

UC Davis

UC Davis Previously Published Works

Title

Identification of genetic elements in metabolism by high-throughput mouse phenotyping

Permalink

<https://escholarship.org/uc/item/6g05h65s>

Journal

Nature Communications, 9(1)

ISSN

2041-1723

Authors

Rozman, Jan

Rathkolb, Birgit

Oestereicher, Manuela A

et al.

Publication Date

2018

DOI

10.1038/s41467-017-01995-2

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>


Peer reviewed

ARTICLE

DOI: 10.1038/s41467-017-01995-2

OPEN

Identification of genetic elements in metabolism by high-throughput mouse phenotyping

Jan Rozman et al. 

Metabolic diseases are a worldwide problem but the underlying genetic factors and their relevance to metabolic disease remain incompletely understood. Genome-wide research is needed to characterize so-far unannotated mammalian metabolic genes. Here, we generate and analyze metabolic phenotypic data of 2016 knockout mouse strains under the aegis of the International Mouse Phenotyping Consortium (IMPC) and find 974 gene knockouts with strong metabolic phenotypes. 429 of those had no previous link to metabolism and 51 genes remain functionally completely unannotated. We compared human orthologues of these uncharacterized genes in five GWAS consortia and indeed 23 candidate genes are associated with metabolic disease. We further identify common regulatory elements in promoters of candidate genes. As each regulatory element is composed of several transcription factor binding sites, our data reveal an extensive metabolic phenotype-associated network of co-regulated genes. Our systematic mouse phenotype analysis thus paves the way for full functional annotation of the genome.

Correspondence and requests for materials should be addressed to M.H.d.A. (email: hrabe@helmholtz-muenchen.de)

#A full list of authors and their affiliations appears at the end of the paper

Metabolic disorders, including obesity and type 2 diabetes mellitus, are major challenges for public health. High initial treatment costs are compounded by complications that arise after diagnosis, including a number of consequential diseases, which together generate a significant burden for health care systems^{1–5}. Genetic variations play established roles in the susceptibility and pathogenesis of these diseases^{6–11}. However, identification of the underlying gene variants and their pathogenic roles is difficult because (1) functional annotation is still not available for many genes, especially genes that may be involved in disease but currently lack biological characterization¹². (2) It also became clear that gene variants do not work in isolation but as parts of networks¹³. (3) There are concerns regarding the reproducibility, predictability, and relevance of

results obtained from genotype–phenotype associations in disease model organisms¹⁴.

At the International Mouse Phenotyping Consortium (IMPC), we are generating a comprehensive catalog of mammalian gene functions to gain functional insights for every protein-coding gene, by producing and phenotyping more than 20,000 knockout mouse strains^{15,16}. Knockout strains are analyzed in a comprehensive, standardized phenotyping screen that covers multiple areas of biology and disease with 14 compulsory test procedures and several additional optional tests (www.mousephenotype.org/impress). Phenotyping is conducted in 10 research centers in Europe, North America, and Asia. Each site follows the IMPC's standardized operating procedures (IMPreSS), which were developed during the pilot programs EUMORPHIA and

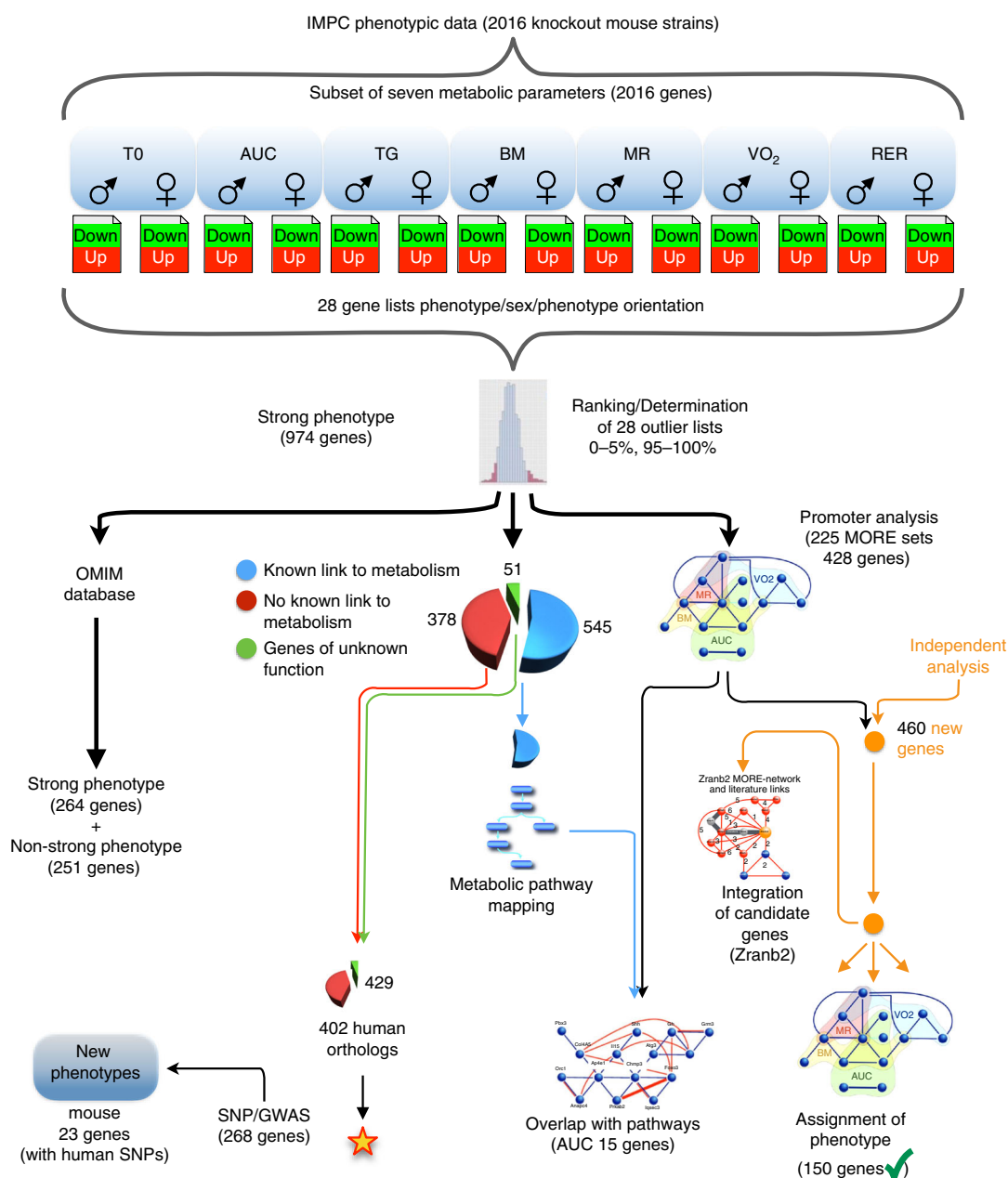


Fig. 1 Strategical abstract depicting the research strategy to identify new genetic elements in metabolism. The IMPC phenotyping data of 2016 knockout mouse strains was systematically evaluated for new links to human metabolic disorders. Nine hundred seventy-four knockout strains showed a strong metabolic phenotype. This set of genes was used as data mining resource. In a multiple line of evidence approach, we finally identified 23 genes that were linked to human disease-related SNPs

Table 1 Number of genes analyzed and candidate hits identified by phenotyping mutant and wild-type mice from both sexes

Parameter	Number of genes			Females and males (in brackets: expected number if sex equally affected)		
	Females	Males	Both	Outlier <5%	Outlier >95%	Outlier <5% and >95%
TO	1843	1840	1839	162 (96)	166 (96)	324 (192)
AUC 0-120	1846	1840	1839	172 (96)	163 (96)	334 (192)
Triglycerides	1384	1383	1380	124 (73)	129 (72)	249 (145)
Body mass	1649	1645	1645	121 (79)	143 (86)	264 (165)
Metabolic rate	335	910	329	55 (48)	53 (48)	108 (96)
VO ₂	335	910	329	58 (48)	52 (46)	110 (94)
RER	319	901	313	55 (47)	54 (46)	109 (93)
Total	1995	2012	2016	575	587	974

EUMODIC^{17,18}. Standardization, data quality control, an automated statistical analysis pipeline, and the phenotyping of reference strains to assess inter-center variation all help to ensure robust and reproducible data^{19–22}. Adherence to each of these standards enables derivation of a high-quality, powerful hypothesis-generating resource that includes genes from a substantial proportion of the entire mouse genome. Here we analyze phenotyping data of 2016 knockout strains for strong metabolic phenotypes and identify 974 known and previously unannotated genes with relevance for metabolic diseases. Twenty-three genes for which we find new strong metabolic phenotypes also have links to human metabolic disorders. We introduce a focus on network/context analysis to demonstrate how systemic phenotype information can be linked to known metabolic pathways. This pathway mapping revealed an unexpected degree of metabolic dimorphism between sexes. In addition, phenotype-associated regulatory networks allow the prediction of previously unknown gene functions. Therefore, our results underline the value of the IMPC resource for gene function discovery and augment the translational potential of metabolic phenotypes in mice.

Results

Strong metabolic phenotypes in IMPC mutants. We analyzed a total of 2016 IMPC mouse strains that were homozygous for a single-gene knockout on a C57BL/6N background or heterozygous when homozygotes were lethal or sub-viable (see Fig. 1 for a study overview). We chose seven metabolic parameters with diagnostic relevance in human clinical research for our study: fasting basal blood glucose level before glucose tolerance test (TO), area under the curve of blood glucose level after intraperitoneal glucose administration relative to basal blood glucose level (AUC), plasma triglyceride levels (TG), body mass (BM), metabolic rate (MR), oxygen consumption rate (VO₂), and respiratory exchange ratio (RER), which is a measure of whole-body metabolic fuel utilization. To identify the universal cross-project metabolic characteristics of knockouts, we calculated mean mutant/wild-type ratios separately for all contributing centers²³. Males and females were analyzed separately because sexual dimorphism is common in disease-related phenotypes^{24,25} (Fig. 1, top part). For all parameters, mutant/wild-type ratios were distributed around a modal value of 1.00, which would be expected when mutant and wild-type mice on average did not differ in the respective parameter (Table 1 and Fig. 2a–n). The shape of the metabolism phenotype mutant/wild-type ratio distributions differed between some parameters more than others. Triglycerides, basal blood glucose, and glucose clearance varied more between mutant and wild-type mice than RER, BM, VO₂, and MR. For further analysis, we focused on gene knockout strains with a strong metabolic phenotype and compiled lists of genes for which the knockout resulted in mutant/wild-type ratios below the 5th percentile and above the 95th

percentile of the ratio distributions (shown as the filled areas in Fig. 2). Based on these thresholds, we generated 28 gene lists, one for every sex parameter combination. Our 28 lists included a total of 974 “strong metabolic phenotype” genes (Fig. 1, second part from top). We used these genes as a data mining resource for further investigation into potential links to human metabolic disorders (see Supplementary Data 1—Mutant wildtype ratios for complete gene lists and Supplementary Data 2—Strong metabolic phenotype genes for strong phenotype genes).

Evaluation of false discovery rates. The reproducibility and robustness of large-scale biology projects depends on minimizing the risks of false-positive or false-negative results. Using a post hoc approach to evaluate the risk of false-negative findings, we compiled a list of 666 mouse and human genes with published links to obesity and type 2 diabetes^{25–29}. We then used these candidate genes to estimate the rate of false-negative discoveries in our project (see Supplementary Data 3—Candidate genes for false negative discovery). Hundred and one knockout mutants of these genes had been phenotyped by the IMPC at the time when we conducted our analysis. Hundred of 101 had a detectable metabolic phenotype: 58 (57.4%) scored as “strong metabolic phenotype genes” (*p* value 0.04, null hypothesis was defined as no enrichment of “strong metabolic phenotype” genes, see “Methods” section for description of the permutation simulation). With only one exception (*Serpinf1*, 210 MGI: 108080), knockout mice of all other 43 remaining genes even though not scoring as “strong metabolic phenotype genes” had phenotypic deviations in metabolic parameters in the range of low 20th or high 80th percentiles of their respective frequency distributions (*p* value 0.047).

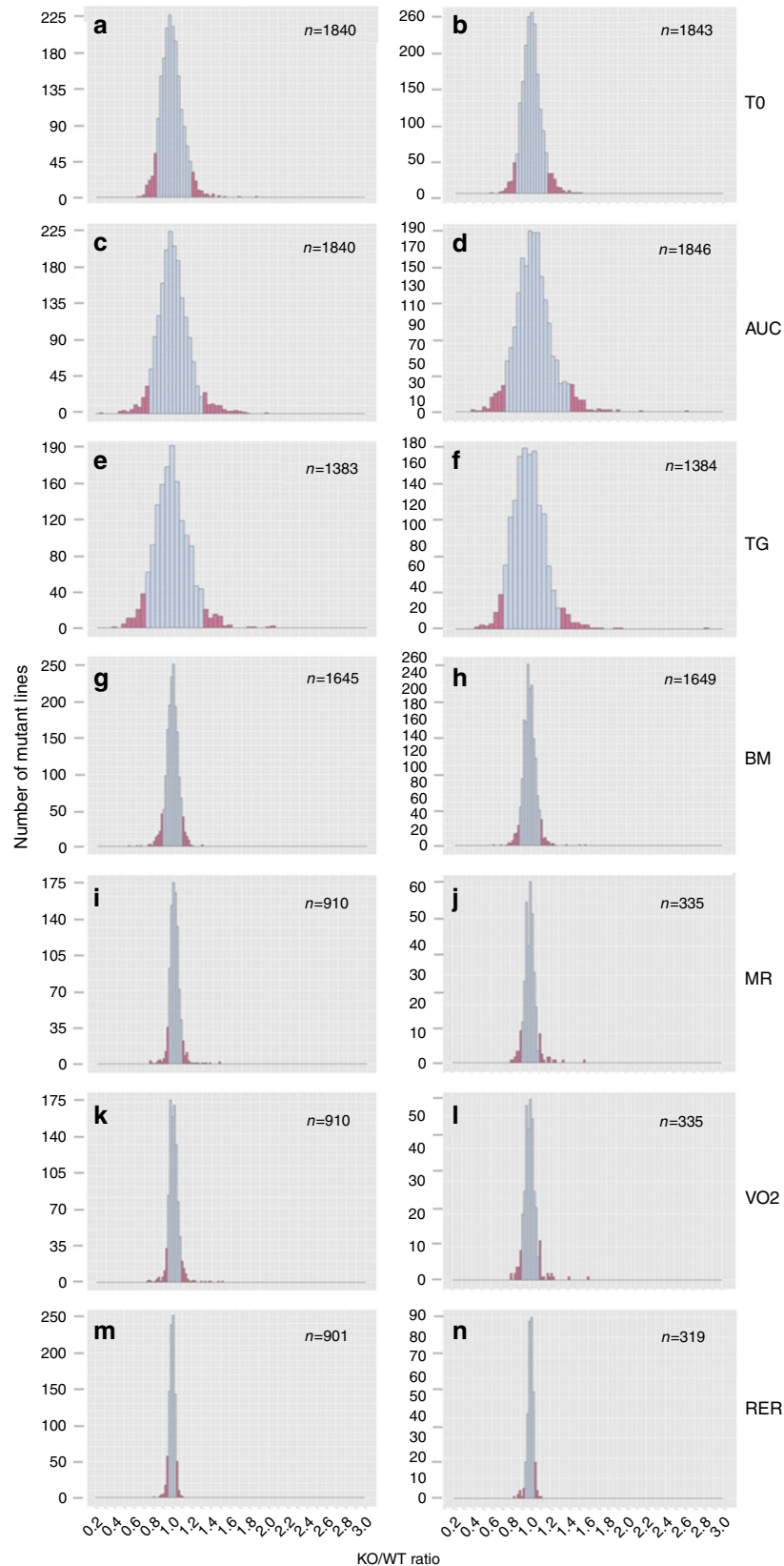
New knockout mouse models for metabolic diseases. For 58 of our 666 candidate genes with known links to metabolic functions, the link was based only on GWAS data with no further evidence available. Based on a systematic search of the mouse genome informatics (MGI) curated database, knockout mouse models had not yet been described for 4 of these 58 candidate genes before our study: *Lypla1*²⁸, *Rfx6*²⁶, *Slc6a14*, and *Slc6a3*²⁹. Knockout mouse models had been generated previously for another 15 candidate genes but these had either no proven link to metabolism or were not reported to have been tested for metabolic parameters. Our phenotype data can therefore identify new genetic disease mouse models that allow in-depth investigation of disease mechanisms and fill the gap between genome-wide association studies and functional validation in a mammalian model organism.

Illuminating unexplored mouse metabolic genes. Many genes still lack any functional annotation. We conducted a systematic search of the MGI curated database to identify murine genes in our “strong metabolic phenotype” list with no previously known

links to metabolism (see Supplementary Fig. 1 and Supplementary Data 4—Search results for unexplored metabolic genes). By careful stepwise evaluation, we found 429 of the 974 strong metabolic phenotype genes had no link to metabolic functions in mice. Fifty-one of these genes had no functional annotation at all (Fig. 1, center part). Their knockout caused strong phenotypes

regarding glucose and energy metabolism (Fig. 3). Our data analysis provides evidence for new links to metabolic functions for these hitherto uncharacterized genes.

Function of new metabolism genes is linked to human disease.
 We were interested in whether the strong metabolic phenotype



genes could be mapped to diseases in the OMIM database. Of all 515 genes that were linked to at least one OMIM disease, 264 (27.1%) belonged to the strong metabolic phenotype genes, whereas 251 (24.1%) were “non-strong metabolic phenotype” genes (Fig. 1, lower left branch; also see Supplementary Data 5—OMIM mapping table).

As a next step, we searched for non-annotated disease links to connect previously unknown strong metabolic phenotype genes to human disease. We found human orthologues for 402 of the 429 strong metabolic phenotype mouse genes. Note these 429 genes had no previously described link to metabolism and that includes the 51 completely unexplored mouse genes and 378 genes, which had no link to metabolism so far. Next, we extracted single-nucleotide polymorphisms (SNPs) in the SNIIPA database³⁰, from a ± 2 kb region around the 402 human orthologues of strong phenotype genes (Fig. 1, second lower branch from the left). The SNIIPA database has both functional annotations and linkage disequilibrium information for bi-allelic genomic variants (SNPs and single-nucleotide variations), based on the 1000 Genome Project. We found 19,253 SNPs for 268 of the 402 orthologous genes. SNIIPA SNPs have various annotation layers such as gene annotations, associated phenotypic traits, and expression based quantitative trait loci. For each SNP, we evaluated the extent of association across 16 type 2 diabetes-related traits using the cross phenotype meta-analysis method on data from the DIAGRAM, MAGIC, GIANT, GLGC, and ICBP GWAS consortia^{30–40} (Fig. 4). We applied different levels of evidence and p value thresholds to infer association with disease traits (Table 2). By applying a standard genome-wide significance level of $p \leq 5 \times 10^{-8}$, we found SNPs in 17 gene regions (*MTNR1B*, *MTCH2*, *SLC39A8*, *NUTF2*, *PABPC4*, *DNAJC5G*, *TCF19*, *PACIN3*, *EVI5*, *EPB41L4B*, *DMXL2*, *RPTOR*, *CCDC18*, *RPGRIP1L*, *PCNXL3*, *WNT3*, *ELMO3*) that were associated with human metabolic phenotypes. By applying a significance level of 2.6×10^{-6} and correcting for all the SNPs we looked up in GWAS data, we found that more SNPs in *CCDC18* and in four additional genes (*CFAP69*, *IQCE*, *LYPD6B*, and *NRDE2*) were strongly linked to human phenotypes (Fig. 4). Regarding the remaining genes, all SNPs with a p value above 0.05 nevertheless had weak links to metabolic phenotypes in humans and only seven of the 268 genes with SNPs (*CMTM5*, *PPP1R14A*, *DUSP5*, *FGFBP3*, *TIMM22*, *WNT6*, and *PPP1R35*) lacked potential links to metabolic phenotypes.

As a next step, we corroborated the role of so-far unexplored mouse metabolic genes in a well-phenotyped clinical cohort. We analyzed the associations of common SNPs (i.e., those with minor allele frequencies ≥ 0.05) in the human orthologues of these murine genes with five human metabolic phenotypes: body fat content/distribution, blood glucose, insulin sensitivity, insulin secretion, and plasma lipids. To do this, we assimilated genome-wide genotype data, generated with Illumina’s Infinium[®] Global Screening Array, from 2788 participants of the Tübingen Family (TÜF) study for type 2 diabetes⁴¹. We identified 240 common, bi-allelic, and non-linked SNPs (call rates ≥ 0.75) in 37 orthologues and analyzed them in the additive inheritance model by multiple linear regression analysis. Potential confounders were accounted for, e.g., gender, age, BMI, and insulin sensitivity, where

appropriate. After Bonferroni correction for the number of SNPs tested, we found two SNPs below the study-wide significance threshold of $p < 0.000213$. These are the minor allele of SNP rs11734172 in *C4orf22*, which is associated with reduced insulin sensitivity (HOMA-IR, $p = 8.8 \times 10^{-5}$; ISI Matsuda, $p = 6.2 \times 10^{-5}$), and the minor allele of SNP rs76378941 in *CNBD1* which is associated with reduced plasma triglycerides ($p = 0.00020$). We identified seven additional non-linked SNPs in *C4orf22* and two additional non-linked SNPs in *CNBD1* with nominal associations ($p < 0.05$) with insulin sensitivity and triglycerides, respectively (Fig. 1, bottom left; all association results are shown in Supplementary Data 6—Association results of the TÜF study). To address whether the identified SNPs in *C4orf22* and *CNBD1* likely affect other nearby genes, we explored linkage disequilibrium data of the CEU population (99 Utah residents of Central European origin) from phase 3 of the 1000 Genomes Project (http://grch37.ensembl.org/Homo_sapiens/Info/Index). Neither the *C4orf22* nor the *CNBD1* gene locus had any adjacent disequilibrium. Thus, the identified SNPs in *C4orf22* and *CNBD1* likely do not affect nearby genes. So in conclusion, our single-gene based analysis identified a total of 23 orthologous genes with a link to human disease traits. We next focused on context and network aspects of our findings.

Pronounced metabolic sexual dimorphism on gene level. As we had discovered connections between several strong metabolic phenotype murine genes and human disease, we next focused on the biological gene networks that underlie these phenotypes and their interconnections. First, we mapped our strong metabolic phenotype genes to metabolic pathways in the KEGG repository⁴². At least one knockout gene for each metabolic phenotype analyzed mapped to a metabolic pathway (males and females, Supplementary Data 7—Metabolic pathways in KEGG). We present three interesting findings: (i) genes from different pathways caused comparable metabolic phenotypes in males and females (Supplementary Data 8–14—Pathway maps). (ii) The 13 male and 13 female strong metabolic phenotype genes for triglyceride levels that were linked to the global metabolic pathway map of the KEGG database had one gene in common, *Hpse*. However, 13 male and 13 female strong metabolic phenotype genes mapped to the same pathway (out of 121 male and 121 female genes). (iii) Some metabolic pathways were only found in the set of female strong metabolic phenotype genes, and not in males affecting plasma triglyceride levels. These pathways were “glycerophospholipid metabolism,” “linoleic acid metabolism,” and “ether lipid metabolism” pathways. We mapped the strong triglyceride-phenotype genes to non-metabolic biological pathways in the KEGG database, this also revealed sex-specific differences (Supplementary Data 7—Metabolic pathways in KEGG). These results give insight into and confirm sexual dimorphism in genes and pathways associated with metabolic phenotypes in mice.

Regulatory networks of strong metabolic phenotype genes. Metabolic pathways require simultaneous and coordinated presence of functionally connected proteins. Therefore, we analyzed next

Fig. 2 Frequency distribution of mutant/wild-type ratios for metabolic parameters, separated for males and females. **a** TO females, basal blood glucose after overnight food deprivation, **b** TO males, basal blood glucose after overnight food deprivation, **c** AUC females, area under the curve of blood glucose excursions after glucose injection in a glucose tolerance test, **d** AUC females, area under the curve of blood glucose excursions after glucose injection in a glucose tolerance test, **e** TG females, plasma triglyceride concentrations, **f** TG males, plasma triglyceride concentrations, **g** body mass females, **h** body mass males, **i** MR females, metabolic rate obtained from a 21 h indirect calorimetry trial, **j** MR females, metabolic rate obtained from a 21 h indirect calorimetry trial, **k** VO₂ females, oxygen consumption obtained from a 21 h indirect calorimetry trial, **l** VO₂ males, oxygen consumption obtained from a 21 h indirect calorimetry trial, **m** RER females, respiratory exchange ratio, **n** RER females, respiratory exchange ratio. Filled areas of the distributions cover the <5% and >95% strong metabolic phenotype genes, n provides number of mutant lines

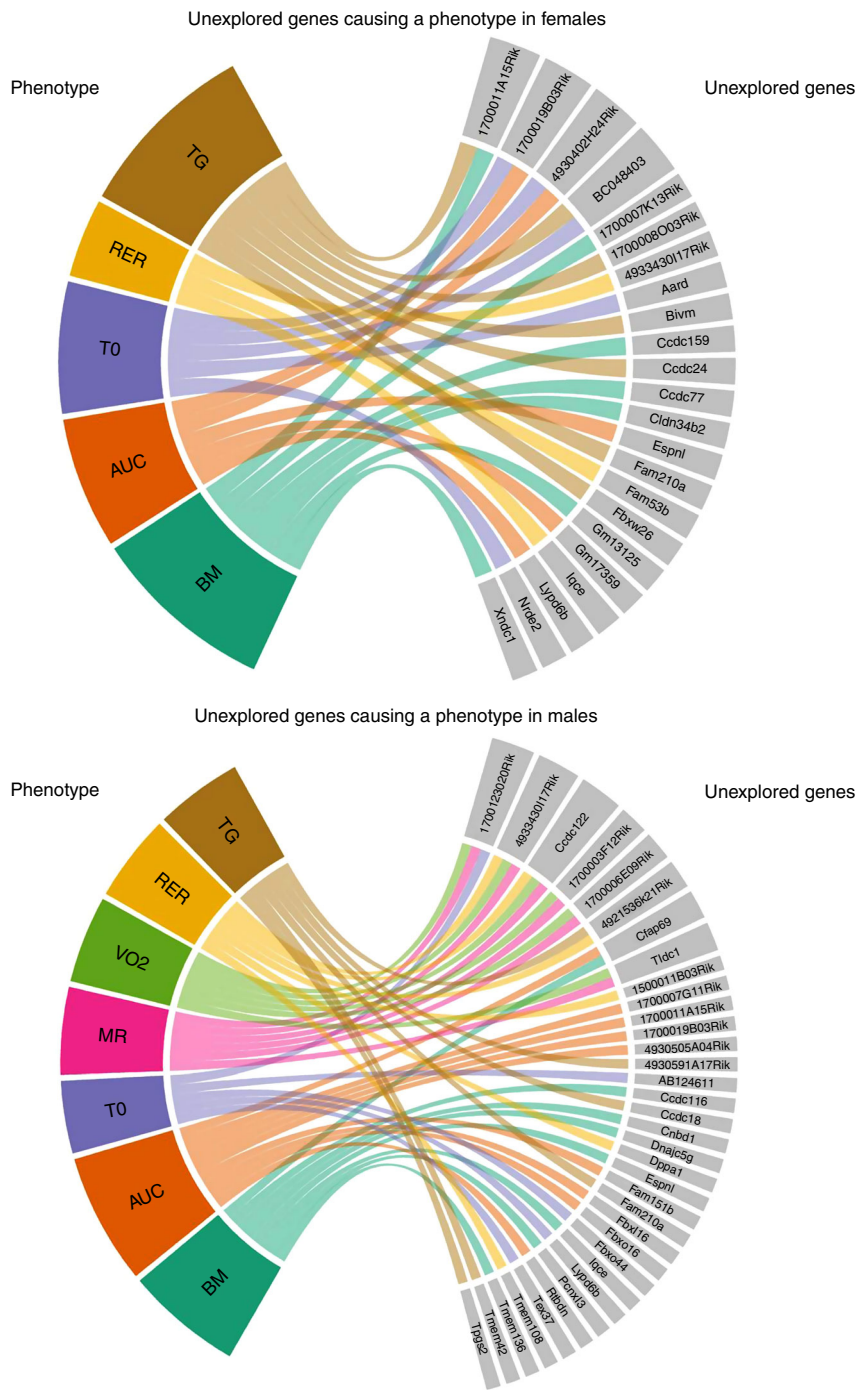


Fig. 3 Links between unexplored metabolic genes and parameters that contribute to strong metabolic phenotypes in females (upper) and males (lower)

whether our strong metabolic phenotype genes have common molecular regulatory features within their promoters. Transcriptional co-regulation often involves a common set of transcription factor-binding sites shared between co-regulated promoters. This concept of higher-level organized promoter wiring is known as multiple organized regulatory element (MORE) cassettes (see Supplementary Note 1). First, we extracted the promoter sequences of all our strong metabolic phenotype genes from the mouse genome sequence and analyzed them for the presence of identical MORE cassettes. We required promoter sets to contain at least three promoters. Analysis of all promoter sets for shared MORE cassettes resulted in a total of 225 sets of one or more MORE cassettes. These MORE sets were associated with at least one of the

four sub-phenotypes (male, female, up or down). They were subdivided into these four subtypes by sex, up or down (top and bottom percentiles, respectively) of the seven metabolic phenotypes, e.g., AUC female low or VO₂ female high. These associated MORE sets were present in 428 promoters of unique genes (Fig. 1, central part right). Genes sharing the same MORE set in at least one of their promoters thus belong to one regulatory network. We obtained a grouping of the collection of knockout genes independent of prior knowledge by the regulatory promotor analysis.

We next selected genes from our strong metabolic phenotype list that have two or more phenotypic associations with glucose homeostasis (18/20), body mass regulation (10/20), metabolic rate (2/20), and substrate utilization (2/20). The list of genes selected

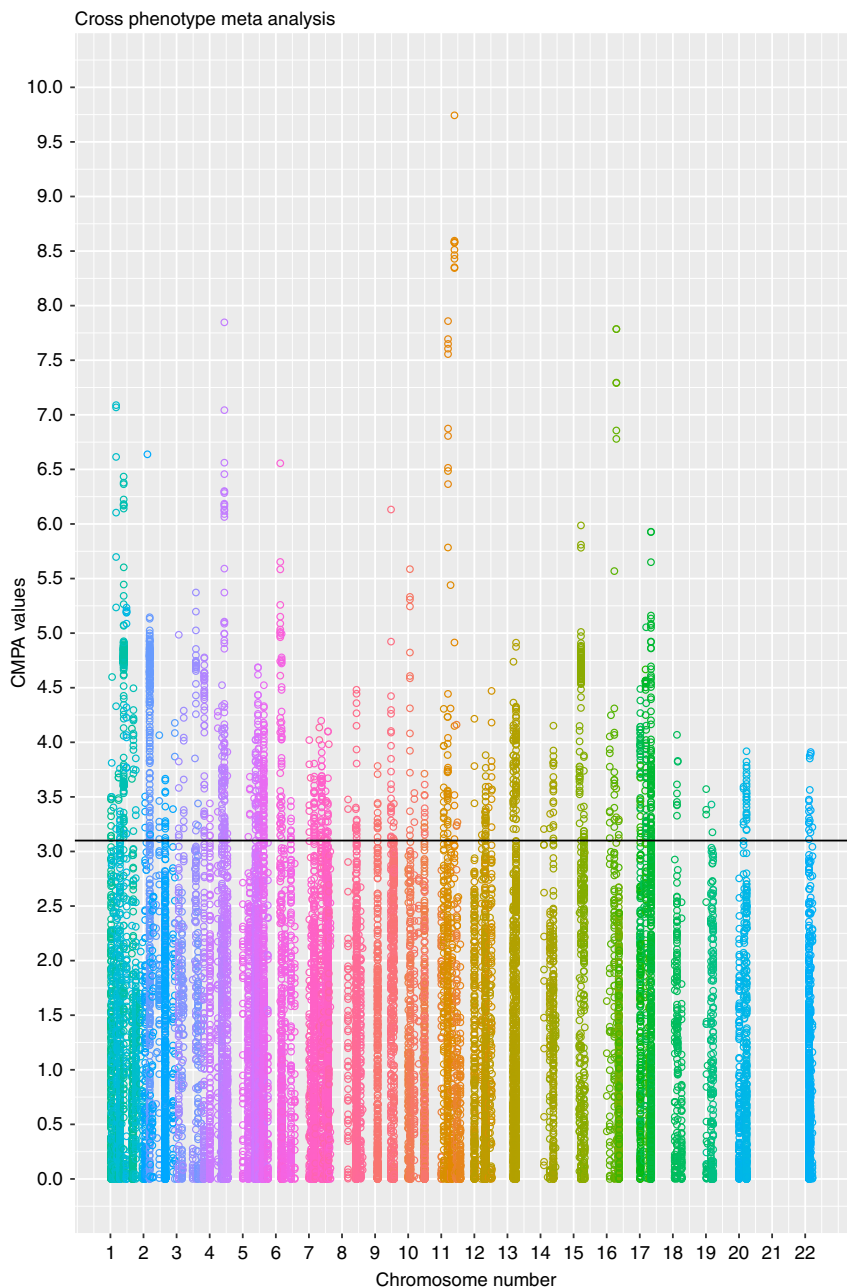


Fig. 4 Cross phenotype meta-analysis of murine genes without prior link to metabolism. SNPs are on the x-axis ordered as per chromosome and the CPM values (log transferred) are on the y-axis. The chromosomes are shown in different colors. Each column represents the genes and SNPs stacking vertically. The higher the CPM measure, the higher the significance of SNPs across different phenotypes. SNPs above CPM = 3.1 were considered to have a significant link to the disease traits

was *Asf1a*, *Atp2a2*, *Bbs5*, *Cir1*, *Commd9*, *Cpe*, *Dpm2*, *Dtnbp1*, *Epha5*, *Ggnbp2*, *Golga3*, *Il31ra*, *Lbp*, *Mboat7*, *Mrap2*, *Rabl2*, *Scrib*, *Slc2a2*, *Zranb1*, and *Zfp11* (see Supplementary Data 15—Genes with metabolic phenotype). We chose these 20 genes as targets to examine shared links to regulatory network links. These links were defined by MORE cassettes and MORE sets. Overall, we found several MORE sets in corresponding promoters and the majority of the MORE sets were in more than one gene promoter in the 20 genes set (Table 3). This overlap allowed us to construct an association network with 14 of the 20 genes (Fig. 5). Interestingly, 12 of these genes were present in one single network indicating potentially coordinated expression of these genes (Fig. 5a). The overrepresented MORE sets thus connect

genes within a specific phenotype, which suggests that individual regulatory mechanisms are confined to phenotypic subtypes (Fig. 5b). The overlap between different phenotypes is mainly due to genes that have more than one associated phenotype (Fig. 5c). Twelve of the 20 examined genes have common regulatory structures in their promoters that link them to a network. The edges of the network in Fig. 5c are shared MORE cassettes between the connected nodes (genes). Each edge is a phenotype or sub-phenotype association of the connected genes (nodes). Since MORE sets are molecular regulatory mechanisms, common MORE sets that link phenotype genes to each other indicate potential common regulatory mechanisms for the linked genes.

Table 2 Translation to human disorders

p value—class across metabolic phenotypes in human GWAS data	SNPs	Genes	Unexplored metabolic genes
>5.0e-08	89	17	3
2.59e-06 to 5.0e-08	52	12	1
1.0e-03 to 2.59e-06	1071	94	5
0.05 to 1.0e-03	9107	254	20
>0.05	8912	233(7)	17(0)

Search for single-nucleotide polymorphisms in prioritized genes from IMPC in a cohort of pre-diabetic patients
The number of genes in the category >0.05 not overlapping with genes from other categories is displayed in brackets

MORE sets and KEGG pathway mappings are overlapping networks. MORE sets are solely derived from promoter analysis without the use of prior knowledge of the phenotypes or the genes involved. Therefore, the resulting networks are independent of knowledge, often exceeding current knowledge. Because of that, validation of the regulatory network connection by existing current knowledge is rather limited. However, the genes within these regulatory networks are well-known and many are already associated with known pathways. We thus selected the genes that mapped to KEGG pathways and that had a strong glucose clearance phenotype to construct a MORE set network of 15 genes (Fig. 1, third lower branch from the left and Fig. 6a). A network for those 15 genes was constructed in which the edges were KEGG pathway networks to which both connected genes (nodes) mapped (Fig. 6b). Importantly, the two networks could be superposed, which thus validates combining our data-driven MORE set approach with experimental knowledge-based KEGG pathway mapping (Fig. 6c). In summary, our analysis supports the concept that the various initially unconnected knockout genes form well-organized functional metabolic networks. These networks are supported at several system levels (phenotype, regulation, and metabolism).

Prediction of metabolic genes. We hypothesized that the presence of MORE cassettes in currently unannotated genes could be used to functionally characterize them. Therefore, we proceeded determining which of the promoter sequences of the ~22,000 protein-coding genes in the mouse genome matched to the MORE cassettes defined in this study. We analyzed promoters of an independent set of genes that was not part of the original data set used to define the initial MORE cassettes (original data set: IMPC Release 4.2 from Dec 2015; additional set of genes: IMPC Release 4.3 from Apr 2016). The complementary data set contained 757 genes that had completed IMPC phenotyping after the earlier Release 4.2. A batch query of phenotype terms, from the IMPC portal (www.mousephenotype.org), for these genes showed that 150 of 460 genes (32.6%) with 1–11 MORE cassette matches had a metabolic phenotype, whereas only 68 of the total 297 genes (22.9%) with no MORE cassette matches had such a metabolic phenotype (32.6% vs. 22.9%, $p = 0.004$, Fisher's exact test; Fig. 1, right lower branch). Thus, the presence of phenotype-associated MORE cassettes does indeed predict genes with a corresponding phenotype.

The 9 genes with 6–11 MORE cassette matches associated with metabolic phenotypes were linked to the MORE set network defined in this study. This analysis also provided additional support for the relative position of the gene in the network in several cases. Figure 7 shows the *Zranb2* gene, which has six MORE sets in its promoters. *Zranb2* shares two central MORE sets with the *Dtnbp1* and *Epha5* genes, which confirms their

previously reported functional links with *Zranb2*^{42–45}. Gauging the presence of phenotype-associated MORE cassettes in promoters is thus currently the only method to predict whether an uncharacterized gene has a candidate phenotype, or whether disruption of the gene will cause the respective phenotype (Fig. 1, second from right lower branch).

Discussion

In this study, we have established and evaluated analysis and visualization tools that identify and select candidate genes with roles in glucose and energy metabolism starting from the pool of IMPC phenotype data of mouse knockout strains, which covers ~10% of mammalian protein-coding genes. This data set links new genes to disease-relevant metabolic phenotypes. As with all large-scale approaches, false discovery could be a major issue. We used 101 genes with a known or hypothesized link to obesity or type 2 diabetes to assess the false-negative discovery rate and our approach missed only one of the 101 genes. Therefore, metabolic phenotypes can be detected in almost all IMPC mouse models of published candidate genes for obesity and type 2 diabetes.

The pioneering systematic approach of our IMPC database is an advance toward the final frontier of genome functional annotation as 44% of the strong metabolic phenotype genes have not been linked to metabolism in mice (429 of the 974). Furthermore, 51 of these new metabolism genes had no previous functional annotation at all. Strikingly, orthologues of 23 of these genes had single-nucleotide polymorphisms associated with human metabolic disease phenotypes. With large-scale projects like the IMPC, adding functional annotation and disease links to so-far unannotated genes will continuously contribute to closing the systems biology knowledge gap in all fields, e.g., for unexplored metabolic genes. The data analysis approaches we describe are based on snapshot computational evaluation of current knowledge that is changing fast. Since our initial analysis, new suggested links to physiological functions of three so-far unexplored genes (*AB124611*, *Dppa1*, and *GM13125*) were found in an updated literature search.

The problem of candidate genes without established links to existing knowledge is as “old” as high-throughput analysis itself. These “out of context” genes are very hard to characterize functionally and pure statistical association (e.g., gene ontology analysis) is insufficient to understand their biological context. Since it is well established that genes only work in collaboration with other genes, our analysis focuses on elucidating the genetic context of newly discovered candidate genes. Our approach allows targeted verification of a simple “guilt-by-association” hypothesis that infers that a gene likely contributes to a specified pathway if it is frequently found in our lists alongside genes with a common function.

Our findings of wide-spread pathway and network structures underlying individual gene findings provide a new integrative view to our results. Our pathway network approach pinpointed genes that caused specific and interrelated metabolic phenotypes in specific genetic backgrounds. For example, we confirmed recently published findings regarding a strong sexual dimorphism with minimal or no shared genes between males and females with the same associated phenotypes⁴⁶. The metabolic discordance between males and females was generally lower at the pathway level than at the gene level. For example, only 1 out of 13 genes was common between the 13 strong metabolic phenotype genes for triglyceride levels for each sex, whereas both sets of 13 genes mapped to the same pathways.

Our pathway mapping necessarily used prior knowledge, like the metabolic (KEGG) pathways, which cannot be derived from newly inputted experimental data. We complemented this prior

Table 3 Transcription factor-binding site alignment of MORE sets comprising a regulatory network of 14 genes

MORE cassette set															
AUC-mh MHS	dtnpb1														golga3
AUC-mh PZE														asfa1	atp2a2
AUC-mh XSSh	dtnpb1				bbs5										
AUC-mh PSH	dtnpb1														dpm2
BM-fh CFS	dtnpb1			ggnbp2	slcs2a2	bbs5									mrap2
BM-mh GEgE				ggnbp2											cir1
MR-mh PBLs	dtnpb1			ggnbp2				zranb1							
MR-mh ASF	dtnpb1	epha5													
RER-mh XEE		epha5	ggnbp2												
RER-mh GSO		epha5							rabl2						
RER-mh MHS					slcs2a2	bbs5									
TG-fh XEgEg	dtnpb1			ggnbp2											
TG-fl NSF		epha5	ggnbp2												
VO2-mh XXCS		epha5						zranb1							
VO2-mh XXLSS		epha5			slcs2a2										
VO2-fh XCHS		epha5						zranb1							
VO2-fl XSSsf					slcs2a2				rabl2						cpe

knowledge-based approach by a purely data-driven (prior knowledge-independent) analysis using MORE sets. This MORE set analysis deciphered the molecular regulatory networks in the promoters of the strong metabolic phenotype genes. Our approach therefore detects functionally connected gene networks that are supported by several techniques.

Data-driven analysis has another important feature. MORE cassettes are invariant features like reading frames. It is therefore very likely that the promoters of some uncharacterized genes with the same phenotypes will have phenotype-associated MORE cassettes. If this is the case, it should be possible to use the presence of MORE cassettes for a priori prediction of phenotypes. We validated this hypothesis based on a set of newly phenotyped genes that were not part of our initial analysis. A considerable number of these genes also had metabolic phenotype-associated MORE sets. We found a significant correlation between the presence of these MORE sets and the link of genes to metabolic functions. Of course, also genes newly predicted based on MORE sets would be expected to link into the known context of genes for this phenotype. Therefore, we analyzed one example of a de novo predicted potential phenotype-causing gene (*Zranb2*) in more detail. We not only found additional verification for the pre-defined network, but also additional links in the literature supporting the network association of this gene. Therefore, the presence of characterized MORE sets provides opportunities for experimental planning of verification experiments.

In conclusion, we show here that our new multiple-line-of-evidence functional discovery approach, which is based on the IMPC phenotyping program, can identify new disease genes related to energy metabolism and glucose homeostasis. Our platform will thus enable researchers to prioritize research on so far uncharacterized genes to fill the gap in functional annotation for metabolic genes. We predicted gene functions in a set of 757 subsequently phenotyped genes, based on the integration of functionally uncharacterized genes into established regulatory networks and functional contexts. By linking gene functions to metabolic disorders, our protocol and the identification of metabolic relevant genetic elements will accelerate the understanding of human disease.

Methods

Mouse husbandry and phenotyping. We used the IMPC data resource to identify genes associated with strong metabolic phenotypes in mice. The IMPC phenotyping pipeline includes 14 mandatory and several optional tests that cover all major disease areas (www.mousephenotype.org/impress). Phenotyping procedures

are being conducted in 10 centers (Baylor College of Medicine BCM, Helmholtz-Zentrum München HMGU, PHENOMIN Institut Clinique de la Souris ICS, Jackson Laboratory JAX, Medical Research Council Harwell MRC Harwell, MARC Nanjing University NING, RIKEN BioResource Center RBRC, Toronto Centre for Phenogenomics TCP, Mouse Biology Program University of California Davis UC Davis, and Wellcome Trust Sanger Institute WTSI) in Europe, North America, and Asia (for details, see <http://www.mousephenotype.org/about-impc-impc-members>). Mice were generated on a C57BL/6N background and phenotyping data were collected between the age of 4 and 16 weeks following approved animal ethics protocols in every institution (see Supplementary Table 3 for license numbers).

Phenotyping parameters. From the 509 phenotyping parameters assessed by the IMPC early adult phenotyping screen (<http://www.mousephenotype.org/impress>), we chose seven parameters: (1) basal blood glucose levels after overnight food deprivation (T0) and (2) the area under the curve (AUC) of glucose excursions during the intraperitoneal glucose tolerance test as a read out for glucose-stimulated insulin secretion and insulin sensitivity (IPGTT) [IMPC_IPG_001]; (3) non-fasted triglyceride levels from clinical chemistry [IMPC_CBC_003]; (4) body weight [IMPC_DEXA_001]; (5) metabolic rate (MR) normalized to body mass (see below); (6) oxygen consumption (VO_2) normalized to body mass (see below); and (7) respiratory exchange ratio ($RER = VCO_2/VO_2$) from the indirect calorimetry trial [IMPC_CAL_003].

Bioinformatics. In general, only unique genes were included, i.e., in the case that data for two different zygositys were available, we selected homozygotes instead of heterozygotes. We included lines in which data of males and females with the same genotype was available instead of having different genotypes for either males or females. Multiple entries for the same gene only occurred in reference strains. For each test procedure (indirect calorimetry, clinical chemistry, and glucose tolerance test), we received a csv file containing phenotyping data (one row per mouse, including characteristics such as sex, center, etc., parameters and metadata in columns). We conducted careful quality-control checks on each of the selected seven parameters and excluded obviously invalid values, which was only the case for VO_2 and metabolic rate.

VO_2 and metabolic rate parameters. For VO_2 and metabolic rate parameters, we used the body weight-independent residuals since body mass is the major determinant for variability in absolute VO_2 and metabolic rate. For this, we calculated a linear model with mean body mass (mean individual mass before and after the calorimetry test) as predictor and VO_2 , respectively, metabolic rate as response variable separate for each phenotyping center and for each sex. The residuals of these models represent the difference between each individual's actual VO_2 /metabolic rate and the response value predicted by their mean body weight. By adding a constant (e.g., the predicted value for the mean body weight) to the residual, the sense of an actual VO_2 /metabolic rate value is conveyed. The new residual response variable is finally uncorrelated with the mean body weight⁴⁷.

General statistics. The number of mutant strains assessed for each parameter varied depending upon (1) center-wide test implementation; and (2) whether both males and females were phenotyped; e.g., indirect calorimetry providing VO_2 , MR, and RER was not conducted in every center and not always using both sexes. The statistical power of the phenotyping approach implemented by the IMPC alliance

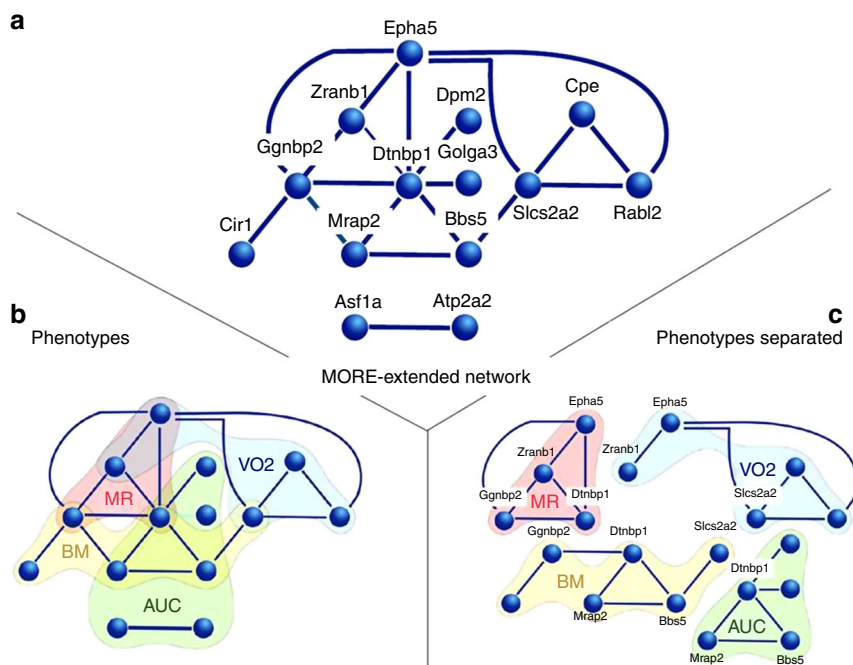


Fig. 5 MORE set-derived network of the 20 genes having two or more phenotypic associations. Genes were selected from the strong metabolic phenotype list with two or more phenotypic associations to metabolic traits. Genes were chosen as targets to examine shared links to regulatory networks. **a** Connection of 12 genes and 2 other genes by MORE sets found in their promoters. **b** The MORE-derived network from **a** with gene-phenotype associations superimposed as colored areas. This superposition joins the 14 genes to one network. **c** Four phenotypes are shown separated from each other along with the links between these strong metabolic phenotype genes. This will facilitate recognition of the individual phenotypes

was evaluated by data obtained from the pilot program EUMODIC²¹. Here a sample size of seven mutant mice per sex was found to be required to detect genotype effects. This is the outcome of a trade-off between sufficient statistical power and technical and workflow constraints, which are linked in any high-throughput phenotyping screen. As described above mean mutant/wild-type ratios were calculated for all parameters split for phenotyping center and sex to identify universal cross-project metabolic characteristics of knockouts. No further quantitative statistical measure was assigned to these pragmatic but restrictive selection criteria. However, critical values for the 5% tails could be computed by upper/lower threshold = mean \pm 1.645 s.d., approximating a significance value of $p < 0.05$ for each tail or $p < 0.1$ for both higher and lower tails together.

Permutation analysis. We performed permutation tests in order to estimate p values for the fraction of strong metabolic phenotype genes within the set of 101 IMPC genes with documented links to obesity and type 2 diabetes. In particular, we shuffled the phenotype ratios of the gene set analyzed by IMPC one million times and then we obtained sample distributions of the fraction of genes within the <5th >95th and the <20th and >80th range percentiles of the 101 genes in these random sets. These distributions were subsequently used to obtain the p values.

Identification of unexplored genes. Starting with all genes with strong metabolic phenotype of the IMPC project (IMPC Release 4.2 from Dec 2015), several filters were applied to discard genes with known molecular or functional annotation. First, we removed genes linked to mammalian phenotype (MP) terms related to metabolism (as described in the next section “Mapping to MP ontology terms”). Next, genes with annotated information in KEGG or gene ontology (GO) databases were removed following the procedure described in “Representation of genes in KEGG graphical pathways” and “Selection of GO terms describing metabolic processes” Method sections, respectively. Literature information on the function of the resulting list of genes was inspected using PubMed, LitInspector⁴⁸, and GeneCards⁴⁹. This iterative process produced a list of 51 genes with no information. The additional 20 genes represent those genes with known molecular information (based on GO molecular properties) but unknown biological processes (See Supplement Fig. 1).

Representation of genes in KEGG graphical pathways. The Interactive Pathways Explorer v2 web-based tool⁵⁰ was used to visualize the strong metabolic phenotype genes on a map representing global metabolism in mice. To that aim, we mapped the strong metabolic phenotype and “non-strong” metabolic phenotype genes for each parameter and sex to the global “Metabolic pathways” overview map from *Mus musculus* organism constructed using 146 KEGG pathways.

To get an overview of the representation of the strong metabolic phenotype genes of each parameter and sex in biological pathways, we mapped the genes to the graphical KEGG pathways provided by KEGG online website. In these pathway maps, several functionally related genes might be grouped in the same node. When strong metabolic phenotype and “normal” genes of the same parameter and sex mapped to the same group, we highlighted the strong metabolic phenotype genes.

To link genes to KEGG pathways, the KEGG website search tool (http://www.genome.jp/kegg/tool/map_pathway1.html) was used applying the filter “organism: *Mus musculus*.” The information available from the page was then downloaded and processed using Bash shell and R scripts (R version 3.3.2). To identify those genes linked to metabolism, we selected those mapping to the “Metabolism” class of KEGG classification.

Selection of GO terms describing metabolic processes. Associations of genes to GO terms were extracted from Gene Ontology Consortium website⁵¹. We analyzed terms of the “Molecular function” and “Biological processes” domains. To select genes involved in metabolism based on GO term annotation, we selected all GO terms under the “metabolic process” category and mapped them to the genes of interest.

Mapping to MP ontology terms. MP terms for the genes of interest were extracted from the MGI database (<http://www.informatics.jax.org/>) using the file (MGI_PhenoGenoMP.rpt), which contains information on genes and their annotated phenotypes. In order to avoid circular discoveries, as the MGI database includes the IMPC phenotype data, IMPC entries were removed using internal filters.

Human subjects, GWAS, and SNP analysis and statistics. Orthologous/paralogous genes of mouse metabolism genes mapping to the human genome were used for analysis. We searched for SNPs in a ± 2 kb region in SNIQA database. For each SNP occurring in or around genes, we evaluated the extent of this sharing for 19238 SNPs in 16 metabolic phenotype GWAs from various consortia, including DIAGRAM, MAGIC, GIANT, GLGC, and ICBP^{30–36}. All participants of the studies contributing to the consortia have given their informed consent for genetic studies, which was confirmed by the appropriate ethics committees. For our present study, we used metadata without any link to individual IDs or data only. We used cross phenotype meta-analysis (CPMA), which detects association of a SNP to multiple, but not necessarily all, phenotypes³⁷. The CPMA analysis applies the likelihood ratio test that measures the likelihood of the null hypothesis (i.e., that the significant SNP is uniformly distributed across consortiums) over the alternative hypothesis.

The ongoing Tübingen family (TÜF) study for type 2 diabetes currently comprises more than 3000 unrelated non-diabetic Caucasian individuals at increased risk of type 2 diabetes (subjects with family history of diabetes, BMI ≥ 27 , impaired fasting glycemia, and/or previous gestational diabetes). The participants were comprehensively characterized by anthropometrics, five-point oral glucose tolerance tests, and clinical chemistry⁴¹ and were genotyped on Illumina's Infinium® Global Screening Array-24 v1.0 BeadChip, which was developed based on Phase-III data of the 1000 Genomes Project and which has 700,078 single-nucleotide polymorphisms (SNPs). The study followed the principles laid down in the Declaration of Helsinki and was approved by the Ethics Committee of the University of Tübingen. Informed written consent was obtained from all participants. From the TÜF study, we selected 2788 subjects with complete phenotypic data sets (body fat content/distribution, blood glucose, insulin sensitivity, insulin secretion, and plasma lipids) as the study population for the analysis of the human orthologues of unexplored murine metabolic genes. Of these 51 murine genes, four (*1500011B03Rik*, *4930591A17Rik*, *Dppa1*, and *Cldn34b2*) had no human orthologues, and three orthologues (*PRAMEF20*, *C17orf105*, and *TMEM42*) were not on the array. In the remaining 44 orthologues, only SNPs with minor allele frequencies ≥ 0.05 were analyzed, due to statistical power limitations of the study population. In consequence, seven orthologues (*C5orf52*, *CCDC24*,

CCDC116, *FBXW12*, *TMEM136*, *DNAJC5G*, and *TEX37*) with no common SNPs were excluded from the analyses. In the remaining 37 genes, 240 common, bi-allelic, and non-linked ($r < 0.8$) SNPs with call rates ≥ 0.75 were identified and ultimately analyzed. Analysis of association with the aforementioned phenotypes was carried out by multiple linear regression analysis (least squares method) to account for potential confounders (gender, age, BMI, and insulin sensitivity) whenever appropriate. The SNPs were analyzed in the additive inheritance model. According to Bonferroni correction for the number of SNPs tested in parallel, p values < 0.000213 were considered significant. Associations were indicated as nominal if p values were ≥ 0.000213 and < 0.05 .

MORE cassette enrichment analysis (promoter analysis). Regarding promoter selection, sets of genes were collected from genes in the 5% outlier range (sex, high and low outliers were separately analyzed). The promoters of these gene sets were collected with the ElDorado database and the program Gene2Promoter (both Genomatix, Munich).

Multiple organized regulatory elements (MORE) form a (partial) "fingerprint" that is present within a group of regulatory regions (promoters, enhancers, etc.). Transcriptional MORE cassettes use transcription factor-binding sites (TFBSs) as elements, which are defined by a weight-matrix-based detection method (MatInspector, Genomatix, Munich).

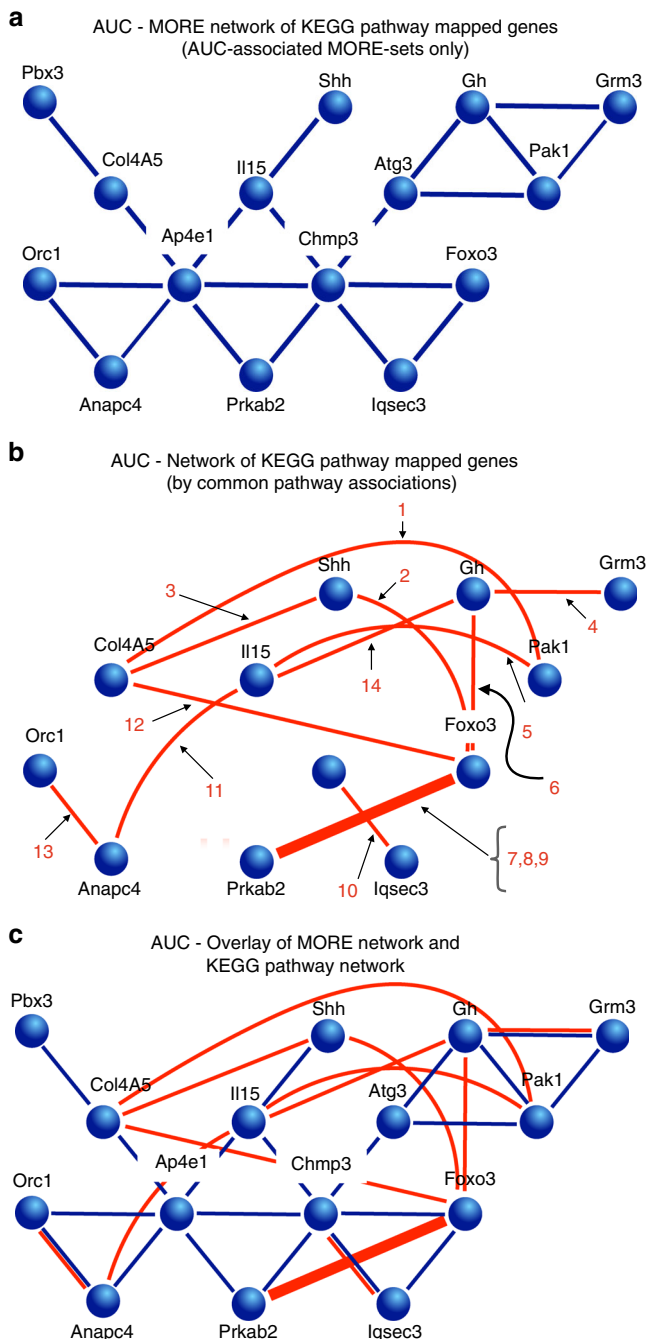
A MORE cassette is defined by several individual TFBSs and corresponding detection thresholds, their order and strand orientation in the DNA sequence, and the distance ranges and distance variations between pairs of TFBS elements (See Supplementary Notes 1 and 2 for more details).

Since there are too many potential MORE cassettes to collect in a database for standard enrichment analysis (see Supplementary Note 1 for details), we determined the number and structures of the MORE cassettes actually present in at least three promoters of each set. Each promoter set was analyzed separately for the presence of MORE cassettes shared by at least three promoters per set using the program FrameWorker (Genomatix, Munich). Parameters were set according to the supplier's defaults, except for Sequence quorum which was adapted top-down until MORE cassettes were found. The distance range variation was set to 20 bp, the minimum distance between elements was set to 10 and the minimum range of elements in MORE cassettes was set to 3–6. No further adjustments were made unless the minimum setting of three sequences did not reveal any MORE cassettes. In this case, the distance variation was increased to 30. Promoter sets that still had no MORE cassettes with at least three elements were considered negative.

To define MORE cassette sets following procedure was applied: In cases where multiple MORE cassettes were detected, a manual alignment of the MORE cassettes was carried out and all MORE cassettes that were identical in all elements and in their order were collected into a MORE cassette set (only varying in the distance definitions). In cases of four or more elements, MORE cassettes were also collected into one bin if they differed only in one element at the exact same relative position within the MORE (e.g., A-B-C-D and A-B-F-D), however, in such cases the overlap of the promoters harboring one of these MORE cassettes was required to exceed 80%. The resulting MORE cassette sets were treated as if they were individual MORE cassettes. However, detection in promoters by ModelInspector was always based on actual individual MORE cassettes.

All MORE cassettes and the respective sets were located in promoter sequences by the program ModelInspector (Genomatix, Munich) in the following promoter sets: the set initially used for the detection of these MORE cassettes (to detect additional matches missed in the detection process, e.g., located on the opposite strand), the promoter set derived from the opposite outlier group (plus or minus, respectively), the corresponding promoter sets from the other sex, and finally the set of all promoters from the mouse genome as provided in the ElDorado database.

Fig. 6 The joint MORE set and KEGG pathway network. Fifteen genes from the gene list AUC female impaired (KEGG) for which MORE set connections were also found. Male impaired did not yield results. **a** AUC gene network derived from KEGG pathway-mapped genes showing AUC-associated MORE sets only. **b** AUC gene network of KEGG pathway-mapped strong knockout genes by common pathway associations. **c** AUC overlay of MORE network and KEGG pathway network. (Legend extension: 1 = Focal adhesion—*Mus musculus* (mouse), 2 = JAK-STAT signaling pathway—*Mus musculus* (mouse), 3 = Pathways in cancer—*Mus musculus* (mouse), 4 = Neuroactive ligand-receptor interaction—*Mus musculus* (mouse), 5 = Chemokine signaling pathway—*Mus musculus* (mouse), 6 = PI3K-AKT signaling pathway—*Mus musculus* (mouse), 7 = FoxO signaling pathway—*Mus musculus* (mouse), 8 = AMPK signaling pathway—*Mus musculus* (mouse), 9 = Longevity regulating pathway—*Mus musculus* (mouse), 10 = Endocytosis—*Mus musculus* (mouse), 11 = HTLV-I infection—*Mus musculus* (mouse), 12 = PI3K-AKT signaling pathway—*Mus musculus* (mouse), 13 = Cell cycle—*Mus musculus* (mouse), 14 = Basal cell carcinoma—*Mus musculus* (mouse))



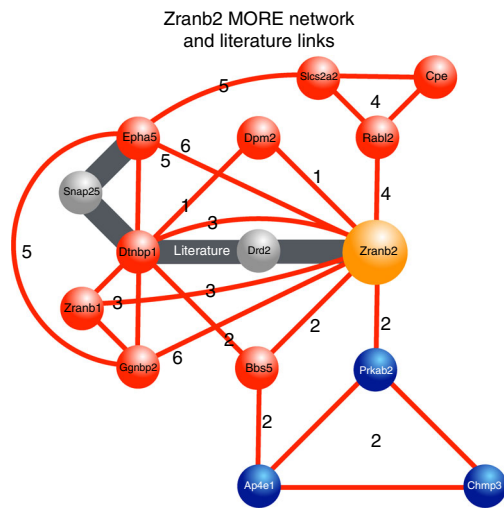


Fig. 7 *Zranb2* MORE network. Validation of predicted gene functions based on six shared MORE sets and functional links from literature (further explanation see text)

To analyze enrichment of MORE cassettes, we applied following procedures. From the total number of promoters containing a particular MORE cassette or MORE cassette set, as determined above, an expectation value was determined for a random subset of the same size as the test set (e.g., the original promoter set) and an over-representation against an expected random draw of a subset of the size of the sub-phenotype (male, female, up or down) gene promoters, comparing the promoter set of interest to all of the mouse promoters (and the total number of MORE sets actually present in all promoters). Any over-representation ≥ 2 was regarded as indication of an association of the MORE cassette or MORE cassette set with the specific promoter set analyzed. Enrichment was also analyzed in the set of genes with the corresponding opposite phenotype as a control (e.g., for female low, female high was checked).

To determine regulatory connections between phenotypes, all promoters from all genes selected for all phenotypes were analyzed for presence of each enriched MORE cassette or MORE set as determined from that phenotype. This analysis was carried out across all phenotypes for both sexes and regardless of the high/low differentiation. Genes with any of the enriched MORE cassettes or MORE sets in at least one of their promoters were collected. In this case, enrichment with phenotypes other than the initial one from which the MORE cassettes were determined was not required.

The list of the 20 most interesting genes was checked for the presence of associated MORE cassettes in their promoters. Then a network was constructed in which genes associated with the same MORE sets in the same sub-phenotype (male, female, up or down) were “connected” via this shared MORE structure in their promoters. The resulting network was superimposed onto closed areas that represent the respective phenotype, the MORE sets or the genes they were associated with.

Ethical approval. All details regarding animal ethics approval of mouse production, breeding and phenotyping at each center are provided in Supplementary Table 3. All procedures were conducted in compliance with each center’s ethical animal care and use guidelines. All procedures were in accordance with the respective national legislation. In addition, we confirm that all efforts were made to minimize suffering by considerate housing and husbandry. Animal welfare was assessed routinely for all mice.

Data availability. All phenotyping data and mouse lines presented in this paper are openly available from the IMPC portal and via our FTP site (<ftp://ftp.ebi.ac.uk/pub/databases/impc/latest/>). Information on strong metabolic phenotype gene, GWAS results and network analysis are provided in Supplementary Data files. The complete list of MORE cassette descriptions is available from the authors, consisting of the salient features: individual TFBS elements, their order, strand orientation, distance range and distance range variations between the pairs. This allows location of the MORE cassettes, using any suitable TFBS program, without requiring access to the Genomatix Suite—for individual promoters. The list of MORE sets, consisting of the MORE cassettes belonging to each set is also available.

Received: 24 November 2016 Accepted: 30 October 2017

Published online: 18 January 2018

References

- Ahmed, M. Non-alcoholic fatty liver disease in 2015. *World J. Hepatol.* **7**, 1450–1459 (2015).
- Boehme, M. W. et al. Prevalence, incidence and concomitant co-morbidities of type 2 diabetes mellitus in South Western Germany—a retrospective cohort and case control study in claims data of a large statutory health insurance. *BMC Public Health* **15**, 855 (2015).
- Forouhi, N. G. & Wareham, N. J. Epidemiology of diabetes. *Medicine* **42**, 698–702 (2014).
- Kharroubi, A. T. & Darwish, H. M. Diabetes mellitus: the epidemic of the century. *World J. Diabetes* **6**, 850–867 (2015).
- Stevens, G. A. et al. National, regional, and global trends in adult overweight and obesity prevalences. *Popul. Health Metr.* **10**, 22 (2012).
- Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
- Hattersley, A. T. & Patel, K. A. Precision diabetes: learning from monogenic diabetes. *Diabetologia* **60**, 769–777 (2017).
- Kraja, A. T. et al. Pleiotropic genes for metabolic syndrome and inflammation. *Mol. Genet. Metab.* **112**, 317–338 (2014).
- Kunes, J. et al. Epigenetics and a new look on metabolic syndrome. *Physiol. Res.* **64**, 611–620 (2015).
- Mamtani, M. et al. Genome- and epigenome-wide association study of hypertriglyceridemic waist in Mexican American families. *Clin. Epigenetics* **8**, 6 (2016).
- Somer, R. A. & Thummel, C. S. Epigenetic inheritance of metabolic state. *Curr. Opin. Genet. Dev.* **27**, 43–47 (2014).
- Pandey, A. K. et al. Functionally enigmatic genes: a case study of the brain ignorome. *PLoS ONE* **9**, e88889 (2014).
- Sahni, N. et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660 (2015).
- Steckler, T. et al. The preclinical data forum network: a new ECNP initiative to improve data quality and robustness for (preclinical) neuroscience. *Eur. Neuropsychopharmacol.* **25**, 1803–1807 (2015).
- Brown, S. D. & Moore, M. W. The international mouse phenotyping consortium: past and future perspectives on mouse phenotyping. *Mamm. Genome* **23**, 632–640 (2012).
- Ring, N. et al. A mouse informatics platform for phenotypic and translational discovery. *Mamm. Genome* **26**, 413–421 (2015).
- Gailus-Durner, V. et al. Introducing the German mouse clinic: open access platform for standardized phenotyping. *Nat. Methods* **2**, 403–404 (2005).
- Mallon, A. M., Blake, A. & Hancock, J. M. EuroPhenome and EMPReSS: online mouse phenotyping resource. *Nucleic Acids Res.* **36**, D715–D718 (2008).
- Meehan, T. F. et al. Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat. Genet.* **49**, 1231–1238 (2017).
- Bowl, M. R. et al. A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nat. Commun.* **8**, 886 (2017).
- Hrade de Angelis, M. et al. Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nat. Genet.* **47**, 969–978 (2015).
- Karp, N. A. et al. Applying the ARRIVE Guidelines to an in vivo database. *PLoS Biol.* **13**, e1002151 (2015).
- Brommage, R. et al. High-throughput screening of mouse gene knockouts identifies established and novel skeletal phenotypes. *Bone Res.* **2**, 14034 (2014).
- Karp, N. A. et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat. Commun.* **8**, 15475 (2017).
- Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nat. Rev. Genet.* **9**, 911–922 (2008).
- Bonnefond, A. & Froguel, P. Rare and common genetic events in type 2 diabetes: what should biologists know?. *Cell Metab.* **21**, 357–368 (2015).
- Dauriz, M. et al. Association of a 62 variants type 2 diabetes genetic risk score with markers of subclinical atherosclerosis: a transethnic, multicenter study. *Circ. Cardiovasc. Genet.* **8**, 507–515 (2015).
- Hara, K., Kadowaki, T. & Odawara, M. Genes associated with diabetes: potential for novel therapeutic targets?. *Expert. Opin. Ther. Targets* **20**, 255–267 (2016).
- Vimaleswaran, K. S. et al. Candidate genes for obesity-susceptibility show enriched association within a large genome-wide association study for BMI. *Hum. Mol. Genet.* **21**, 4537–4542 (2012).
- Arnold, M. et al. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334–1336 (2015).
- Cotsapas, C. et al. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7**, e1002254 (2011).
- Ehret, G. B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
- Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

34. DIAGRAM Consortium 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).
35. Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycaemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
36. Manning, A. K. et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet. Epidemiol.* **35**, 11–18 (2011).
37. Scott, R. A. et al. Large-scale association analyses identify new loci influencing glycaemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
38. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
39. Soranzo, N. et al. Common variants at 10 genomic loci influence hemoglobin A (1)(C) levels via glycaemic and nonglycaemic pathways. *Diabetes* **59**, 3229–3239 (2010).
40. Global Lipids Genetics, C. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
41. Stefan, N. et al. Polymorphisms in the gene encoding adiponectin receptor 1 are associated with insulin resistance and high liver fat. *Diabetologia* **48**, 2282–2291 (2005).
42. Kanehisa, M. et al. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
43. Carroll, L. S. et al. Evidence that putative ADHD low risk alleles at SNAP25 may increase the risk of schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **150B**, 893–899 (2009).
44. Cohen, O. S. et al. A splicing-regulatory polymorphism in DRD2 disrupts ZRANB2 binding, impairs cognitive functioning and increases risk for schizophrenia in six Han Chinese samples. *Mol. Psychiatry* **21**, 975–982 (2016).
45. Jia, J. M. et al. Age-dependent regulation of synaptic connections by dopamine D2 receptors. *Nat. Neurosci.* **16**, 1627–1636 (2013).
46. Karp, N. A. et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat. Commun.* **8**, 15475 (2017).
47. Willett, W. C., Howe, G. R. & Kushi, L. H. Adjustment for total energy intake in epidemiologic studies. *Am. J. Clin. Nutr.* **65**, 1220S–1228S (1997).
48. Frisch, M. et al. LitInspector: literature and signal transduction pathway mining in PubMed abstracts. *Nucleic Acids Res.* **37**, W135–W140 (2009).
49. Stelzer, G. et al. The genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* **54**, 1 30 1–1 30 33 (2016).
50. Yamada, T. et al. iPath2.0: interactive pathway explorer. *Nucleic Acids Res.* **39**, W412–W415 (2011).
51. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research: Infrafrontier [no. 01KX1012] (M.HdA.), the German Center for Diabetes Research (DZD), EU Horizon2020: IPAD-MD [no. 653961] (M.HdA.), by NIH grants U54 HG006370 (T.F.M., P.F., A.-M.M., H.E.P., D.S., and S.D.M.B.), U42 OD011185 (S.A.M.), UM1 OD023222 (S.A.M., K.L.S., and R.E.B.), U54 HG006332 (R.E.B. and K.S.), U54 HG006348-S1 and OD011174 (A.L.B.), 1R24OD011883 (C.J.M.), U54 HG006364, and U42 OD011175, and UM1 OD023221 (K.C.K.L.), and additional support provided by the The Wellcome Trust, Medical Research Council Strategic Award 53658 (S.W. and S.D.M.B.), Government of Canada through Genome Canada and Ontario Genomics (OGI-051) (C.M. and S.D.M.B.), Wellcome Trust Strategic Award, National Centre for Scientific Research (CNRS), the French National Institute of Health and Medical Research (INSERM), the University of Strasbourg (UDS), the “Centre Européen de Recherche en Biologie et en Médecine,” the “Agence Nationale de la Recherche” under the frame program “Investissements d’Avenir” labeled ANR-10-IDEX-0002-02, ANR-10-INBS-07 PHENOMIN to (Y.H.), “EUCOMM: Tools for Functional Annotation of the Mouse Genome” (EUCOMMTOOLS) project—grant agreement no. [FP7-HEALTH-F4-2010-261492] (W.G.W.), the Government of Australia through the National Collaborative Research Infrastructure Strategy to the Australian Phenomics Network Project (M.S.D.)

and by the Grants for Korea Mouse Phenotype Consortium (2013M3A9D5072550) funded by National Research Foundation, Korean Government (J.K.S.). Regarding the NIH grants listed: “The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.”

Author contributions

J.A.A.-P., M.O., C.S.-W., T.St., A.Ö.Y., O.Ei., L.B., T.K., I.T., D.H.B., J.B., R.Be., Kr.M., L. G., S.M.H., Ann.Z., O.V.A., J.G., F.N., J.C.-W., L.Br., A.Z., R.B., J.R., B.R., W.W., M.K., E. W., M.Wu., V.G.-D., H.F., M.A.Ö., Ch.S., Ch.L., L.Z., Y.-L.C., P.S.-B., M.J.K., P.M.-K., C. S., K.K.G., L.S., T.F., N.R., H.W., S.G., D.S., H.Mo., S.A.Ma., D.L., V.E.V., S.A.P., A.S.G., M.S., C.Sh., L.F.E.A., M.T.S., R.S.B.Mc.L., A.S., C.M.I., E.L.C., H.M.W., S.S.C., A.K.B.M., J.D., A.G., A.O.S., C.J.L., S.Wa., T.Su., M.T., H.K., T.Fu., K.K., I.M., I.Y., N.T., D.W., and K.L.S. obtained phenotyping and performed data QC. Ch.L., M.A.Ö., A.Ga., E.M., J.S., I. L., K.M., R.Bo., F.C., C.D.P., G.D.S., O.E., F.F., P.Fr., A.I.G., S.Ga., E.G., G.L.S., S.Ma., D. Ma., M.M., R.M., T.O., M.Pa., M.Ra., A.R., G.R., N.Ro., S.Pu., F.S., G.D.T.-V., H.L.C., H. P.N., J.C., S.C.-M., Z.S.-K., A.P.W., S.J.J., M.E.S., J.N.C., K.C., M.N.D.G., L.K., I.V., Z.B., C.O., D.Q., R.G., S.N., L.M., N.L., X.S., P.Fe., Y.W., M.E., Y.Z., J.H., G.C., M.B., D.M., K. So., H.A., M.Ga., A.Be., S.Y.C., J.A.W., D.C., P.H., T.T., H.T., M.Sch., S.P., A.T., E.D., M. P., A.Hu., V.D.V., I.Mo., F.B., N.C., J.W., C.-K.C., I.Tu., M.R., P.M., L.P., S.R., A.A., P.R., A.Ay., M.Se., and J.O.J. acquired and processed data. C.Sei., A.B., J.Bei., R.K., J.Sch., A.H., O.O., F.G., W.W., S.M., G.F.C., J.L., L.M.T., M.D.F., B.D., H.W.-J., J.Bo., E.R., D.G., P.P., V.L., S.C., Q.L., G.S., A.C., E.J., O.D., M.G., M.Pe., S.Mac.M., S.T., T.C., J.Ca., A.Y., S.A., H.-C.K., H.L., and S.A.M. planned and performed mouse production. T.W. and Ch.S. analyzed the MORE cassettes data set. S.Sh., F.M., A.P., H.S., H.-U.H., and H.G. analyzed the SNP data set. H.M., Ch.S., M.A.Ö., J.R., B.R., M.C., A.C.R., M.Ki., S.L., T.F.M., and J. M. analyzed the IMPC data set. J.R., B.R., M.A.Ö., Ch.S., A.C.R., S.L., S.Sh., M.Ki., M.W., R.B., M.H.T., W.W., M.K., E.W., J.B., F.M., A.P., H.S., H.-U.H., H.G., M.C., H.M., H.F., V.G.-D., T.W., M.HdA., C.L.R., A.L.B., G.P.T.-V., T.H., A.-M.M., S.W., S.D.B., C.J.L., J.K. W., A.M.F., C.Mc.K., L.M.J.N., Hi.M., J.K.S., L.R.B., A.M., B.M., K.C.K.L., S.A.M., K.L.S., R.B.B., T.F.M., J.M., H.H., P.F., H.E.P., T.S., Y.H., M.-F.C., M.S.D., F.B., X.G., C.K.L.W., M.Mo., and D.Sm. drafted and revised the manuscript. J.R., M.C., H.M., H.F., V.G.-D., T. W., M.HdA., H.G., H.-U.H., C.L.R., J.S., M.E.D., G.P.T.-V., S.D.B., T.H., A.-M.M., S.W., R.S., C.J.L., A.Br., W.C.S., R.R.-S., K.S., D.J.A., J.K.W., C.Mc.K., A.M.F., Y.O., D.W., J.K. S., K.C.K.L., M.M., C.K.L.W., P.F., H.E.P., and Y.H. designed and supervised the study.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-017-01995-2>.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as 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 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, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

Jan Rozman^{1,2}, Birgit Rathkolb^{1,2,3}, Manuela A. Oestreich¹, Christine Schütt¹, Aakash Chavan Ravindranath^{2,4}, Stefanie Leuchtenberger¹, Sapna Sharma^{2,5}, Martin Kistler¹, Monja Willershäuser^{6,7,8}, Robert Brommage¹, Terrence F. Meehan⁹, Jeremy Mason⁹, Hamed Haselimashhadi⁹, Tertius Hough¹⁰, Ann-Marie Mallon¹⁰, Sara Wells¹⁰, Luis Santos¹⁰, Christopher J. Lelliott¹¹, Jacqueline K. White^{11,12}, Tania Sorg^{13,14,15,16,17}, Marie-France Champy^{13,14,15,16,17}, Lynette R. Bower¹⁸,

Corey L. Reynolds¹⁹, Ann M. Flenniken^{20,21,22}, Stephen A. Murray¹², Lauryl M.J. Nutter^{20,21}, Karen L. Svenson¹², David West²³, Glauco P. Tocchini-Valentini²⁴, Arthur L. Beaudet^{20,21}, Fatima Bosch²⁵, Robert B. Braun¹², Michael S. Dobbie²⁶, Xiang Gao²⁷, Yann Herault^{13,14,15,16,17}, Ala Moshiri²⁸, Bret A. Moore²⁹, K.C. Kent Lloyd¹⁸, Colin McKerlie^{20,21}, Hiroshi Masuya³⁰, Nobuhiko Tanaka³⁰, Paul Flicek⁹, Helen E. Parkinson⁹, Radislav Sedlacek³¹, Je Kyung Seong³², Chi-Kuang Leo Wang³³, Mark Moore³⁴, Steve D. Brown¹⁰, Matthias H. Tschöp^{2,35,36}, Wolfgang Wurst^{37,38,39,40}, Martin Klingenspor^{6,7,8}, Eckhard Wolf^{2,3}, Johannes Beckers^{1,2,41}, Fausto Machicao⁴², Andreas Peter^{2,42,43}, Harald Staiger^{2,43,44}, Hans-Ulrich Häring^{2,42,43}, Harald Grallert^{2,5,45}, Monica Campillos^{2,4}, Holger Maier¹, Helmut Fuchs¹, Valerie Gailus-Durner¹, Thomas Werner⁴⁶ & Martin Hrabe de Angelis^{1,2,41}

¹German Mouse Clinic, Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. ²German Center for Diabetes Research (DZD), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. ³Ludwig-Maximilians-Universität München, Gene Center, Institute of Molecular Animal Breeding and Biotechnology, Feodor-Lynen Strasse 25, 81377 Munich, Germany. ⁴Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. ⁵Research Unit of Molecular Epidemiology, Institute of Epidemiology II, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. ⁶Chair of Molecular Nutritional Medicine, Technical University of Munich, TUM School of Life Sciences Weihenstephan, 85354 Freising, Germany. ⁷EKFZ - Else Kröner-Fresenius Center for Nutritional Medicine, Technical University of Munich, 85354 Freising, Germany. ⁸ZIEL - Institute for Food & Health, Technical University of Munich, 85354 Freising, Germany. ⁹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ¹⁰Medical Research Council Harwell (Mammalian Genetics Unit and Mary Lyon Centre), Oxfordshire OX11 0RD, UK. ¹¹The Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ¹²The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA. ¹³CELPEDIA, PHENOMIN, Institut Clinique de la Souris (ICS), 1 Rue Laurent Fries, 67404 Illkirch-Graffenstaden, France. ¹⁴Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Parc d'innovation, 1 Rue Laurent Fries - BP 10142, 67404 Illkirch, France. ¹⁵Centre National de la Recherche Scientifique, UMR7104, 67404 Illkirch, France. ¹⁶Institut National de la Santé et de la Recherche Médicale, U964, 67404 Illkirch, France. ¹⁷Université de Strasbourg, 67404 Illkirch, France. ¹⁸Mouse Biology Program, University of California, One Shields Avenue, Davis, CA 95616, USA. ¹⁹Department of Molecular and Human Genetics, Baylor College of Medicine, 7702 Main St, Houston Medical Center, Houston, TX 77030-4406, USA. ²⁰The Centre for Phenogenomics, 25 Orde St, Toronto, M5T 3H7, ON, Canada. ²¹The Hospital for Sick Children, 600 University Avenue, Toronto, ON M5G 1X5, Canada. ²²Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Joseph and Wolf Lebovic Health Complex, 600 University Avenue, Toronto, ON M5G 1X5, Canada. ²³Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA. ²⁴Monterotondo Mouse Clinic, Italian National Research Council (CNR), Institute of Cell Biology and Neurobiology, Adriano Buzzati-Traverso Campus, Via E. Ramarini 32, Monterotondo Scalo, RM 00015, Italy. ²⁵Center of Animal Biotechnology and Gene Therapy and Department of Biochemistry and Molecular Biology, Universitat Autònoma de Barcelona, Bellaterra, Spain. ²⁶Australian Phenomics Network, John Curtin School of Medical Research, Australian National University, 131 Garran Road, Canberra, ACT 2601, Australia. ²⁷SKL of Pharmaceutical Biotechnology and Model Animal Research Center, Collaborative Innovation Center for Genetics and Development, Nanjing Biomedical Research Institute, Nanjing University, Nanjing 210061, China. ²⁸Department of Ophthalmology & Vision Science, School of Medicine, U.C. Davis, 77 Cadillac Drive, Sacramento 95825 CA, USA. ²⁹William R. Pritchard Veterinary Medical Teaching Hospital, School of Veterinary Medicine, U.C. Davis, One Shields Avenue, Davis 95616 CA, USA. ³⁰RIKEN BioResource Center, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan. ³¹Czech Centre for Phenogenomics, Institute of Molecular Genetics, Prumyslova 595, 252 50 Vestec, Czech Republic. ³²Korea Mouse Phenotyping Consortium (KMPC) and BK21 Program for Veterinary Science, Research Institute for Veterinary Science, College of Veterinary Medicine, Seoul National University, 599 Gwanangno, Gwanak-gu, Seoul 151-742, South Korea. ³³National Laboratory Animal Center, National Applied Research Laboratories (NARLabs), 128 Yen-Chiou-Yuan Rd., Sec. 2, Nankang, Taipei 11529, Taiwan. ³⁴IMPC, San Anselmo, CA 94960, USA. ³⁵Institute for Diabetes and Obesity, Helmholtz Diabetes Center at Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), 85764 Neuherberg, Germany. ³⁶Division of Metabolic Diseases, Department of Medicine, Technische Universität München, 80333 Munich, Germany. ³⁷Institute of Developmental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health GmbH, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. ³⁸Chair of Developmental Genetics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. ³⁹Deutsches Institut für Neurodegenerative Erkrankungen (DZNE) Site Munich, Feodor-Lynen-Str. 17, 81377 Munich, Germany. ⁴⁰Munich Cluster for Systems Neurology (SyNergy), Adolf-Butenandt-Institut, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 17, 81377 Munich, Germany. ⁴¹Chair of Experimental Genetics, School of Life Science Weihenstephan, Technische Universität München, Alte Akademie 8, 85354 Freising, Germany. ⁴²Department of Internal Medicine, Division of Endocrinology, Diabetology, Vascular Medicine, Nephrology and Clinical Chemistry, University of Tübingen, 72076 Tübingen, Germany. ⁴³Institute for Diabetes Research and Metabolic Diseases of the Helmholtz Center Munich at the Eberhard-Karls-University of Tübingen, Otfried-Müller-Str. 10, 72076 Tübingen, Germany. ⁴⁴Institute of Pharmaceutical Sciences, Department of Pharmacy and Biochemistry, Eberhard Karls University Tübingen, 72076 Tübingen, Germany. ⁴⁵Clinical Cooperation Group Type 2 Diabetes, Helmholtz Zentrum München and Ludwig-Maximilians Universität München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. ⁴⁶Internal Medicine Nephrology and Center for Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

IMPC Consortium

Antonio Aguilar-Pimentel¹, Lore Becker¹, Irina Treise¹, Kristin Moreth¹, Lillian Garrett^{1,37}, Sabine M. Hölter^{1,37}, Annemarie Zimprich^{1,37}, Susan Marschall¹, Oana V. Amarie^{1,37}, Julia Calzada-Wack¹, Frauke Neff^{1,48}, Laura Brachthäuser^{1,48}, Christoph Lengger¹, Claudia Stoeger¹, Lilly Zapf¹, Yi-Li Cho¹, Patricia da Silva-Buttkus¹

Markus J. Kraiger¹, Philipp Mayer-Kuckuk¹, Karen Kristine Gampe¹, Moya Wu¹, Nathalie Conte⁹, Jonathan Warren⁹, Chao-Kung Chen⁹, Ilinca Tudose⁹, Mike Relac⁹, Peter Matthews⁹, Heather L. Cater¹⁰, Helen P. Natukunda¹⁰, James Cleak¹⁰, Lydia M. Teboul¹⁰, Sharon Clementson-Mobbs¹⁰, Zsombor Szoke-Kovacs¹⁰, Alison P. Walling¹⁰, Sara J. Johnson¹⁰, Gemma F. Codner¹⁰, Tanja Fiegel¹⁰, Natalie Ring¹⁰, Henrik Westerberg¹⁰, Simon Greenaway¹⁰, Duncan Sneddon¹⁰, Hugh Morgan¹⁰, Jorik Loeffler¹⁰, Michelle E. Stewart¹⁰, Ramiro Ramirez-Solis¹¹, Allan Bradley¹¹, William C. Skarnes¹¹, Karen P. Steel¹¹, Simon A. Maguire¹¹, Joshua Dench¹¹, David Lafont¹¹, Valerie E. Vancollie¹¹, Selina A. Pearson¹¹, Amy S. Gates¹¹, Mark Sanderson¹¹, Carl Shannon¹¹, Lauren F.E. Anthony¹¹, Maksymilian T. Sumowski¹¹, Robbie S.B. McLaren¹¹, Brendan Doe¹¹, Hannah Wardle-Jones¹¹, Mark N.D. Griffiths¹¹, Antonella Galli¹¹, Agnieszka Swiatkowska¹¹, Christopher M. Isherwood¹¹, Anneliese O. Speak¹¹, Emma L. Cambridge¹¹, Heather M. Wilson¹¹, Susana S. Caetano¹¹, Anna Karin B. Maguire¹¹, David J. Adams¹¹, Joanna Bottomley¹¹, Ed Ryder¹¹, Diane Gleeson¹¹, Laurent Pouilly^{13,14,15,16,17}, Stephane Rousseau^{13,14,15,16,17}, Aurélie Auburtin^{13,14,15,16,17}, Patrick Reilly^{13,14,15,16,17}, Abdel Ayadi^{13,14,15,16,17}, Mohammed Selloum^{13,14,15,16,17}, Joshua A. Wood¹⁸, Dave Clary¹⁸, Peter Havel¹⁸, Todd Tolentino¹⁸, Heather Tolentino¹⁸, Mike Schuchbauer¹⁸, Sheryl Pedroia¹⁸, Amanda Trainor¹⁸, Esi Djan¹⁸, Milton Pham¹⁸, Alison Huynh¹⁸, Vincent De Vera¹⁸, John Seavitt¹⁹, Juan Gallegos¹⁹, Arturo Garza¹⁹, Elise Mangin¹⁹, Joel Senderstrom¹⁹, Iride Lazo¹⁹, Kate Mowrey¹⁹, Ritu Bohat¹⁹, Rodney Samaco¹⁹, Surabi Veeraragavan¹⁹, Christine Beeton¹⁹, Sowmya Kalaga¹⁹, Lois Kelsey^{20,21}, Igor Vukobradovic^{20,21}, Zorana Berberovic^{20,21}, Celeste Owen^{20,21}, Dawei Qu^{20,21}, Ruolin Guo^{20,21}, Susan Newbigging^{20,21}, Lily Morikawa^{20,21}, Napoleon Law^{20,21}, Xueyuan Shang^{20,21}, Patricia Feugas^{20,21}, Yanchun Wang^{20,21}, Mohammad Eskandarian^{20,21}, Yingchun Zhu^{20,21}, Patricia Penton^{20,21}, Valerie Laurin^{20,21}, Shannon Clarke^{20,21}, Qing Lan^{20,21}, Gillian Sleep^{20,21}, Amie Creighton^{20,21}, Elsa Jacob^{20,21}, Ozge Danisment^{20,21}, Marina Gertsenstein^{20,21}, Monica Pereira^{20,21}, Suzanne MacMaster^{20,21}, Sandra Tondat^{20,21}, Tracy Carroll^{20,21}, Jorge Cabezas^{20,21}, Jane Hunter^{20,21}, Greg Clark^{20,21}, Mohammed Bubshait^{20,21}, David Miller^{20,21}, Khondoker Sohel^{20,21}, Hibret Adissu^{20,21}, Milan Ganguly^{20,21}, Alexandr Bezginov^{20,21}, Francesco Chiani²⁴, Chiara Di Pietro²⁴, Gianfranco Di Segni²⁴, Olga Ermakova²⁴, Filomena Ferrara²⁴, Paolo Fruscoloni²⁴, Aalessia Gambadoro²⁴, Serena Gastaldi²⁴, Elisabetta Golini²⁴, Gina La Sala²⁴, Silvia Mandillo²⁴, Daniela Marazziti²⁴, Marzia Massimi²⁴, Raffaele Matteoni²⁴, Tiziana Orsini²⁴, Miriam Pasquini²⁴, Marcello Raspa²⁴, Aline Rauch²⁴, Gianfranco Rossi²⁴, Nicoletta Rossi²⁴, Sabrina Putti²⁴, Ferdinando Scavizzi²⁴, Giuseppe D. Tocchini-Valentini²⁴, Shigeharu Wakana³⁰, Tomohiro Suzuki³⁰, Masaru Tamura³⁰, Hideki Kaneda³⁰, Tamio Furuse³⁰, Kimio Kobayashi³⁰, Ikuo Miura³⁰, Ikuko Yamada³⁰, Yuichi Obata³⁰, Atsushi Yoshiki³⁰, Shinya Ayabe³⁰, J. Nicole Chambers³¹, Karel Chalupsky³¹, Claudia Seisenberger³⁷, Antje Bürger³⁷, Joachim Beig³⁷, Ralf Kühn^{37,49,50}, Andreas Hörlein³⁷, Joel Schick^{37,51}, Oskar Ortiz³⁷, Florian Giesert³⁷, Jochen Graw³⁷, Markus Ollert^{52,53}, Carsten Schmidt-Weber⁵⁴, Tobias Stoeger⁵⁵, Ali Önder Yildirim⁵⁵, Oliver Eickelberg⁵⁵, Thomas Klopstock^{39,40,56,57}, Dirk H. Busch⁵⁸, Raffi Bekerredjian⁵⁹, Andreas Zimmer⁶⁰, Jules O. Jacobsen⁶¹, Damian Smedley⁶¹, Mary E. Dickinson⁶², Frank Benso⁶³, Iva Morse⁶³, Hyoung-Chin Kim⁶⁴, Ho Lee⁶⁵ & Soo Young Cho⁶⁵

⁴⁸Institute of Pathology, Helmholtz Zentrum München, German Research Center for Environmental Health GmbH, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. ⁴⁹Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Str. 10, 13125 Berlin, Germany. ⁵⁰Berlin Institute of Health, Anna-Louisa-Karsch Str. 2, 10178 Berlin, Germany. ⁵¹Helmholtz Zentrum Munich, Institute of Molecular Toxicology and Pharmacology, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. ⁵²Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette 4354, Luxembourg. ⁵³Department of Dermatology and Allergy Center, Odense Research Center for Anaphylaxis, University of Southern Denmark, 5000 Odense, Denmark. ⁵⁴Center of Allergy & Environment (ZAUM), Technische Universität München, and Helmholtz Zentrum München, Ingolstädter Landstrasse, 85764 Neuherberg, Germany. ⁵⁵Comprehensive Pneumology Center, Institute of Lung Biology and Disease, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH) and Member of the German Center for Lung Research, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. ⁵⁶Department of Neurology, Friedrich-Baur-Institut, Ludwig-Maximilians-Universität München, Ziemssenstrasse 1a, 80336 Munich, Germany. ⁵⁷German Center for Vertigo and Balance Disorders, University Hospital Munich, Campus

Grosshadern, Marchioninistraße 15, 81377 Munich, Germany. ⁵⁸Institute for Medical Microbiology, Immunology and Hygiene, Technical University of Munich, Trogerstrasse 30, 81675 Munich, Germany. ⁵⁹Department of Cardiology, University of Heidelberg, Im Neuenheimer Feld 410, 69120 Heidelberg, Germany. ⁶⁰Institute of Molecular Psychiatry, Medical Faculty, University of Bonn, Sigmund-Freud-Strasse 25, 53127 Bonn, Germany. ⁶¹Queen Mary University of London, London E1 2AD, UK. ⁶²Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX 77030, USA. ⁶³Charles River Laboratories, Wilmington, MA 01887, USA. ⁶⁴Korea Research Institute of Bioscience and Biotechnology, 30 Yeongudanji-ro, Ochang-eup, Cheongwon-gu, Cheongju-si, Chungcheongbuk-do 28116, Korea. ⁶⁵National Cancer Center, 323 Ilsan-ro, Ilsandong-gu, Koyang 410-769, Korea.