisms to hyperglycemia and complications through impaired insulin secretion in pancreatic β cells and insulin resistance in central and peripheral tissues [9, 10]. Although genome-wide association studies (GWAS) have identified many diabetes risk variants [6, 7, 11–13], the underlying mechanisms remain elusive. Functional annotation of disease risk loci can progress with advances in high-density molecular phenotyping approaches, mainly transcriptomics [14–17] and metabonomics [11, 18], which inform about gene and metabolite networks for various tissues. Combining these high-throughput technologies generates complementary, and potentially convergent, multidimensional information on the function of the genome and individual genes.

Metabolic phenotyping (i.e., metabotyping) [19, 20] relates to quantitative physiological and biochemical changes to both phenotypic and genetic variation. Metabotypes provide a read-out for gene–environment interactions including microbiome influences [21, 22]. Metabotyping is particularly appropriate to define biomarkers associated with diabetes risk [9, 11, 23]. In previous studies, we and others have demonstrated genome mapping of ^{1}H NMR quantitative metabotypes in mouse and rat genetic crosses and defined causal relationships between segregating genetic polymorphisms and variations in metabolite abundance [18, 24–28]. Implementation of this strategy in humans remains limited to genetic associations in blood [29–33] and urine [9, 32, 34] but will most likely progress with genetic analysis of metabolic regulation in many organs from many individuals [35, 36]. Meanwhile, rodent models of complex disorders represent useful systems to investigate the direct molecular consequences of naturally occurring genetic polymorphisms and to understand the genetic architecture of metabolic regulation in biofluids [18, 26, 37] and organs [38–40].

The integrative analysis of expression quantitative trait loci (eQTL)-associated gene networks [15, 41] and metabolomic quantitative trait loci (mQTL)-associated metabolite and candidate gene networks remains a challenge [2, 42]. Here, we present an integrative systems genetics approach designed to identify mechanistic relationships between linked alleles in genomic blocks and metabolism using ^{1}H NMR-based metabotyping combined with genome-wide transcriptomic analyses. In a previous study we identified QTLs for adiposity (retroperitoneal fat pad weight) and insulin secretion on chromosome 1 [9, 43], which enabled the congenic breeding program to confirm the original associations [11, 39]. We selected a series of 12 congenic strains carrying contiguous regions of various lengths (1–177 Mb) of chromosome 1 of the spontaneously diabetic (type 2) Goto-Kakizaki (GK) rat transferred onto the genomic background of normoglycemic Brown-Norway (BN) rats for metabolic and gene expression profiling in adipose tissue. We then carried out a joint eQTL and mQTL analysis and mapped the associated genes and metabolites onto metabolic networks. Through topological analysis, we implemented a parsimonious approach (i.e., Occam's razor) by minimizing shortest paths across the resulting networks to identify pairs of mechanistically connected eQTL-associated genes and mQTL-associated metabolites for rapid validation in cell-based assays. Our approach indicates that distinct blocks of genetic polymorphisms differentially impact the adipose tissue metabolic network, thereby prioritizing candidates for gene silencing in adipocytes, exemplified by *Galm* and *Asns*.

Methods

Animals

A colony of GK/Ox rats bred locally and derived in 1995 from the GK/Par colony was used to produce the congenic strains. BN rats were obtained from a commercial supplier (Charles River Laboratories, Margate, UK). All congenics were derived from these strains using a genetic marker-assisted breeding strategy ("speed congenics") as previously described [11, 44] and maintained by brother–sister mating. They were specifically designed to contain GK alleles over genomic regions of various lengths (1–176 Mb) of rat chromosome 1, introgressed onto the genetic background of the BN strain (Table 1, Fig. 1). The targeted GK genomic intervals between markers D1Rat27 (90.3 Mb) and D1Got254 (264.37 Mb) cover several cardiometabolic disease (CMD) -relevant QTLs originally mapped in F2 (GKxBN) genetic crosses [9, 45, 46]. GK alleles on the X chromosome were lost early in the breeding

Table 1 Genomic details of the BN.GK congenic strains

Congenic name	Last marker BN allele	First marker GK allele	Last marker GK allele	First marker BN allele	Genomic length (Mb)
1cns	D1Got96 (88.6)	D1Rat27 (90.3)	D1Got353 (266.9)	- (Telomere)	176.6–179.3
1b	D1Rat77 (238.0)	D1Got237 (246.7)	D1Got353 (266.9)	- (Telomere)	20.2–29.9
1d	J576143 (225.8)	J337594 (227.9)	D1Got231 (231.6)	D1Got224 (233.0)	3.7–7.2
1f	J576143 (225.8)	J337594 (227.9)	D1Got353 (266.9)	- (Telomere)	39.0–42.1
1h	D1Got337 (193.5)	D1Got338 (197.0)	D1Rat84 (259.3)	D1Cebr4 (263.8)	62.34–70.3
1k	D1Got231 (231.6)	Glis3 (231.9)	XM_219778 (233.1)	Fxna (233.3)	1.2–1.7
1o	D1Got96 (90.3)	D1Got100 (92.1)	D1Wox86 (224.7)	J576143 (225.8)	132.6–137.2
1p	D1Got96 (88.6)	D1Rat27 (90.3)	D1Smu5 (189.5)	D1Got172 (191.7)	99.2–103.1
1q	D1Wox86 (224.7)	J576143 (225.8)	D1Rat75 (235.1)	D1Wox89 (236.5)	9.3–11.8
1t	J576143 (225.8)	J337594 (227.9)	D1Rat223 (228.8)	D1Rat76 (230.4)	0.9–4.6
1u	D1Wox7 (139.0)	D1Wox78 (143.8)	D1Got191 (175.4)	D1Got326 (176.6)	31.6–37.6
1v	D1Wox18 (94.6)	D1Got307 (102.5)	D1Got108 (127.4)	D1Got318 (130.1)	24.9–35.5

Name and genomic position (Mb) of the genetic markers flanking the GK congenic intervals are given, along with the minimum and maximum genomic length (Mb) of the introgressed GK genomic segment. Genomic positions were taken from the rat reference genome assembly (RGSC3.4, Ensembl release 69)

program by two consecutive breedings of male backcross animals to BN females. All animals used in this study were systematically genotyped with markers chosen to accurately monitor retention of GK alleles across the congenic intervals and absence of GK allele contaminants from the genetic background [47, 48]. All strains were co-housed in order to avoid cage-specific microbiome selection [14, 49].

Rats were allowed free access to tap water and standard laboratory chow pellets (B&K Universal Ltd, Grimston, Aldbrough, Hull, UK) and were maintained on a 12-h light–dark cycle. All rats were identified using a microchip (identity chip, Animal Care Ltd, York, UK) linked to a database specifically designed to administer the project (husbandry, phenotype scheduling, and data

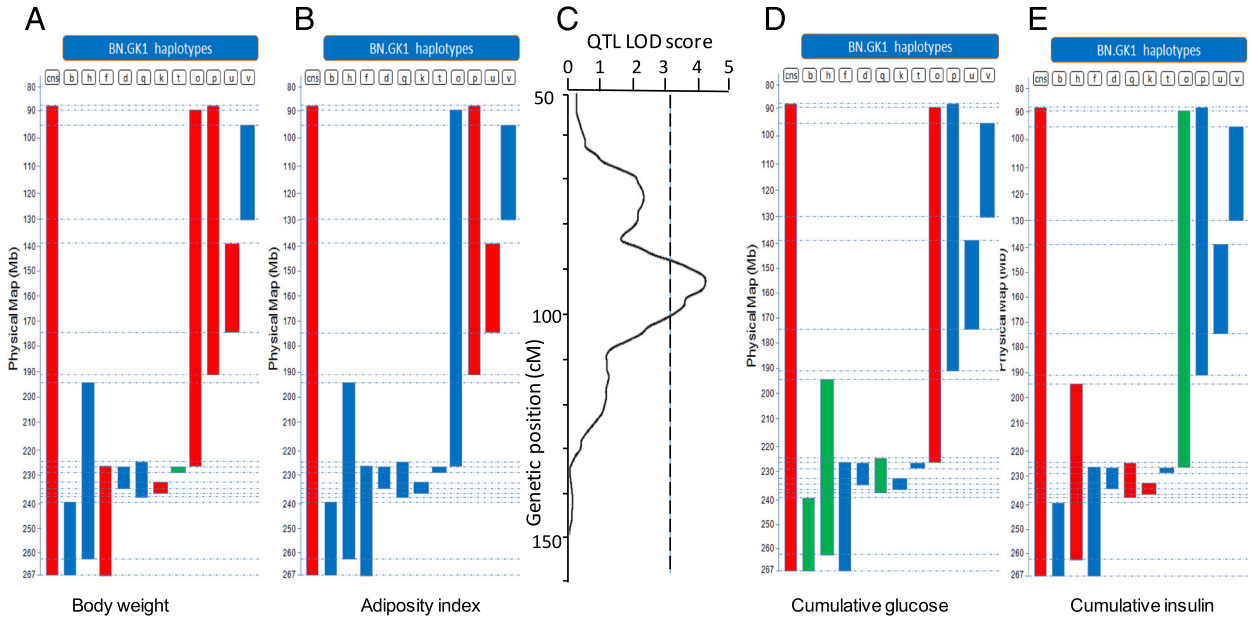


Fig. 1 Phenotypic features of the congenic strains. Association for body weight (**a**), adiposity index (**b**), QTL for adiposity index (**c**), association for cumulative plasma glucose (**d**), and insulin (**e**) during the intravenous glucose tolerance and insulin secretion tests were measured in male rats. *Solid bars* represent the GK genomic segments of chromosome 1 of each congenic strain introgressed onto the genetic background of the BN strain. The *y-axis* shows genomic length (Mb) and boundaries of the genomic region of GK origin. The location of the adipose tissue QTL mapped to chromosome 1 in the GK × BN F2 cross [9] is reported with significance threshold shown with a *dashed line* (**c**). Details of GK chromosomal regions introgressed in each congenic strain are given in Table 1. Significantly ($P < 0.05$) increased and decreased values of the phenotypes between congenic strains and the BN control are indicated in *red* and *green*, respectively. Phenotype data are available in Additional file 1: Table S1

storage) and store genetic information and phentoypic data [48, 50].

Phenotype analysis

Three-month-old male congenic rats and BN controls were used for all experiments. Intravenous glucose tolerance and insulin secretion tests (IVGTTs) were performed following procedures strictly identical to those consistently applied in both F2 (GK × BN) hybrids [9, 20], which we used to map diabetes QTLs in the GK, and BN.GK congenic strains derived for several GK QTLs [11, 22, 51–54]. Briefly, rats in the post-absorptive state at the end of the post-prandial glycemic response (4.5 to 5 h fasting from 9–9:30 am when food was removed until 2 pm when the IVGTT procedures were initiated) were anesthetized by injection of 95 mg/kg body weight intraperitoneal ketamine hydrochloride (Ketalar, Parke-Davies, UK). Rats were injected with a solution of 0.8 g glucose/kg body weight via the saphenous vein. Blood samples were collected into heparinized tubes before glucose injection and 5, 10, 15, 20, and 30 min afterward. Plasma was separated by centrifugation prior to glucose assays using a diagnostic kit (ABX, Shefford, UK) on a Cobas Mira Plus automatic analyzer (ABX, Shefford, UK) and assay of immunoreactive insulin (IRI) using an ELISA kit (Mercodia, Uppsala, Sweden). Cumulative values of plasma glucose and plasma insulin during the IVGTTs were calculated to evaluate overall glucose tolerance and insulin secretion capacity in response to glucose, respectively. At 6 months, rats were killed by terminal anesthesia following an overnight fast (16–18 h). Retroperitoneal fat pads (RFPs) were collected, weighed, snap frozen in liquid nitrogen, and stored at −80 °C until preparation of tissue extracts and RNA for analysis of the metabolome and the transcriptome, respectively. The adiposity index was calculated as the ratio between RFP weight and body weight.

Metabotyping of white adipose tissue by ^{1}H NMR spectroscopy

Tissues samples (30–50 mg) were weighed into 2-mL Eppendorf tubes and were each homogenized in 1.5 ml 50 % methanol using TissueLyser (5 min at 25 Hz; QIAGEN, Germany). The mixtures were each transferred into 3-mL glass tubes and 0.7 mL chloroform was added into each sample. The mixtures were vortexed for 1 min followed by centrifugation at ~3500 g for 25 min at 10 °C. The aqueous phase was decanted and the methanol was removed under a fume cupboard before freeze drying. The lipid phase was pipetted out and chloroform was removed under a fume cupboard. Dried extracts were reconstituted using 500 μL of 0.1 M phosphate buffer solution (10 % $^{2}H_2O/H_2O$ v/v, with 0.05 % sodium 3-trimethylsilyl-(2,2,3,3-^{2}H4)-1-propionate for chemical shift reference at δ0.0) in 5-mm tubes for NMR acquisition. Standard ^{1}H NMR spectra were measured on a Bruker spectrometer (Rheinstetten, Germany) operating at 600.22 MHz ^{1}H frequency, as described previously [18, 55]. The ^{1}H NMR spectra were imported into Matlab and phase- and baseline-corrected at high resolution. The region δ5.0–4.5 was removed to eliminate baseline effects of imperfect water signal presaturation. Each spectrum was normalized to a constant intensity sum and each variable was mean centered. Analyses were carried out using R and Matlab.

Orthogonal partial least squares discriminant analysis

The method allows enhanced focus on strain and diet intervention while minimizing other biological/analytical variation. Sample classes were modeled using the orthogonal partial least squares (O-PLS) algorithm. This algorithm derives from the partial least squares (PLS) regression method. In linkage analysis version, the model explains the maximum separation between genotypes Y (*coded as 0, 1, 2 for GK allele numbers*) using the NMR data X. Further details on standard O-PLS implementation in metabonomics have been given previously [56, 57]. The model coefficients locate the NMR signals significantly associated with genotypic variation in a specific genomic region Y.

RNA preparation and Illumina bead array hybridization

Total RNA was individually isolated from 100 mg of RFP (four biological replicates per strain) using the RNeasy® 96 Universal Tissue kit (Qiagen, Crawley, UK): frozen tissue samples were transferred into cooled RNeasy® 96 Universal Tissue plates and homogenized in QIAzol Lysis Reagent using a Qiagen Tissue Lyser. Following phase separation after addition of chloroform, total RNA was purified with RNeasy columns using a spin technology according to the manufacturer's guidelines and eluted in RNase-free water. RNA concentrations were determined using a NanoDrop spectrophotometer and RNA quality, purity, and integrity were assessed using an Agilent 2100 Bioanalyser (Agilent Technologies, Waldbronn, Germany).

Samples were independently used to hybridize Gene Expression Sentrix® BeadChip RatRef-12 v1 arrays (Illumina Inc., San Diego, CA, USA) containing 22,523 oligonucleotide probes (replicated 30 times on average). They allowed interrogation of transcript levels for 21,910 genes (6274 RefSeq NM transcripts, 15,983 Refseq XM transcripts, 12 Refseq XR transcripts, 250 Unigene clusters). Double-stranded cDNA and purified biotin-labeled cRNA were synthesized from 300 ng high quality total RNA using the Illumina® TotalPrep RNA Amplification Kit (Ambion Inc., Austin, TX, USA). cRNA concentrations were determined using a NanoDrop spectrophotometer whilst cRNA quality and integrity were assessed using an Agilent 2100 Bioanalyser (Agilent Technologies, Waldbronn, Germany).

Hybridizations onto the arrays were carried out using 750 ng of each (132) biotinylated cRNA in a 58 °C hybridization oven for 18 h. Following washing and staining with Streptavidin-Cy3, the BeadChip Arrays were scanned on the Illumia® BeadArray Reader (Illumina Inc., San Diego, CA, USA). Resulting data were then preliminarily analyzed using the Illumina® BeadStudio Application software before undergoing comprehensive statistical analysis. Particular attention was given to the following quality control parameters: $0 \leq G\ sat \leq 1$; Green 95 percentile (GP95) for consistency between arrays (around 2000); GP5 background level in the range of the low 100 s or below.

Microarray experiments were compliant with MIAME (Minimum Information About a Microarray Experiment) and both protocol details and raw data have been deposited in ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-1048.

Network representation of genome–metabolome associations

In order to explore genome–metabolome associations, a functional association network was derived from metabotype and genotype correlation coefficients using the bipartite graph Rgraphsviz package from R to represent the O-PLS correlation matrix derived from the linkage analysis between NMR variables and genotypes. A cutoff was then applied to the *P* value of the Pearson's correlation coefficient adjusted using Benjamini and Hochberg's multiple testing correction ($P_{BHadj} < 0.05$); significant correlations were set to 1 and non-significant correlations were set to 0, defining the adjacency matrix for the graph. Hence, G = (N, E) specifies a graph G with N denoting the two node sets (two types of nodes: genomic regions and metabolites) and E the edge set (link between nodes, here a correlation between metabotypes and genotypes above the cutoff, i.e., $P_{BHadj} < 0.05$).

Integration of eQTL-responsive genes and mQTL-responsive metabolites on KEGG metabolic networks to identify functional candidates

Differentially regulated genes and metabolites that could be associated with GK haplotypes were mapped to KEGG pathways. Metabolic pathways were imported in R using KEGGgraph [55, 58]. An in-house python script was written to generate an adjacency table. The resulting region-specific networks were mathematically formalized as a directed multilabeled graph Gm,n = (Vm,E) composed of two types of nodes Vm, where m = ("metabolic reactions", "metabolites"), and functional equally-weighted edges corresponding to network connectivity (i.e., metabolic reactions) between enzymes and metabolites. To identify pairs of eQTL-associated genes coding for enzymes that are metabolically connected with mQTL-associated metabolites, we computed shortest paths from the eQTL genes mapped on the KEGG network to the target mQTL metabolites using the igraph R package. Defining shortest path lengths (spl) corresponds to counting the minimal number of additional reactions required to connect a given gene and a given metabolite across the metabolic network. For example, *Galm* and β-D-glucose have a spl of 0 as β -D-glucose is the product of the reaction catalyzed by *Galm* (rn:R01602). The gene *Asns* is annotated as AsnsA and AsnsB since it involves two catalytic sites for two different reactions (rn:R00256 for AsnsB and rn:R00578 for AsnsA).

shRNA-based inhibition of *Galm* and *Asns* expression in vitro in 3T3-L1 cells

3T3-L1 fibroblasts (ATCC, Molsheim, France) were cultured in 10 % calf serum (PAA, Velizy, France) containing DMEM high glucose (Life Technologies, Saint Aubin, France). Cells were plated at 10^5 cells/well density until confluence and differentiated into adipocytes in 10 % fetal bovine serum (FBS; Life Technologies, Saint Aubin, France) containing DMEM high glucose, IBMX, dexamethasone (Sigma-Aldrich, Saint-Quentin, France), and insulin. Differentiated 3T3-L1 cells were maintained in 5 % FBS and DMEM high glucose.

We used pGFP-V-RS-shRNA plasmids (Origene, Rockville, MD, USA) containing short hairpin RNA (shRNA) sequences specifically designed to target *Asns* or *Galm* and a puromycin resistance gene cloned between integrative long terminal repeat sequences. The Platinum-Ecotropic Retroviral Packaging Cell Line (Cell Biolabs, San Diego, CA, USA) producing host range recombinant γ-retroviruses was used for shRNA-containing viral production. Platinum E cells were maintained in DMEM supplemented with glucose and 10 % FBS, puromycin, and blasticidin (Sigma-Aldrich, Saint-Quentin, France). Transfection was performed after adapting culture medium for 3T3-L1 cells (DMEM high glucose and 10 % calf serum without antibiotics). Fugene 6 HD® (Roche, Boulogne, France) was used according to the manufacturer's recommendation in 3T3-L1 medium. Plasmid transfection efficiency was determined by GFP fluorescence. Supernatants were collected 48 h after transfection to transduce 3T3-L1 in the presence of polybrene (Sigma-Aldrich, Saint-Quentin, France). Puromycin selection started 24 h post-transduction.

Differentiation and glucose transport in shRNA-transfected 3T3-L1 adipocytes

For differentiation analysis, cells were first incubated in 10 % formaldehyde (Sigma-Aldrich, Saint-Quentin, France), washed with 60 % isopropanol, and dried. A solution Oil Red O (Sigma-Aldrich, Saint-Quentin, France) was added and dishes were washed with distilled water. Quantification of coloration was performed by spectrophotometry at 490 nm with a plate reader (Perkin Elmer, Villebon, France). A separate batch of cells was used for glucose transport

analysis, which was determined by incubation with a solution containing 0.5 μCi tritium labeled 2-deoxy-D-glucose. Briefly, adipocytes were cultured in DMEM high glucose without FBS for 4 h and washed with a buffer containing $CaCl_2$, $MgCl_2$, fatty acid free BSA (PAA, Velizy, France) in PBS (Life Technologies, Saint Aubin, France), followed by 20 min of insulin stimulation (100 nM). After 10 min of incubation with labeled 2-deoxy-D-glucose, cells were washed in ice-cold PBS and collected in a solution of NaOH for radioactivity recording and protein content quantification.

Results

Pathophysiological features of the congenic series

To attach specific patterns of diabetes intermediate phenotypes to each BN.GK congenic strain used in the study (Table 1) glucose tolerance, in vivo insulin secretion, body weight, and adiposity index were determined for each animal at 3 months. Rats from different strains were co-housed in the same cage to avoid cage-specific microbiome selection [14, 17, 59]. As previously observed [11], the strain carrying the largest GK haplotype (179.3 Mb in BN.GK1cns) exhibited increased body weight (270.0 ± 5.7 versus 243.4 ± 3.5) and adiposity (0.416 ± 0.028 versus 0.376 ± 0.031), glucose intolerance (CumG, 5504 ± 135 versus 5104 ± 83) and enhanced glucose-induced insulin secretion (CumI, 70.21 ± 11.50 versus 43.45 ± 3.23). To dissect the genetic basis of body weight and glucose homeostasis on chromosome 1, we validated these phenotypes in congenic substrains containing smaller GK genomic blocks introgressed into the BN normoglycemic genome (Fig. 1). Validation was achieved in nine congenic strains, which exhibited significant changes in body weight (Fig. 1a; BN.GK1f, 1k, 1o, 1p, 1t, 1u) and adiposity (Fig. 1b; BN.GK1p, 1u) when compared to BN (Additional file 1: Table S1). Rats of congenic strains BN.GK1cns, 1p and 1u showed consistent increases in both body weight and adiposity index, suggesting coordinated regulation of these phenotypes by common genetic polymorphisms in the shared GK genomic region (143.8–175.4 Mb) of rat chromosome 1 corresponding to a region of significant linkage to adiposity in the GK × BN F2 cross (Fig. 1c). Impaired glucose tolerance in BN.GK1o (Fig. 1d) may be the consequence of impaired insulin secretion (Fig. 1e), whereas improved glucose tolerance in BN.GK1h and 1q may be caused by enhanced insulin response to glucose (Fig. 1d, e). Rats of strains BN.GK1d and 1v did not show any significant change in any of these parameters when compared to BN controls (Fig. 1; Additional file 1: Table S1).

These data underline the strong phenotypic heterogeneity in these congenic strains and the complexity of the underlying genetic regulation, even though they share 95–99 % genomic homology with the BN control. Strain-specific phenotypic features can be attached to GK genomic blocks contained in each congenic line and therefore characterize the systems-wide effects of linked GK genetic polymorphisms in each contiguous region of chromosome 1.

^{1}H NMR-based metabotyping of adipose tissue in congenic strains

To complement pathophysiological phenotypes in the congenics with molecular phenotypes, and to investigate possible relationships between physiological and metabolic variables, we carried out ^{1}H NMR metabotyping of aqueous extracts of white adipose tissue from rats of the 12 BN.GK congenic strains and the BN control (Additional file 2: Figure S1). A total of 34 metabolites were detected, 31 of which could be assigned using published data [9, 19, 46] and in-house databases (Table 2). We then built orthogonal partial least squares discriminant analysis (O-PLS-DA) models to compare metabolite patterns in each congenic strain to the BN normoglycemic control (Additional file 2: Figure S2). The robustness of each O-PLS model was tested by sevenfold cross-validation and resampling under the null hypothesis as described previously [21, 48]. The range of distribution of the Q^2_Y values (0.35–0.88) for each of the congenic models suggests the presence of a clear discrimination of congenic lines (Additional file 1: Table S2). We identified 19 metabolites exhibiting variations in abundance between at least one congenic strain and the BN control (Additional file 2: Figure S3; Table 2).

Definition of strain-specific metabotypes in the congenic series

To simultaneously visualize all metabotypes characterizing each congenic strain, we constructed a binary association map summarizing strain–metabolite associations (Fig. 2a; Additional file 2: Figures S2 and S3). The congenic strains BN.GK1b and 1u exhibited differential regulation of many metabolites when compared to BN. Only a few metabolites showed a strain-specific pattern of regulation, such as succinate in GK1b and taurine and glycerol in BN.GK1u. In contrast, several metabolites were differentially regulated in many congenic strains (e.g., *N*-acetylglutamine, L-glutamate, glycerophosphocholine, *scyllo*-inositol, inosine; Fig. 2a), suggesting consistent effects of common genetic polymorphisms in the shared GK genomic blocks in these strains. For example, inosine, which is consistently more abundant in BN.GK1d, 1h and 1q, may be controlled by a gene in the GK genomic block of BN.GK1d (225.8–233.0 Mb), whereas *scyllo*-inositol (δ3.36, s) was downregulated in three congenic strains (BN.GK1b, 1q, 1u) that carry distinct GK genomic blocks and is therefore controlled by different genes. Discordant trends of metabolic regulation among congenics were found for *N*-acetyl

Table 2 Summary of metabolites detected in ^{1}H NMR spectra from adipose tissue extracts of congenic rats and BN controls

Metabolite	^{1}H chemical shift (δ, ppm) and multiplicity
Formate	8.46 (s)
Inosine[a]	8.34 (s), 8.23 (s), 6.1 (d), 4.44 (t)
Non-assigned	8.27 (s)
Non-assigned[a]	7.68 (s)
Inosine-diphosphate	6.16 (s)
Allantoin[a]	5.4 (s)
Lipids (C[H_2]-CH_2-CO)	5.31 (m)
Glucose[a]	5.23 (m), 3.84
Glycerophosphocholine[a]	4.32 (m), 3.23 (s)
Lactate[a]	4.11 (q), 1.33 (d)
Myo-inositol	4.06 (t), 3.61 (dd), 3.52 (dd)
Creatine	3.93 (s), 3.04 (s)
β-D-glucose[a]	3.91 (dd), 3.73 (d), 3.48 (m)
Glycerol[a]	3.78 (m), 3.66 (dd), 3.64 (dd), 3.56 (dd)
3-Methyl-histidine[a]	3.7 (s)
Taurine 3.42 (t)[a]	3.42 (t), 3.25 (t)
Scyllo-inositol[a]	3.36 (s)
Choline[a]	3.21 (s)
Non-assigned	3.13 (s), 4.04 (d)
Lipids (C = C-CH_2-C = C)	2.76 (m)
Glutamine[a]	2.46 (m), 2.14 (m)
Succinate[a]	2.41 (s)
Glutamate[a]	2.36 (dt), 2.04 (m)
3-Hydroxybutyrate[a]	2.31 (m), 1.2 (d)
Lipids (CH_2-CO)	2.26 (m)
N-acetylglutamine[a]	2.02 (s)
Acetate[a]	1.92 (s)
Lipids (C[H_2]-CH2-CO)	1.58 (m)
Alanine	1.48 (d)
Lipids ((–CH_2-)n)	1.27 (m)
Valine	1.04 (d), 0.99 (d)
Isoleucine[a]	1.01 (d)
Leucine	0.96 (dd)
Lipids (CH_3)	0.89 (m)

[a]Compounds that were statistically significant ($P < 0.05$) between at least a congenic strain and BN. Patterns of significant differential regulation of these metabolites in the relevant congenic strains and contributing chromosomal regions are shown in Fig. 2

S singlet, *d* doublet, *t* triplet, *m* multiplet, *dd* doublet of doublet, *dt* doublet of triplet

glutamine and L-glutamate, which were more abundant in adipose tissue of BN.GK1b, 1f, 1k, 1cns, 1p, 1q, 1t, and 1u than in the BN control and showed an opposite trend of regulation in BN.GK1q. These results illustrate the existence of genetically determined metabotypes and suggest functional redundancy in the regulation of individual metabolites by distinct genes.

Identification of genetically determined metabotypes by mQTL mapping

To connect genomic information with metabolic endpoints, we took advantage of the genetic structure of the congenic panel, which is characterized by overlapping and unique GK genomic blocks across a large region of rat chromosome 1, to perform QTL analysis. The resulting association networks defined 15 independent contiguous genomic regions of GK origin across the rat chromosome 1 (Fig. 2b; Additional file 2: Figures S3 and S4). Genomic regions R15 and R16, which cover 18 Mb and 7 Mb, respectively, at the telomeric end of chromosome 1, were often associated with a similar set of metabolites (choline, L-glutamine, succinate, L-glutamate, acetate, L-alanine, *myo*-inositol; Fig. 2b). Several metabolites (L-glutamine, taurine, *scyllo*-inositol, D-glucose, L-lactate, inosine, formate) were associated with several genomic regions, suggesting the involvement of multiple independent GK variants, whereas glycerophosphocholine was specifically linked to region R5, indicating a specific effect of GK variants in this region.

Identification of genetically determined expression traits by eQTL mapping

To test the existence of functional relationships between metabolic changes mapped to congenic regions and gene expression, we next carried out genome-wide transcriptome analyses of white adipose tissue from rats of the entire congenic panel and from BN and GK controls. As previously applied to metabotypes in congenic series, we performed eQTL mapping to anchor specific gene expression patterns to each of the 15 regions of rat chromosome 1 defined by GK genomic blocks introgressed in the congenics. We found evidence of genetic linkage (LOD >5) between quantitative variation of 378 transcripts and regions of chromosome 1 (Fig. 3; Additional file 1: Table S3). Over 25 % of eQTLs correspond to genes localized within the GK genomic regions in congenics, which may underlie *cis*-mediated regulatory mechanisms. The remaining 287 eQTLs were related to transcripts localized outside the congenic region, which unambiguously underlie distant regulatory mechanisms of transcription (i.e., *trans*-acting eQTLs, as illustrated in Fig. 3).

These transcriptome data demonstrate the effect of GK variants in the congenic intervals on genome-wide gene expression and identify potential positional candidate genes that may be functionally related to changes in metabolite abundance and disease phenotypes.

Mapping of eQTL genes and mQTL metabolites onto metabolic networks

The identification of candidate genes remains difficult because the size of the genomic blocks prevents us from applying a classic GWAS "one-SNP-one-gene-at-a-time" approach. A systems biology approach is thus required to take into account the fact that each genomic block influences multiple genes and metabotypes in a coordinated fashion, which can prove helpful to identify candidate genes relevant to diabetes and obesity. We have therefore developed a systems genetics approach stitching together genetically determined gene expression and metabolic profiles in a tissue-specific network to rank and prioritize the validation of candidate genes for adiposity and glucose homeostasis using cell-based assays.

We built an association network summarizing all significant associations between genomic blocks and metabolites derived from mQTL mapping (BH adjusted $P < 0.05$) and

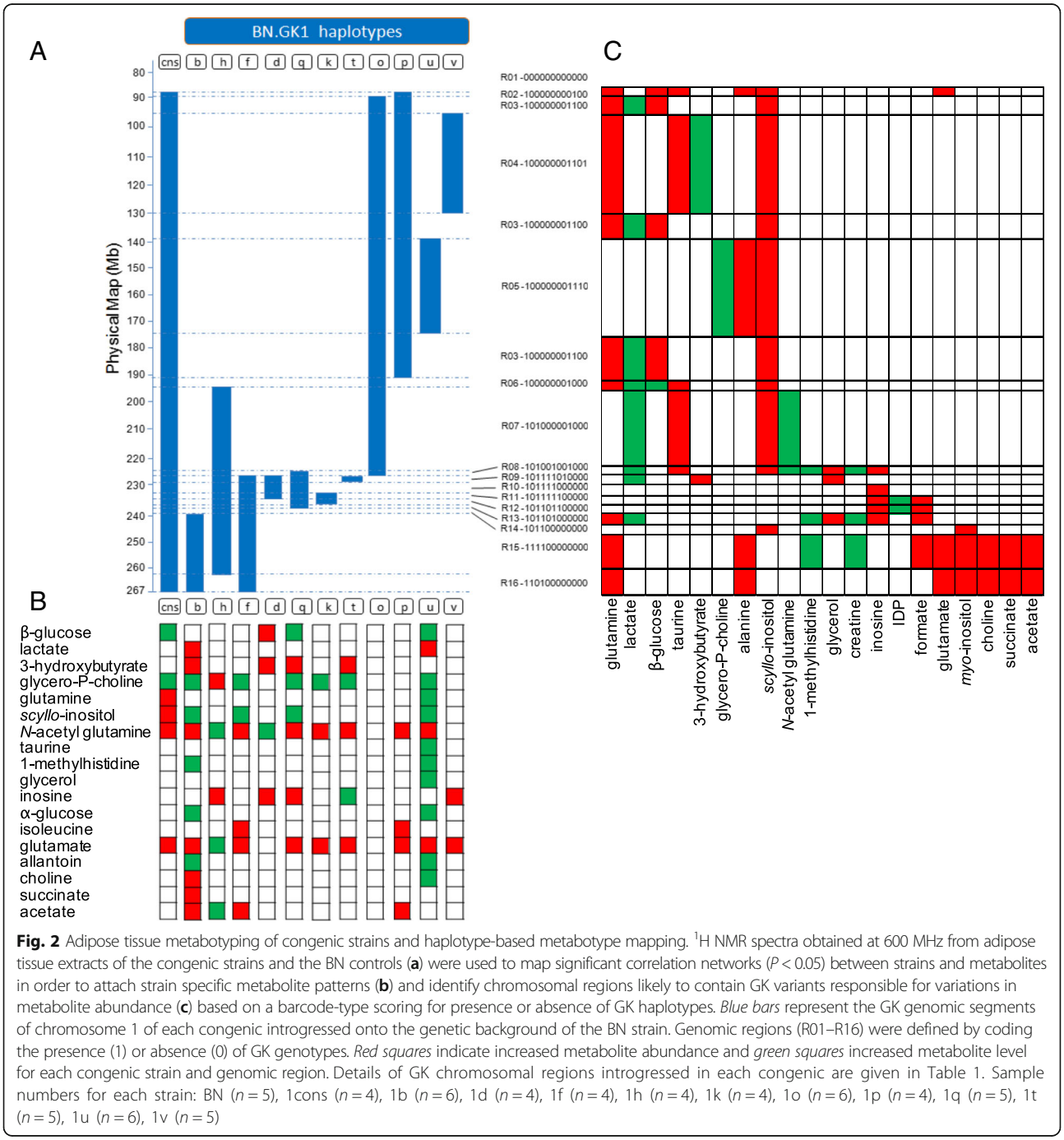


Fig. 2 Adipose tissue metabotyping of congenic strains and haplotype-based metabotype mapping. ^{1}H NMR spectra obtained at 600 MHz from adipose tissue extracts of the congenic strains and the BN controls (**a**) were used to map significant correlation networks ($P < 0.05$) between strains and metabolites in order to attach strain specific metabolite patterns (**b**) and identify chromosomal regions likely to contain GK variants responsible for variations in metabolite abundance (**c**) based on a barcode-type scoring for presence or absence of GK haplotypes. *Blue bars* represent the GK genomic segments of chromosome 1 of each congenic introgressed onto the genetic background of the BN strain. Genomic regions (R01–R16) were defined by coding the presence (1) or absence (0) of GK genotypes. *Red squares* indicate increased metabolite abundance and *green squares* increased metabolite level for each congenic strain and genomic region. Details of GK chromosomal regions introgressed in each congenic are given in Table 1. Sample numbers for each strain: BN ($n = 5$), 1cons ($n = 4$), 1b ($n = 6$), 1d ($n = 4$), 1f ($n = 4$), 1h ($n = 4$), 1k ($n = 4$), 1o ($n = 6$), 1p ($n = 4$), 1q ($n = 5$), 1t ($n = 5$), 1u ($n = 6$), 1v ($n = 5$)

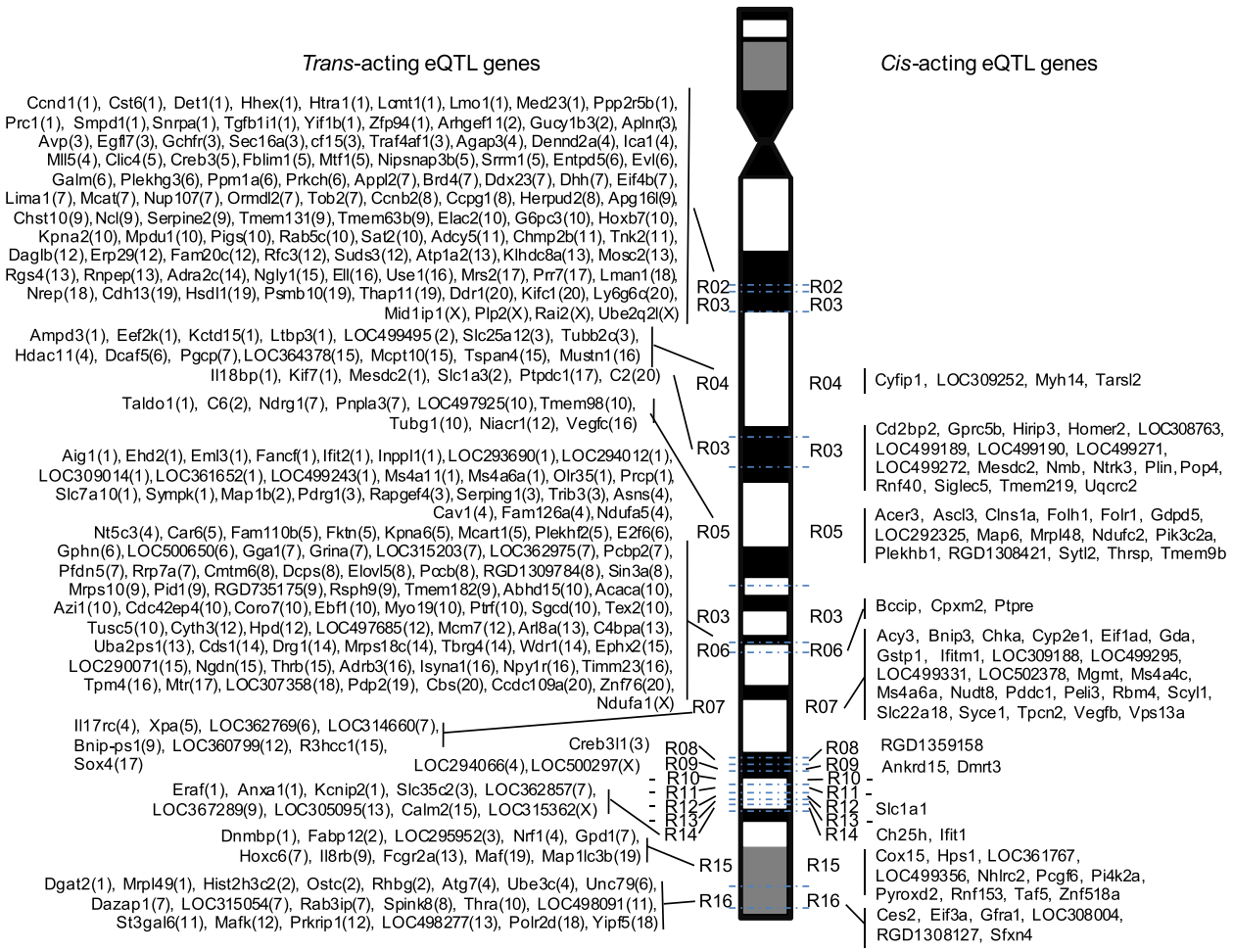


Fig. 3 Genetic mapping of genome-wide gene expression in the adipose tissue of BN.GK congenic rats derived for chromosome 1 loci. Quantitative trait locus (QTL) mapping was applied to define regions of chromosome 1 showing evidence of statistically significant (LOD >5) linkage with changes in the transcription of genes localized in genomic regions defined by GK haplotypes (putative *cis*-acting eQTL effects) or outside the regions (*trans*-acting eQTL effects). Details of the congenic strains and congenic-defined regions are given in Table 1 and associated transcripts in Additional file 3: Table S4. The localization of genes regulated *in trans* is indicated in *parentheses*

genes derived from eQTL mapping (Fig. 4a). To systematically search for the mechanistic links between variations in metabolite abundance and gene expression, we mapped significant eQTL-responsive genes (39 genes encoding 73 reactions) and mQTL-responsive metabolites (20) onto metabolic pathways (Fig. 4b). Objective biological relationships between significant changes in gene expression and metabolite abundance were inferred following mapping of genes and metabolites to KEGG pathways and computational analysis of shortest paths between eQTL-associated genes and mQTL-associated metabolites across the metabolic pathways. Applying a shortest path length threshold of 1 (spl ≤1), we identified pairs of candidate genes and metabolites that are directly mechanistically related (i.e., the gene codes for an enzyme which directly catalyzes the reaction involving its paired metabolite; Fig. 4c; Additional file 3: Table S4). For instance, our network topology analysis showed direct connection between haplotype-associated gene *G6pc3*, encoding glucose-6-phosphatase, and glucose in chromosomal region 2. Glucose was also directly connected to *Galm*, encoding galactose mutarotase (aldose 1-epimerase). Likewise, we identified direct reactions between *Asns*, encoding asparagine synthetase (glutamine-hydrolyzing), and glutamine. To test whether the distance between the 73 reactions and 20 metabolites obtained from eQTL and mQTL mapping was significantly shorter, we permutation-tested our network under the null hypothesis: using random lists of 73 reactions and 20 metabolites for each of the 10,000 permutations and derived shortest paths. The average shortest path length of the original eQTL–mQTL directed network was 6.54 reactions, which was significantly shorter than the average shortest path length (obtained for 10,000 undirected random networks; Fig. 4d).

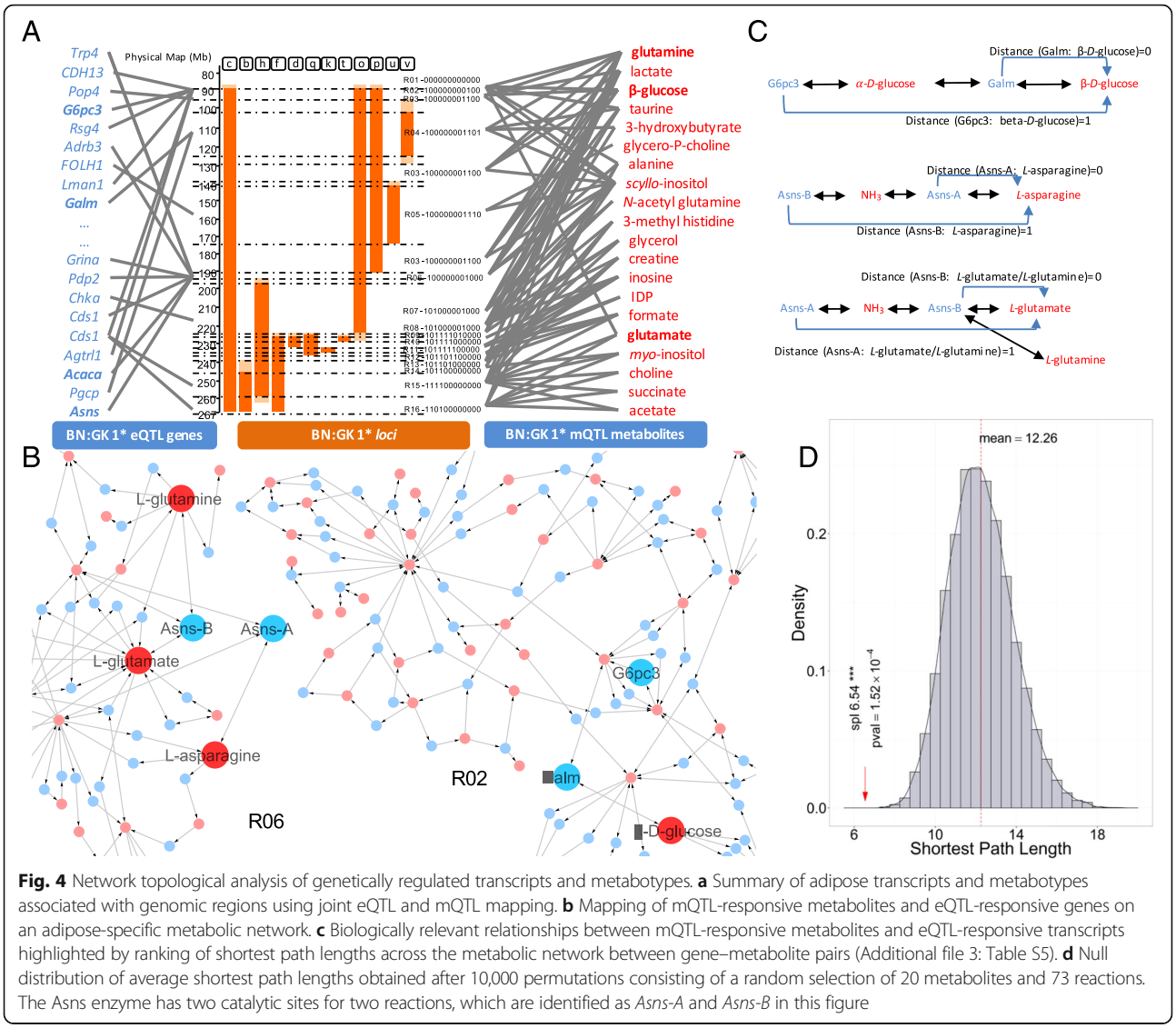


Fig. 4 Network topological analysis of genetically regulated transcripts and metabotypes. **a** Summary of adipose transcripts and metabotypes associated with genomic regions using joint eQTL and mQTL mapping. **b** Mapping of mQTL-responsive metabolites and eQTL-responsive genes on an adipose-specific metabolic network. **c** Biologically relevant relationships between mQTL-responsive metabolites and eQTL-responsive transcripts highlighted by ranking of shortest path lengths across the metabolic network between gene–metabolite pairs (Additional file 3: Table S5). **d** Null distribution of average shortest path lengths obtained after 10,000 permutations consisting of a random selection of 20 metabolites and 73 reactions. The Asns enzyme has two catalytic sites for two reactions, which are identified as *Asns-A* and *Asns-B* in this figure

Through mapping eQTL-responsive genes and mQTL-responsive metabolites onto the adipose tissue metabolic network and analyzing its topology, we identified coordinated regulation of gene transcription and metabolite abundance in the adipose tissue which may account for differences in the pathophysiological phenotypes observed in these congenic strains. We therefore sought to validate by cell-based assays the relevance of the *Asns* and *Galm* genes predicted by our systems genetics approach.

Experimental assessment of *Asns* and *Galm* function in vitro

To exemplify the relevance of our network integrating eQTL-responsive genes and mQTL-responsive metabolites in adipose tissue to prioritize biological validation, we selected two genes highlighted by our topological analysis: *Asns* and *Galm*. Since we had previously identified physiological QTLs mapping to chromosome 1 in the GK rat for adiposity and insulin secretion [9, 11, 14], we developed a system of shRNA-based expression knock-down for these genes in 3T3-L1 cells, which are often used to test cellular phenotypes related to diabetes (insulin-stimulated glucose uptake) and obesity (lipid accumulation through oil red-O straining) [24, 25, 27, 48]. Treatment with shRNA targeting *Asns* and *Galm* resulted in a 50 % reduction in abundance of the respective transcripts (Fig. 5a, b). The aspecific shRNA had no effect on lipid accumulation, a proxy measure of adipocyte differentiation, assessed by oil Red-O staining (Fig. 5c). Lipid accumulation was significantly hampered by *Asns* knockdown, whereas it was unchanged for *Galm* knockdown (Fig. 5c). Insulin induced a significant stimulation of glucose transport in both control cells and cells

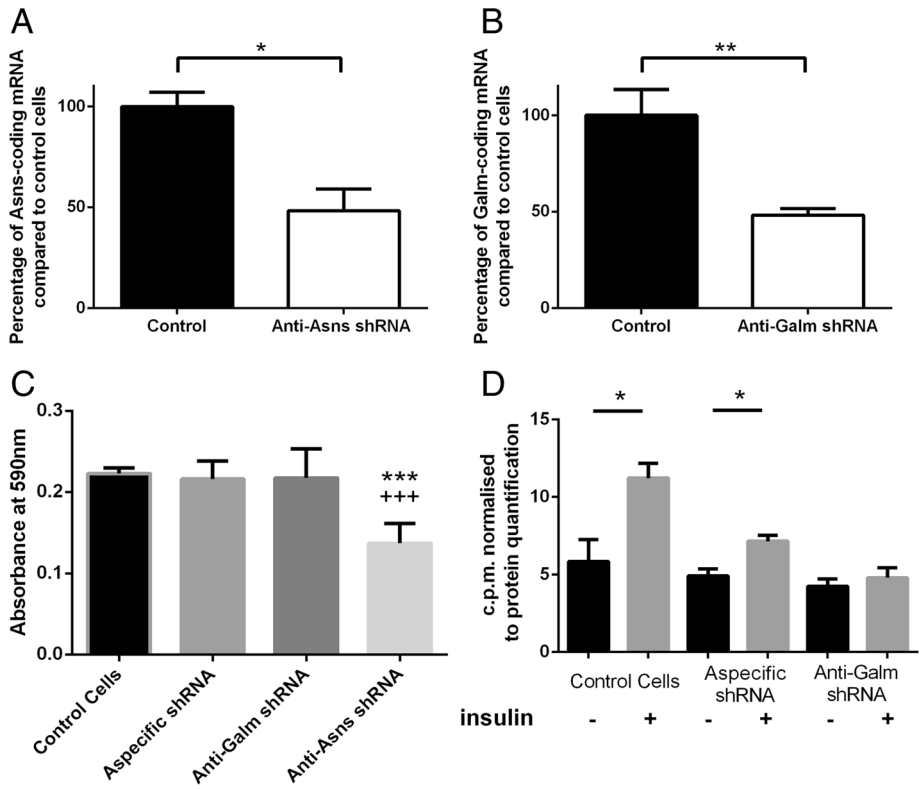


Fig. 5 Functional assessment of in vitro shRNA-mediated inhibition of *Galm* and *Asns* expression in 3T3-L1 adipocytes. **a** *Asns* mRNA expression levels in anti-*Asns* shRNA-treated cells expressed as a percentage of control 3T3-L1 cells. **b** *Galm* mRNA expression levels in anti-*Galm* shRNA-treated cells expressed as a percentage of control 3T3-L1 cells. **c** Intracellular lipid content of differentiated 3T3-L1 cells was measured by absorbance at 590 nm after Oil Red-O staining. **d** Glucose uptake was evaluated by measurement of radiolabeled 2-deoxyglucose present in 3T3-L1 cells following insulin stimulation and normalized to protein level. Shown data represent means ± standard error of the mean. Mann–Whitney tests were performed: $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$ significantly different to control; and $^{+++}P < 0.001$ significantly different to *Galm*-deficient cells

transfected with aspecific shRNA (Fig. 5d). This effect was abolished in *Galm*-deficient cells, thus demonstrating the involvement of *Galm* in the regulation of insulin signaling. The validation of the role of genes directly impacting lipid accumulation and glucose uptake in adipocytes identified in adipose tissue in a context of obesity and diabetes illustrates the efficiency of our approach based on the topology of the metabolic network.

Discussion

Here we report the joint genome mapping of transcripts and metabolites in adipose tissue extracts using a novel network-based integration of several -omics dimensions (genome, transcriptome, and metabolome) to prioritize mechanistic investigations in adipocyte biology in a context of metabolic syndrome.

Our approach can be summarized as follows. First, we used a rat congenic panel to study glucose homeostasis and adiposity whilst limiting epistatic interactions. Second, we profiled the adipose tissue metabolome and transcriptome to identify loci regulating metabolism through differential expression and complex eQTL and mQTL patterns. Third, to tackle the complexity of these genetically determined co-regulation patterns between transcripts and metabolites, we mapped underlying eQTL-responsive genes and mQTL-responsive metabolites onto metabolic networks and minimized the search space through topological analysis within those networks, which highlighted mechanistically relevant pairs of candidate genes and metabolites. Fourth, using shortest path lengths across the metabolic network, we prioritized genes for mechanistic investigation by gene silencing and uncovered novel roles for *Asns* and *Galm* in adipocyte biology, thus validating the pertinence of our network topology analysis.

This systems genetics approach provides insights into possible coordinated mechanisms impacted by distinct series of blocks of genetic polymorphisms in well-defined genomic regions in congenic strains, as well as strain-specific phenotypic patterns. The overall interconnectedness of the association patterns between genomic blocks and metabolites illustrates the complex genetic architecture of metabolic regulation (Fig. 6). In

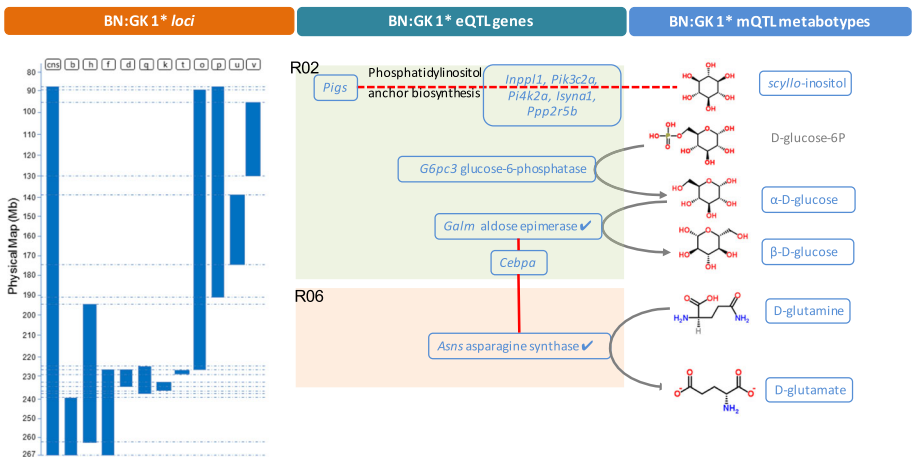


Fig. 6 Illustration of network-based mapping of eQTL and mQTL signals. Synthetic functional map illustrating biological connections between genomic regions *R02* and *R06* of chromosome 1 and differential regulation of transcripts and metabolites

particular, we identified coordinated changes in the regulation of metabolite abundance (D-glucose, L-glutamine) underpinned by *trans*-mediated expression of two genes (*Galm* and *Asns*) (Fig. 6) and uncovered a novel role for these genes in adipocyte function, which may have repercussions on pathophysiological phenotypes.

Our experimental approach in congenic strains allowed the dissection of the biological consequences of overlapping series of contiguous, localized GK genetic polymorphisms, thus limiting gene × gene interactions (epistatic effects) to interactions between homozygous GK variants present in the same genomic blocks. Chromosome substitution strains also demonstrated the impact of epistatic interactions on the detection and significance of genetic associations [9, 29]. We show that, even in the simplified context of congenic series, the regulation of adipose tissue metabolism involves different combinations of metabolites (i.e., specific metabotypes), thus providing experimental evidence for the strong capacity of organ metabolism to adapt to subtle genetic changes.

Such complexity of the metabolic regulation in the congenic panel was illustrated by strain-specific changes in metabolite abundance in adipose tissue, which may account for phenotypes discriminating GK and BN rats, including primarily glucose tolerance and adiposity [9]. Genome mapping of these effects indicate that GK genetic polymorphisms in several regions of rat chromosome 1 independently affect phosphatidylinositol signaling and glucose sensing and metabolism, suggesting functional redundancy of genes to ensure maintenance of essential phenotypes [35]. Prime examples are significant associations of GK haplotypes with *myo*-inositol (region 15) and *scyllo*-inositol (regions 2 and 6), which are stereoisomers of inositol, as well as with glucose and L-glutamine (regions 2 and 6) (Fig. 6). Inositol derivatives regulate insulin signaling in humans [37] and their conversion is reduced in insulin-sensitive tissues in the GK rat [38]. Chronic *myo*-inositol treatment results in improved glucose homeostasis and decreased adipocyte volume [41]. Glutamine is a TCA cycle-replenishing substrate and its analogues regulate insulin sensitivity in cultured adipocytes by preventing the desensitization of the glucose transport system [42].

Transcriptome data in the congenic strains provided possible explanations for changes in metabolite regulation, when the genetic control of metabolites and transcripts encoding biologically relevant proteins co-localizes in the same genomic region. We were able to map to the same regions of chromosome 1 the genetic control of the abundance of inositol compounds and transcripts functionally relevant to inositol metabolism, including phosphatidylinositol glycan anchor biosynthesis class S (*Pigs*; LOD = 7.42), inositol polyphosphate phosphatase-like 1 (*Inppl1*; LOD = 8.04), phosphoinositide-3-kinase, catalytic subunit type 2 alpha (*Pik3c2a*; LOD = 5.26), phosphatidylinositol 4-kinase type 2 alpha (*Pi4k2a*; LOD = 9.87), and inositol-3-phosphate synthase 1 (*Isyna1*; LOD = 5.73), which plays a critical role in *myo*-inositol biosynthesis. Furthermore, association with the transcript encoding protein phosphatase 2, regulatory subunit B' beta (*Ppp2r5b*; LOD = 6.34), regulated by *Pik3c2a*, supports the proposed role of defective regulation of serine/threonine protein phosphatases on insulin resistance in GK adipocytes [43].

Integrative analysis of multi-level -omic datasets in QTL mapping contexts remains a challenge [39], which has been best addressed in segregating populations in yeast [44] and mice [45] through the application of network biology tools. Through search space minimization within metabolic networks, we designed our metabolic network topology analysis to highlight direct coordinated functional relationships between genetically determined transcripts and metabolites and to prioritize them

for experimental validation. This approach is particularly relevant compared to recent advances in the field. For instance, MetaboNetworks [47] computes the minimal network interconnecting a metabolite list but ignores gene lists. IMPaLA [49] integrates gene and metabolite lists to perform over-enrichment analyses whilst Ambient [50] agnostically identifies modules from metabolite and gene lists in the absence of arbitrarily defined pathway boundaries. These approaches can be applied to eQTL-associated genes and mQTL-associated metabolites but do not specifically rely on minimization of the search space within the network.

Through minimization of shortest path lengths between candidate eQTL genes and associated mQTL metabolites, we successfully prioritized and validated the mechanistic relevance of pairs of transcripts and metabolites, such as L-glutamine and asparagine synthetase (glutamine-hydrolyzing) (*Asns*) in region 6 of chromosome 1 corresponding to the QTL linked to adiposity in the GK × BN F2 cross [9] (Fig. 1c) or D-glucose and galactose mutarotase (aldose 1-epimerase; *Galm*) in region 2. Loci in these regions accounted for *trans*-mediated genetic control of transcripts for *Asns* (LOD = 5.70) and *Galm* (LOD = 5.34). ASNS converts glutamate and asparagine into glutamine and aspartate and GALM catalyses the interconversion of the two glucose anomers. We were able to demonstrate the roles of these genes in vitro in 3T3-L1 cells in cellular differentiation (*Asns*) and glucose uptake (*Galm*), providing new insights into their function in (pre)adipocyte physiology.

Results from functional genomic analyses in BN.GK congenics illustrate the molecular consequences of naturally occurring DNA polymorphisms originally selected in the GK strain for their role in the regulation of glucose homeostasis [51]. The >175-Mb GK genomic region in the main congenic strain (1cns) contains over 362,300 DNA variants when compared to BN, including 264 non-synonymous coding variants and exonic deletions [55], which were dissected out in shorter haplotypes (down to 1 Mb) in congenic sub-strains. A report of polymorphisms in *Inppl1* in the GK affecting insulin sensitivity [56] is consistent with our finding of disrupted regulation of metabolites involved in phosphatidylinositol signaling in adipose tissue in congenics. Furthermore, analysis of genome sequence data in 27 inbred rat strains showed that the promoter region of *Galm* contains a series of DNA variants unique to the GK [55], suggesting that they could be etiologically relevant to glucose intolerance and adiposity in the GK strain.

Conclusions

We developed a novel network-based systems genetics [59] framework for joint mQTL and eQTL analyses of metabolic and gene expression profiles in adipose tissue of a series of rat congenic strains to dissect metabolic regulations and identified underlying physical and functional links between significant genes and metabolites, best exemplified by the novel biological roles we describe for *Galm* and *Asns* in adipocytes. Our metabolic network topology analysis approach integrates haplotype-specific co-regulated metabolites and gene transcripts, thus providing crucial information for functional annotation of genomes and for deciphering disease-associated molecular mechanisms.

Additional files

Additional file 1: Supplementary Tables 1-3. (PDF 109 kb)
Additional file 2: Supplementary Figures S1–S5. (PDF 2442 kb)
Additional file 3: Supplementary Tables 4-6. (XLSX 28452 kb)

Funding
This work is supported by the Wellcome Trust and grants from the European Commission (FGENTCARD — Functional genomic diagnostic tools for coronary artery disease; LSHG-CT-2006-037683), the Fondation pour le Recherche Médicale (FRM) and the Agence Nationale pour la Recherche (ANR-08-GENOPAT-030). This research received funding from the European Community Seventh Framework Programme (FP7/2007-2013) under grant agreement number HEALTH-F4-2010-241504 (EURATRANS) and the UK Medical Research Council (MRC) under grant agreements (MR/M501797/1, MR/K501281/1). MED was the recipient of a Young Investigator Award from ANR (ANR-07-JCJC-0042) and DG held a Wellcome Trust senior fellowship in basic biomedical science (057733).

Availability of data and materials
The transcriptomic dataset generated during and/or analyzed during the current study are available from ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-1048. The ^{1}H NMR metabolomic dataset is included in this published article (Additional file 3: Table S6).

Authors' contributions
MED and DG conceived the study; SC, NSZ, SCC, RBH, YW, CH, and KA generated experimental data; MED, CD, ARM, RA, SPW, QG, GWO, VN, SCM, JCL, EH JBC, JKN, and DG analyzed data; MED, CD, and DG wrote the paper. All authors approved the manuscript.

Competing interests
MED, YW, and SCM have performed consultancy work for Metabometrix Ltd, which was funded by EH, JCL, and JKN. The remaining authors declare that they have no competing interests.

Ethics approval and consent to participate
Work on animals was carried out under the project license 30/2324 issued by the UK Home Office and reviewed annually by the Animal Ethics Committee of the University of Oxford.

Author details
[1]Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Sir Alexander Fleming Building, Imperial College, London SW7 2AZ, UK. [2]Centre de Résonance Magnétique Nucléaire à Très Hauts Champs, 5 rue de la Doua, Villeurbanne 69100, France. [3]Metabometrix Ltd, Prince Consort Road, London SW7 2BP, UK. [4]UMR Modélisation Systémique Appliquée aux Ruminants, INRA, AgroParisTech, Université Paris-Saclay, Paris 75005, France. [5]Sorbonne Universities, University Pierre & Marie Curie, University Paris Descartes, Sorbonne Paris Cité, INSERM, UMR_S 1138, Cordeliers Research Centre, Paris 75006, France. [6]The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, UK. [7]MRC-University of Glasgow Centre for Virus Research, Glasgow G61 1QH, UK. [8]Key Laboratory of Magnetic Resonance in Biological Systems, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Centre for Magnetic Resonance, Wuhan Institute of Physics and Mathematics, University of

Chinese Academy of Sciences, Wuhan 430071, China. [9]Centre for Computational Biology, University of Birmingham, Haworth Building, Birmingham B15 2TT, UK.

Received: 10 March 2016 Accepted: 12 September 2016
Published online: 30 September 2016

References

1. Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498:99–103.
2. Dumas M-E. Metabolome 2.0: quantitative genetics and network biology of metabolic phenotypes. Mol Biosyst. 2012;8:2494–502.
3. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes. 2014;63:2158–71.
4. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. Living with water stress: evolution of osmolyte systems. Science. 1982;217:1214–22.
5. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–53.
6. Ozcan U, Yilmaz E, Ozcan L, Furuhashi M, Vaillancourt E, Smith RO, et al. Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. Science. 2006;313:1137–40.
7. Bandyopadhyay A, Saxena K, Kasturia N, Dalal V, Bhatt N, Rajkumar A, et al. Chemical chaperones assist intracellular folding to buffer mutational variations. Nat Chem Biol. 2012;8:238–45.
8. Ma J, Pazos IM, Gai F. Microscopic insights into the protein-stabilizing effect of trimethylamine N-oxide (TMAO). Proc Natl Acad Sci U S A. 2014;111:8476–81.
9. Gauguier D, Froguel P, Parent V, Bernard C, Bihoreau MT, Portha B, et al. Chromosomal mapping of genetic loci associated with non-insulin dependent diabetes in the GK rat. Nat Genet. 1996;12:38–43.
10. Tripathy D, Chavez AO. Defects in insulin secretion and action in the pathogenesis of type 2 diabetes mellitus. Curr Diab Rep. 2010;10:184–91.
11. Wallis RH, Collins SC, Kaisaki PJ, Argoud K, Wilder SP, Wallace KJ, et al. Pathophysiological, genetic and gene expression features of a novel rodent model of the cardio-metabolic syndrome. PLoS One. 2008;3:e2962.
12. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518:187–96.
13. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44:981–90.
14. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science. 2013; 341:1241214.
15. Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Lu H, et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature. 2010;467:460–4.
16. Akerfeldt MC, Howes J, Chan JY, Stevens VA, Boubenna N, McGuire HM, et al. Cytokine-induced beta-cell death is independent of endoplasmic reticulum stress signaling. Diabetes. 2008;57:3034–44.
17. Lees H, Swann J, Poucher SM, Nicholson JK, Holmes E, Wilson ID, et al. Age and microenvironment outweigh genetic influence on the Zucker rat microbiome. PLoS One. 2014;9:e100916.
18. Dumas M-E, Wilder SP, Bihoreau M-T, Barton RH, Fearnside JF, Argoud K, et al. Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models. Nat Genet. 2007;39:666–72.
19. Nicholson JK, Foxall PJ, Spraul M, Farrant RD, Lindon JC. 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. Anal Chem. 1995;67: 793–811.
20. Gavaghan CL, Holmes E, Lenz E, Wilson ID, Nicholson JK. An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. FEBS Lett. 2000;484:169–74.
21. Blaise BJ, Giacomotto J, Elena B, Dumas M-E, Toulhoat P, Ségalat L, et al. Metabotyping of Caenorhabditis elegans reveals latent phenotypes. Proc Natl Acad Sci U S A. 2007;104:19808–12.
22. Elliott P, Posma JM, Chan Q, Garcia-Perez I, Wijeyesekera A, Bictash M, et al. Urinary metabolic signatures of human adiposity. Sci Transl Med. 2015;7:285ra62.
23. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17:448–53.
24. Huang-Doran I, Bicknell LS, Finucane FM, Rocha N, Porter KM, Tung YCL, et al. Genetic defects in human pericentrin are associated with severe insulin resistance and diabetes. Diabetes. 2011;60:925–35.
25. Tiller G, Laumen H, Fischer-Posovszky P, Finck A, Skurk T, Keuper M, et al. LIGHT (TNFSF14) inhibits adipose differentiation without affecting adipocyte metabolism. Int J Obes (Lond). 2011;35:208–16.
26. Cazier J-B, Kaisaki PJ, Argoud K, Blaise BJ, Veselkov K, Ebbels TMD, et al. Untargeted metabolome quantitative trait locus mapping associates variation in urine glycerate to mutant glycerate kinase. J Proteome Res. 2012;11:631–42.
27. Song DH, Getty-Kaushik L, Tseng E, Simon J, Corkey BE, Wolfe MM. Glucose-dependent insulinotropic polypeptide enhances adipocyte development and glucose uptake in part through Akt activation. Gastroenterology. 2007; 133:1796–805.
28. Hedjazi L, Gauguier D, Zalloua PA, Nicholson JK, Dumas M-E, Cazier J-B. mQTL. NMR: an integrated suite for genetic mapping of quantitative variations of (1)H NMR-based metabolic profiles. Anal Chem. 2015;87:4377–84.
29. Buchner DA, Nadeau JH. Contrasting genetic architectures in different mouse reference populations used for studying complex traits. Genome Res. 2015;25:775–91.
30. Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46:543–50.
31. Kettunen J, Tukiainen T, Sarin A-P, Ortega-Alonso A, Tikkanen E, Lyytikäinen L-P, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet. 2012;44:269–76.
32. Nicholson G, Rantalainen M, Li JV, Maher AD, Malmodin D, Ahmadi KR, et al. A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. PLoS Genet. 2011;7:e1002270.
33. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet. 2010;42:137–41.
34. Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, Michael K, et al. A genome-wide association study of metabolic traits in human urine. Nat Genet. 2011;43:565–9.
35. Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, et al. Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell. 2006;21:319–30.
36. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.
37. Heimark D, McAllister J, Larner J. Decreased myo-inositol to chiro-inositol (M/C) ratios and increased M/C epimerase activity in PCOS theca cells demonstrate increased insulin sensitivity compared to controls. Endocr J. 2014;61:111–7.
38. Pak Y, Hong Y, Kim S, Piccariello T, Farese RV, Larner J. In vivo chiro-inositol metabolism in the rat: a defect in chiro-inositol synthesis from myo-inositol and an increased incorporation of chiro-[3H]inositol into phospholipid in the Goto-Kakizaki (G.K) rat. Mol Cells. 1998;8:301–9.
39. Ghazalpour A, Bennett BJ, Shih D, Che N, Orozco L, Pan C, et al. Genetic regulation of mouse liver metabolite levels. Mol Syst Biol. 2014;10:730.
40. Davidovic L, Navratil V, Bonaccorso CM, Catania MV, Bardoni B, Dumas M-E. A metabolomic and systems biology perspective on the brain of the fragile X syndrome mouse model. Genome Res. 2011;21:2190–202.
41. Croze ML, Vella RE, Pillon NJ, Soula HA, Hadji L, Guichardant M, et al. Chronic treatment with myo-inositol reduces white adipose tissue accretion and improves insulin sensitivity in female mice. J Nutr Biochem. 2013;24:457–66.
42. Marshall S, Bacote V, Traxinger RR. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. J Biol Chem. 1991;266:4706–12.
43. Begum N, Ragolia L. Altered regulation of insulin signaling components in adipocytes of insulin-resistant type II diabetic Goto-Kakizaki rats. Metab Clin Exp. 1998;47:54–62.
44. Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, et al. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. PLoS Biol. 2012;10:e1001301.
45. Ferrara CT, Wang P, Neto EC, Stevens RD, Bain JR, Wenner BR, et al. Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling. PLoS Genet. 2008;4:e1000034.
46. Argoud K, Wilder SP, McAteer MA, Bihoreau MT, Ouali F, Woon PY, et al. Genetic control of plasma lipid levels in a cross derived from normoglycaemic

Brown Norway and spontaneously diabetic Goto-Kakizaki rats. Diabetologia. 2006;49:2679–88.
47. Posma JM, Robinette SL, Holmes E, Nicholson JK. MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. Bioinformatics. 2014;30:893–5.
48. Collins SC, Wallis RH, Wallace K, Bihoreau M-T, Gauguier D. Marker-assisted congenic screening (MACS): a database tool for the efficient production and characterization of congenic lines. Mamm Genome. 2003;14:350–6.
49. Kamburov A, Cavill R, Ebbels TMD, Herwig R, Keun HC. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. Bioinformatics. 2011;27:2917–8.
50. Bryant WA, Sternberg MJE, Pinney JW. AMBIENT: Active Modules for Bipartite Networks–using high-throughput transcriptomic data to dissect metabolic response. BMC Syst Biol. 2013;7:26.
51. Goto Y, Kakizaki M, Masaki N. Production of spontaneous diabetic rats by repetition of selective breeding. Tohoku J Exp Med. 1976;119:85–90.
52. Collins SC, Wallis RH, Wilder SP, Wallace KJ, Argoud K, Kaisaki PJ, et al. Mapping diabetes QTL in an intercross derived from a congenic strain of the Brown Norway and Goto-Kakizaki rats. Mamm Genome. 2006;17:538–47.
53. Wallis RH, Wallace KJ, Collins SC, McAteer M, Argoud K, Bihoreau MT, et al. Enhanced insulin secretion and cholesterol metabolism in congenic strains of the spontaneously diabetic (Type 2) Goto Kakizaki rat are controlled by independent genetic loci in rat chromosome 8. Diabetologia. 2004;47:1096–106.
54. Wallace KJ, Wallis RH, Collins SC, Argoud K, Kaisaki PJ, Ktorza A, et al. Quantitative trait locus dissection in congenic strains of the Goto-Kakizaki rat identifies a region conserved with diabetes loci in human chromosome 1q. Physiol Genomics. 2004;19:1–10.
55. Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, et al. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. Cell. 2013;154:691–703.
56. Marion E, Kaisaki PJ, Pouillon V, Gueydan C, Levy JC, Bodson A, et al. The gene INPPL1, encoding the lipid phosphatase SHIP2, is a candidate for type 2 diabetes in rat and man. Diabetes. 2002;51:2012–7.
57. Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, Lindon JC, et al. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. Anal Chem. 2005;77:517–26.
58. Zhang JD, Wiemann S. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. Bioinformatics. 2009;25:1470–1.
59. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. Nat Rev Genet. 2014;15:34–48.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit


BioMed Central