Leveraging Cross-Species Transcription Factor Binding Site Patterns
Claussnitzer, Melina; Dankel, Simon N.; Klocke, Bernward; Grallert, Harald; Glunk, Viktoria; Berulava, Tea; Lee, Heekyoung; Oskolkov, Nikolay; Fadista, Joao; Ehlers, Kerstin

Published in:
Cell

DOI:
10.1016/j.cell.2013.10.058

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

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

Publication date:
2014

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

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

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

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

Download date: 28-05-2021
Leveraging Cross-Species Transcription Factor Binding Site Patterns: From Diabetes Risk Loci to Disease Mechanisms

Melina Claussnitzer, Simon N. Dankei, Bernward Klocke, Harald Grallert, Viktoria Glunk, Christoph Hoffmann, Kun Qian, Tina Rönn, Helene Riese, Martina Müller-Nurasyid, Nancy Bretschneider, Timm Schroeder, Thomas Skurk, Bernhard Horsthemke, DIAGRAM+Consortium, Derek Spieler, Martin Klingenspor, Martin Seifert, Michael J. Kern, Niklas Mejert, Ingrid Dahlman, Ola Hansson, Stefanie M. Hauck, Matthias Blüher, Peter Amer, Leif Groop, Thomas Illig, Karsten Suhre, Yi-Hsiang Hsu, Gunnar Mellgren, Hans Hauner, Helmut Laumen, Chair of Nutritional Medicine, Technische Universität München, Else Kröner-Fresenius-Center for Nutritional Medicine, 85350 Freising-Weihenstephan, Germany.

*Correspondence: melinaclaussnitzer@hsl.harvard.edu (M.C.), helmut.laumen@tum.de (H.L.)

http://dx.doi.org/10.1016/j.cell.2013.10.058
SUMMARY

Genome-wide association studies have revealed numerous risk loci associated with diverse diseases. However, identification of disease-causing variants within association loci remains a major challenge. Divergence in gene expression due to cis-regulatory variants in noncoding regions is central to disease susceptibility. We show that integrative computational analysis of phylogenetic conservation with a complexity assessment of co-occurring transcription factor binding sites (TFBS) can identify cis-regulatory variants and elucidate their mechanistic role in disease. Analysis of established type 2 diabetes risk loci revealed a striking clustering of distinct homeobox TFBS. We identified the PRRX1 homeobox factor as a repressor of PPARG2 expression in adipose cells and demonstrate its adverse effect on lipid metabolism and systemic insulin sensitivity, dependent on the rs4684847 risk allele that triggers PRRX1 binding. Thus, cross-species conservation analysis at the level of co-occurring TFBS provides a valuable contribution to the translation of genetic association signals to disease-related molecular mechanisms.

INTRODUCTION

Recent advances in genome-wide association studies (GWAS) have yielded a plethora of loci associated with diverse human diseases and traits (Hindorff et al., 2009). However, signals emerging from GWAS, which involve typically dozens of variants in linkage disequilibrium (LD), have rarely been traced to the disease-causing variants and even more rarely to the mechanisms by which they may increase disease risk (Califano et al., 2012). The majority of common genetic variants are located in noncoding regions (1000 Genomes Project Consortium et al., 2012), and disease-associated loci are enriched for expression quantitative trait loci (eQTLs) (Nica et al., 2010). DNase I hypersensitive sites sequencing (DHSseq) peaks, and chromatin immunoprecipitation sequencing (ChIP-seq) peaks (Maurano et al., 2012; Encode Project Consortium et al., 2012), suggesting that variants modulating gene regulation are major contributors to common disease risk.

Experimental DHS, RNA, and ChIP-seq approaches have been used to prioritize candidate cis-regulatory variants (Maurano et al., 2012; Encode Project Consortium et al., 2012; Ward and Kellis, 2012b). However, such functional approaches require access to appropriate human tissues and are further hampered by the spatial, temporal, environmental, and epigenetic complexity of gene regulation. These limitations emphasize the need for bioinformatics approaches that reliably assess the regulatory role of noncoding variants. So far, phylogenetic conservation has been a common denominator in the search for noncoding regulatory regions (Waterston et al., 2002; Pennacchio et al., 2006; Encode Project Consortium et al., 2007, 2012; Visel et al., 2009b; Blow et al., 2010; Lindblad-Toh et al., 2011). However, intra- and cross-species differences in gene expression are often driven by changes in transcription factor binding sites (TFBS), and their rapid evolutionary turnover results in lineage-specific regulatory regions that are functionally conserved but have low phylogenetic conservation (Ward and Kellis, 2012a), thus challenging the use of these algorithms. Importantly, gene regulatory regions in eukaryotes tend to be organized in cis-regulatory modules (CRMs), comprising complex patterns of co-occurring TFBS for combinatorial binding of transcription factors (TFs) (Arnone and Davidson, 1997; Pennacchio et al., 2006; Visel et al., 2013). CRMs integrate upstream signals to regulate expression of coordinated gene sets, making them a prime target to achieve phenotypic changes as a result of adaptive evolution (Junion et al., 2012). Despite the critical importance of CRMs, no algorithms have so far been developed to harness the potential power of conserved TFBS patterns within CRMs to predict regulatory variants in disease genetics.

We show that cross-species conservation at the level of the CRMs—rather than at the level of the regulatory sequence that comprises them—identifies cis-regulatory variants within disease-associated GWAS loci. Exploiting phylogenetic conservation of TFBS co-occurrences, we found homeobox TFBS as a cis-regulatory feature of type 2 diabetes (T2D) risk loci, for which the specific causal variants have rarely been pinpointed (Stitzel et al., 2010). Detailed analysis at the PPARG risk locus revealed the rs4684847 risk allele and, by changing binding of the homeobox TF PRRX1, its genotype-dependent effect on PPARG2 expression and insulin sensitivity.

RESULTS

Cross-Species Analysis of TFBS Modularity Discovers cis-Regulatory SNPs at T2D Risk Loci

We developed a method, phylogenetic module complexity analysis (PMCA), which leverages conserved co-occurring TFBS patterns within CRMs to predict cis-regulatory variants, i.e., variants affecting gene expression (Figure 1A; Extended Experimental Procedures available online). To systematically identify cis-regulatory variants at GWAS risk loci, we extracted GWAS tagSNPs and consequently all noncoding (nc) SNPs that are in high LD with these tagSNPs. PMCA individually tests each nc variant by analyzing the flanking region for cross-species conserved TFBS patterns, regardless of global sequence conservation. This requires first the extraction of the region surrounding an nc SNP (±60 bp) from the human genome and consequent identification of orthologous regions in 15 vertebrate species. Within each SNP-specific set of orthologous regions, phylogenetically conserved TFBS, TFBS modules (a cross-species conserved pattern of two or more TFBS occurring in the same order and in a certain distance range), and TFBS in those TFBS modules were identified and then counted. SNP-flanking regions with a significant enrichment of phylogenetically conserved TFBS modules are classified as complex regions, as compared to noncomplex regions (example in Figure 1B) wherein the occurrence of TFBS modules does not exceed expectation by chance. To compute this enrichment we estimate background probabilities using randomizations of orthologous
sets (details on scoring cut-offs in Extended Experimental Procedures).

We applied PMCA to a set of eight GWAS T2D risk loci (MTNR1B, TCF7L2, PPARG, CETPD2, FTO, GCK, CAMK1D, and KLF14) (Dupuis et al., 2010; Voight et al., 2010) covering strong and weaker GWAS signals and reflecting the different T2D features, i.e., insulin resistance and impaired insulin secretion (Doria et al., 2008). Using noncoding sequence data, we defined 200 SNPs in LD with the tagSNPs \((r^2 \geq 0.7, 1000 Genomes)\) (1000 Genomes Project Consortium et al., 2012) (Figure S1A). PMCA predicted 64 complex and 136 noncomplex regions (Figures 1C–1G; Table S1). We ranked complex regions based on the count of TFBS in conserved TFBS modules (Table S2) and examined the allele-dependent cis-regulatory potential of the 25% highest scoring SNPs using in vitro electrophoretic mobility shift assay (EMSA) and reporter assays. As predicted, SNPs in complex regions significantly differed in allele-dependent cis-regulatory activity compared to noncomplex regions (Figures 1H and 1I; Table S3). Indeed, the regulatory variants revealed effects ranging from 3.1- to 101-fold change in DNA-protein binding and 1.3- to 3.5-fold change in reporter activity. Moreover, the identified variants operated in a cell type-specific manner (Figure S1B).

To examine if the identified cis-regulatory variants in complex regions associate with T2D in vivo, we performed look-ups in the MAGIC and DIAGRAM cohorts (Dupuis et al., 2010; Voight et al., 2010). The variants in complex regions revealed a similar or stronger association compared to the initial GWAS signal (Table S4), and a look-up in a recent fine-mapping study (Maller et al., 2012) confirmed that our cis-regulatory SNPs belong to the predicted T2D-disease SNP set. GWAS signals are enriched for regulatory variants (Nica et al., 2010). Comparing random SNPs from the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2012) to a limited representation of GWAS signals for 19 human diseases (Hindorff et al., 2009) (Table S5A), we defined 200 SNPs in LD with the tagSNPs \((r^2 \geq 0.7, 1000 Genomes)\) (1000 Genomes Project Consortium et al., 2012) to a limited representation of GWAS signals for 19 human diseases (Hindorff et al., 2009) (Table S5A), we defined 200 SNPs in LD with the tagSNPs \((r^2 \geq 0.7, 1000 Genomes)\) (1000 Genomes Project Consortium et al., 2012) to a limited representation of GWAS signals. Using noncoding sequence data, we extended our initial PMCA analysis of eight T2D loci to a set of 47 T2D risk loci comprising all GWAS-reported autosomal variants (Hindorff et al., 2009) including 487 complex and 978 noncomplex regions (Figure S2; Table S7). Noncomplex regions were depleted of constrained elements in their close proximity (Figure 2A). Conversely, complex regions were enriched for nearby constrained elements, consistent with a 1.37-fold enrichment of GWAS SNPs relative to HapMap SNPs (Lindblad-Toh et al., 2011). Although complex regions overlapped 1.88-fold more with constrained elements than noncomplex regions \((p = 2.4 \times 10^{-6}, \text{hypergeometric distribution, right sided})\), strikingly the majority of complex regions lacked an overlap with constrained elements (Figure 2B; Table S8). This lack of overlap was true for all variants that we experimentally characterized as cis-regulatory (example in Figure 2C). In essence, considering sequence conservation helps to prioritize genomic regions that harbor potential causal variants, yet seems insufficient to pinpoint them. This underscores the importance of exploiting conservation in terms of a complexity assessment of co-occurring TFBS, in the search for cis-regulatory variants involved in human diseases.

To further support PMCA predictions at T2D risk loci, we merged our analysis with functional genomics data from The ENCODE Project Consortium (2011) (chromatin state and TF binding). We found complex regions highly enriched for both DHSseq peaks \((p = 3.52 \times 10^{-10})\) (Figure 2D) and ChIP-seq peaks \((p = 4.68 \times 10^{-10})\) (Figure 2E; Table S9). Additionally, crossing our regulatory predictions for T2D risk SNPs with RegulomeDB, a data repository of multiple types of functional ENCODE data (Schaub et al., 2012), confirmed that complex regions are significantly enriched for functional annotations \((p = 3 \times 10^{-24}, \text{hypergeometric distribution, right-sided})\) (Table S10).

**Clustering of Distinct Homeobox TFBS Is a Specific Feature of T2D-Related Complex Regions**

TFBS clustering relative to transcription start sites indicates biological significance (FitzGerald et al., 2004), and TFBS combination coupled with the TFs recruited to a CRM determines CRM function (Zinzen et al., 2009). Thus, we sought evidence for a discerning T2D functional feature by exploring TFBS characteristics in evolutionary conserved complex regions at T2D risk loci. Given a SNP genomic region we used positional bias analysis, scanning 1,000 bp with the SNP at midposition for the occurrence of putative TF binding sequences (883 TFBS matrices grouped in 192 TFBS matrix families) (Table S11). First, for the set of eight T2D risk loci selected for in-depth analysis above, we observed a significant positional bias for distinct TFBS families \((−\log10(p) > 6)\) exactly at SNP positions of complex contrary to noncomplex regions (Figure 3A). This striking SNP-directed overrepresentation in T2D complex regions was restricted to specific TFBS in the homeobox superfamily, including the matrix families CART \((−\log10(p) = 6.52)\) and PDX1 \((−\log10(p) = 6.18)\) (Table S12A). To test whether these
A Workflow of PMCA

1. Select region around noncoding SNP in human sequence
2. Identify orthologous regions in vertebrate species
3. Identify conserved TFBS
4. Infer TFBS modules and TFBS in modules
5. Count $\Omega_{\text{TFBS}}, \Omega_{\text{restr-TFBS}}$, $\Omega_{\text{modules}}, \Omega_{\text{TFBS in modules}}$
6. Repeated counting of $\Omega_{\text{TFBS}}, \Omega_{\text{restr-TFBS}}, \Omega_{\text{modules}}, \Omega_{\text{TFBS in modules}}$
7. Randomized set
8. Scoring of SNPs $p$-est $\text{TFBS}$, $p$-est $\Omega_{\text{TFBS}}, p$-est $\Omega_{\text{modules}}, p$-est $\Omega_{\text{TFBS in modules}}$
9. Classification
   - Complex regions
   - Noncomplex regions
10. Disease-distinctive overall TFBS clustering

B Comparison of TFBS density in complex and noncomplex regions

C D E Comparison of $\Omega_{\text{TFBS}}, \Omega_{\text{modules}}, \Omega_{\text{TFBS in modules}}$

F G H Comparison of density and $\Omega_{\text{TFBS}}$ in complex and noncomplex regions

I Comparison of reporter gene activity in complex and noncomplex regions

(legend on next page)
findings could be retrieved in a larger set of T2D-associated variants, we extended TFBS clustering analysis to the set of 47 GWAS T2D risk loci (Hindorff et al., 2009). Indeed, this comprehensive analysis reproduced colocalization of T2D risk SNPs exclusively with homebox TFBS matrices in complex regions as opposed to noncomplex regions (Figure 3B; Table S12B). We again found specific clustering of the CART (log10(p) = 13.00) and PDX1 families (−log10(p) = 6.78) together with the homebox matrix families NKX6 (-log10(p) = 8.50), HOMF (−log10(p) = 8.94), HBOX (−log10(p) = 8.54), and BCDF (−log10(p) = 7.24). No other TFBS matrices showed a significant peak in the bias profile at SNP positions. Importantly, when applying PMCA on risk loci of T2D nonrelated traits, asthma, and Crohn’s disease (Moffatt et al., 2010; Schaub et al., 2012) (Figures S3B and S3C; Table S13), we observed disease-distinctive TFBS at SNP positions (Table S12C and S12D). Both complex and noncomplex regions lacked a clustering of homeobox TFBS at asthma risk SNPs (Figure 3C). The specific clustering of the early growth response factor matrix family (ERGF) for asthma risk SNPs in complex regions (−log10(p) = 8.50; Figure 3D) was in strong contrast to T2D (−log10(p) = 3.97; Figure 2E) and Crohn’s (−log10(p) = 2.07; Figure S3D). Of note, the ERGF-binding factor EGR1 regulates asthma-related IL13-induced inflammation (Cho et al., 2006).

Homebox TFs are known to be involved in tissue developmental processes including β-cell development (Jorgensen et al., 2007). However, except for the MODY gene PDX1 (Fajans et al., 2001) and the common T2D-associated loci HHEX1 and ALX4 (Sladek et al., 2007), the PMCA-inferred homebox factors have not been implicated in T2D pathogenesis. T2D is marked by insulin resistance and impaired insulin secretion (Doria et al., 2008). To evaluate a functional role of the homebox TFBS matrix families in T2D pathogenesis, we extracted data for insulin resistance (HOMA-IR) and impaired insulin secretion (HOMA-B) (Dupuis et al., 2010), to compute the enrichment of predicted cis-regulatory T2D risk SNPs that localize in close proximity to those homebox TFBS (±20 bp, permutations on the phenotypes, n = 1,000, 95% confidence interval [CI]; Extended Experimental Procedures). We verified a significant enrichment of SNPs that localize ± 20 bp at inferred homebox TFBS for both insulin resistance (mean = 1.09 × 10−6; 95% CI: 9.59 × 10−7−9.51 × 10−6; p = 3.28 × 10−3, mean permutation background) and impaired insulin secretion (mean = 9.45 × 10−5; 95% CI: 5.37 × 10−5−1.34 × 10−5, p = 1.29 × 10−3). Furthermore, we elucidated a potential effect of their binding TFs on impaired insulin secretion. Assessing mRNA levels in human islets from 51 healthy and eight T2D deceased donors by RNA-seq (L.G., unpublished data), we found a marked expression difference for RAX, PRRX2, BARX1, PITX1, EMX2, NKX6-3, BARX2, MSX2, and PDX1 in islets from T2D patients compared to healthy controls (7.28 × 10−10 < p < 4.02 × 10−6, false discovery rate [FDR] < 1%) (Table S14). By genome-wide coexpression analysis we found significantly coregulated gene sets (p < 5.02 × 10−3; FDR < 5%, n = 51 healthy donors) (Table S15). Except for the gene set coreregulated with PITX1, we found metabolic pathways among the top five significantly enriched pathways (hypergeometric test, FDR corrected p < 0.05) (Figure S3E). Other top five enriched pathways included insulin signaling, MAPK signaling, notch signaling, calcium signaling, and pancreatic secretion. Knockdown of each candidate homeobox TF in pancreatic INS-1 β-cells significantly perturbed glucose-stimulated insulin secretion (Figure S3F). Moreover, except for PDX1 and MSX2 (corrected FDR, p = 0.96 and p = 0.89), all PMCA-inferred homebox TFs were significantly coexpressed with the insulin gene in islets of 26 hyperglycemic individuals (hemoglobin A1C [HbA1C] > 6) (Table S16). Although the result for PDX1 was borderline nonsignificant, it is a well-known regulator of insulin expression (Brissova et al., 2002). The other TFs can be regarded as candidates for regulation of proinsulin production.

Figure 1. Discovery of cis-Regulatory Diabetes SNPs

(A) Workflow of the PMCA methodology: (1) the flanking region of a noncoding SNP is extracted from the human reference genome; (2) orthologous regions are searched in the genomes of 15 vertebrate species; (3) TFBS are identified in each orthologous sequence; (4) TFBS modules are identified in the set of orthologous TFBS, TFBS modules, and occurrences of TFBS in TFBS modules; (5) phylogenetically conserved TFBS matrices are determined; (6) phylogenetic footprinting is performed; (7) 10,000 permutations on the phenotypes, n = 1,000, 95% confidence interval [CI]; Extended Experimental Procedures. See also Tables S2 and S3.
The T2D-Identified Variant rs4684847 Regulates PPARG2 Gene Expression

To establish the informative value of TFBS pattern analysis for pinpointing the cis-regulatory variant and binding TF underlying GWAS association signals, we chose the PPARG locus for detailed study. PPARγ is crucial in adipogenesis, lipid metabolism, and systemic insulin sensitivity (Rosen et al., 1999; Medina-Gomez et al., 2005) and exists as two isoforms: PPARγ1 (PPARG1, PPARG3 mRNA) and PPARγ2 (PPARG2 mRNA) (Fajas et al., 1998), the latter mainly expressed in adipocytes (Tontonoz et al., 1994). There is a robust association of PPARG with T2D (Deeb et al., 1998; Heikkinen et al., 2009; Dupuis et al., 2010; Voight et al., 2010). The T2D GWAS association comes from an LD region mainly tagged by the coding missense mutation Pro12Ala (Figure 4A, upper panel). However, the minor 12Ala allele, associated with enhanced insulin sensitivity in humans, paradoxically blunts the transcriptional activity of the insulin-sensitizing PPARγ2 (Deeb et al., 1998). Hypothesizing that the elusive PPARG T2D signal instead arises from a regulatory variant that affects PPARG2 expression, we first confirmed—before analyzing variants at the PPARG locus with PMCA—a risk allele-dependent 3.8-fold decrease of PPARG2 mRNA in human adipose stromal cells (hASCs) (p = 1.0 × 10⁻⁹) (Figure 4B). This effect was specific for PPARG2, as there was no effect on PPARG1 expression (Figure 4C).

First, to narrow-down the variants that could explain the decrease in PPARG2 expression and thereby the underlying T2D association, we applied PMCA to each of the 23 correlated noncoding variants at the PPARG locus (r² ≥ 0.7, 348 Cell 156, 343–358, January 16, 2014 ©2014 Elsevier Inc.)
Figure 3. Positional Bias of Distinct Homeobox TFBS Families at T2D Risk SNPs

Distribution of TFBS matrices relative to SNP positions (SNP ± 500 bp) at T2D compared to asthma risk loci, calculated using positional bias analysis. One thousand base pair genomic regions with SNPs at midposition were scanned for the occurrence of TFBS matches for 192 TFBS matrix families (sliding 50 bp windows, p from binomial distribution model, Extended Experimental Procedures).

(A and B) TFBS family distribution in a set of eight and an extended set of 47 T2D risk loci. Complex regions reveal clustering of distinct homeobox TFBS matrix families at T2D risk SNP positions (±20 bp, gray dashed lines). All TFBS families displayed equal distributions within T2D non-complex regions (a subset of representative TFBS families is shown).

(C) TFBS family distribution in a set of eight asthma risk loci. Asthma complex and noncomplex regions lack a positional bias at SNP positions for the homeobox TFBS matrix families clustering in complex regions at T2D risk SNPs (see Figure S3 for details on Crohn’s).

(D and E) TFBS family distribution in asthma risk loci revealed a specific EGRF matrix family clustering in complex regions at asthma risk SNPs (D). T2D complex regions lack a clustering of EGRF matrices at SNP positions (E).
Figure 4. The Noncoding SNP rs4684847 by Binding the Homeobox Factor PRRX1, Represses PPARG2 Expression at the PPARG Diabetes Risk Locus

(A) Top panel: an LD regional plot of the PPARG locus. Diamonds, tagSNP Pro12Ala and pairwise correlation of SNPs in LD (MAF ≥ 1%) against genomic position; blue, PPARG gene and exons. Middle/lower panel: classification of SNPs in complex regions (red lines) and noncomplex regions (gray lines) (PMCA steps 1–9, Figure 1A). Scanning of PPARG complex regions for T2D-distinct homeobox TFBS matrix families (CART, HOMF, HBOX, NKX6, BCDF, PDX1; Figure 3B) pinpoints rs4684847 (C/T), based on its overlap with the CART binding matrix for PRRX1 (step 10, Figure 1A). Zoom-in, human PPARG gene; arrows, transcription start site (TSS) of PPARG1-3 mRNA isoforms; boxes, coding exons (filled) and untranslated exons (open); lines, introns. Second zoom-in, CRM at rs4684847; the PRRX1 matrix co-occurs with diverse TFBS matrices in consistent orientation and distance range across species, exemplarily illustrated by one conserved TFBS module (U_TFBS_in_modules = 3; TFBS matrices: PRRX1, TEF, LHXF).

(B and C) Genotype-dependent mRNA expression in undifferentiated hASCs genotyped for Pro12Ala and rs4684847 (r^2 = 1.0). qPCR of PPARG1 and PPARG2 mRNA isoforms (standardized to HPRT) homozygous CC risk (n = 9) and CT nonrisk allele carriers (n = 5) normalized to mean for CC. Mean ± SD, t test.

(D) Validation of cis-regulatory predictions for complex regions at the PPARG locus. Quantified change in reporter activity in 3T3-L1 adipocytes is shown for each SNP, using luciferase constructs harboring the risk or nonrisk alleles, representing an activating or repressing effect of the risk allele on transcriptional activity. Mean ± SD, n = 3–14, paired t test.

(legend continued on next page)
Given the availability of cell-stage-dependent open chromatin data in hASCs reported by Mikkelsen et al. (2010), we sought supportive evidence for rs4684847 as the variant underlying the cell-stage-dependent allelic PPARG2 expression. We integrated all six variants in complex regions at the PPARG locus with genome-wide temporal regulatory annotations estimated by H3K27ac data. Among those six, only the flanking region rs4684847 (C/T) showed consistent cell stage-dependent H3K27ac density distributions (Figure S4D). Thus, the rs4684847-specific match with the T2D homeobox TFBS clustering, informed by conserved TFBS pattern analysis, could be confirmed by cell-stage-dependent regulatory regions estimated by chromatin state data.

Finally, we performed a host of in vitro and in vivo analyses to prove that the rs4684847 risk allele (C allele) mediates the suppression of PPARG2 mRNA expression via the transcriptional regulator PRRX1. By affinity chromatography and liquid chromatography-tandem mass spectrometry (LC-MS/MS), we could demonstrate a 2.3-fold increased binding of PRRX1 to the rs4684847 risk relative to nonrisk allele (Extended Experimental Procedures). Moreover, by EMSA we found rs4684847 risk allele-specific DNA-protein binding (Figure 4F), and competition EMSA and supershift experiments confirmed that PRRX1 was responsible for this allele-specific DNA-protein binding (Figure 4G). Furthermore, consistent with the GWAS signal for insulin resistance rather than insulin secretion (Voight et al., 2010), in luciferase reporter assays we observed rs4684847 cell type-specific effects in 3T3-L1 adipose cells, C2C12 myocytes and HuH7 hepatocytes, whereas pancreatic INS-1 β-cells and 293T cells lacked allelic activity (Figure S4E). Luciferase activity in 3T3-L1 preadipocytes was 5.2-fold lower for the C risk allele (p = 1.0 × 10−4; Figure 4H). This repressive effect was independent of 5′- versus 3′-orientation to the reporter gene (p = 0.03) and forward-reverse orientation (p = 0.03) (Figure S4F), suggesting enhancer function for the nonrisk allelic complex region.

Importantly, perturbing the PRRX1 consensus sequence without affecting the SNP position itself fully abrogated the C risk allelic repression of reporter gene activity (Figure 4H), whereas overexpressing PRRX1 enhanced it (p = 2.0 × 10−4; Figure 4I).

We then sought proof that the rs4684847 risk allele—indeed of correlated sequence variants—causes the suppression of endogenous PPARG2 expression. We used an adopted CRISPR/Cas homology-directed repair genome editing approach (Wang et al., 2013a) to introduce the rs4684847 nonrisk allele in human Simpson-Golabi-Behmel syndrome (SGBS) preadipocytes, replacing the endogenous risk allele. Notably,
the rs4684847 nonrisk allele was sufficient to increase PPAR2 transcript levels 5.4-fold ($p = 0.005$) (Figure 4J, left) (PPARG1 unaffected) (Figure S4G). In parallel experiments, we performed PRRX1 knockdown and confirmed that (1) risk allele-driven suppression of PPAR2 expression was reversed by PRRX1 silencing ($p = 0.005$), and (2) PRRX1 silencing did not affect PPAR2 expression in nonrisk allele cells (Figure 4J, right).

**rs4684847 via PRRX1 Binding Affects FFA Homeostasis and Insulin Sensitivity**

The SNP rs1801282 (Pro12Ala) in PPAR2 associates with BMI, fasting insulin, and insulin sensitivity (Deeb et al., 1998; Voight et al., 2010). rs4684847 is located 6.5 kb upstream of the PPAR2-specific promoter and is in complete LD ($r^2 = 1.0$) with rs1801282. Via PMCA, we found that PRRX1 binds at the rs4684847 C risk allele and thus inhibits PPARG2 expression. On the other hand, the T allele of rs4684847 (minor allele frequency 6.5% in Caucasians) reduces the binding ability of PRRX1 and thus maintains a higher level of PPARG2 expression. Further in vivo evidence was obtained in primary human adipose stromal cells (hASCs) isolated from BMI-matched obese patients, showing rs4684847-dependent PPARG2 mRNA expression ($p = 1.4 \times 10^{-20}$, $n = 32$). PPARγ2 is crucial for maintaining insulin sensitivity: adipose-specific Pparg2 knockout mice develop insulin resistance independently of affecting body weight (Medina-Gomez et al., 2005), and PPARγ is target of the thiazolidinedione (TZD) class of insulin-sensitizing drugs such as Rosiglitazone (Rosiglitazone) (Lehmann et al., 1995). Indeed, we observed rs4684847-dependent association with lower T2D risk (Voight et al., 2010) ($OR = 0.89$, 95% CI = 0.86–0.92, $p = 3.75 \times 10^{-11}$, $n = 80,648$). Further, in hASCs we found rs4684847-dependent increase in adipocyte insulin sensitivity ($p = 1.5 \times 10^{-7}$), ratio insulin-stimulated/basal 2-deoxyglucose uptake, Pearson’s correlation, $n = 32$). We confirmed a significant interaction between the rs4684847 risk allele and adipose PRRX1 mRNA levels to HOMA-IR, independent of BMI ($p = 0.044$, $n = 38$, interaction model; Extended Experimental Procedures). In addition, we observed rs4684847-dependent correlations of PRRX1 mRNA levels with BMI, TG/HDL ratio, and BMI-adjusted HOMA-IR and with glucose infusion rate (GIR) measured by euglycemic hyperinsulinemic clamp in a cohort of 67 BMI- and body fat-matched obese patients (Table 1; Figure S4H).

To further examine PRRX1 as mediator of the repressive rs4684847 risk allele (C allele) effect on PPAR2 expression, we performed knockdown of PRRX1 in primary hASCs and found that PRRX1 silencing was sufficient to reverse the risk allelic suppression ($p = 3.3 \times 10^{-15}$) (Figure 5A; Table 2). Then, to inform on the cellular processes by which PRRX1 may contribute to T2D, we studied the impact of PRRX1 on PPARγ-regulated genes in hASCs from homozygous rs4684847 CC risk allele carriers by microarray analysis ($n = 9$). We found 2,258 transcripts regulated by PRRX1 knockdown ($q < 0.2$), 336 of which were reversely regulated by concomitant PPAR γ knockdown (Figure 5B). Gene set enrichment analysis (GSEA) highlighted an enrichment of those antiregulated genes among the most differentially expressed genes after PRRX1 knockdown (FDR = 0, Figure 5C), revealing that PPARγ2 mediated the primary PRRX1 effect on global gene expression. Ingenuity pathway analysis (IPA) showed the strongest enrichment for lipid metabolism ($p = 2.81 \times 10^{-14}$) followed by adipose tissue function, glucose homeostasis, nutritional disease, and insulin resistance (Figure 5D). Accordingly, an inverse relationship between PRRX1 and adipocyte triglyceride (TG) accumulation was observed in PRRX1-overexpressing SGBS adipocytes (Figure 5E).

By qPCR, we confirmed rs4684847 allele-dependent dysregulation of genes in the identified biological pathways. Notably, the

<table>
<thead>
<tr>
<th>rs4684847 genotypes</th>
<th>PRRX1 mRNA</th>
<th>PRRX1 mRNA</th>
<th>PRRX1 mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>CC</td>
<td>CT and TT</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>$p$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>A</td>
<td>$n = 38$</td>
<td>$n = 20$</td>
<td>$n = 18$</td>
</tr>
<tr>
<td>log(BMI)</td>
<td>1.32</td>
<td>0.05</td>
<td>1.43</td>
</tr>
<tr>
<td>age</td>
<td>1.45</td>
<td>0.03</td>
<td>1.96</td>
</tr>
<tr>
<td>log(TG/HDL)</td>
<td>6.92</td>
<td>7.54 $\times$ $10^{-4}$</td>
<td>6.40</td>
</tr>
<tr>
<td>age</td>
<td>6.97</td>
<td>7.36 $\times$ $10^{-4}$</td>
<td>6.14</td>
</tr>
<tr>
<td>age/BMI</td>
<td>4.86</td>
<td>8.3 $\times$ $10^{-3}$</td>
<td>5.00</td>
</tr>
<tr>
<td>log(HOMA-IR)</td>
<td>2.77</td>
<td>3.52 $\times$ $10^{-3}$</td>
<td>3.13</td>
</tr>
<tr>
<td>age</td>
<td>2.77</td>
<td>3.77 $\times$ $10^{-3}$</td>
<td>3.12</td>
</tr>
<tr>
<td>age/BMI</td>
<td>1.41</td>
<td>0.028</td>
<td>2.1</td>
</tr>
<tr>
<td>B</td>
<td>$n = 67$</td>
<td>$n = 54$</td>
<td>$n = 13$</td>
</tr>
<tr>
<td>log(GIR)</td>
<td>age/BMI</td>
<td>1.83 $\times$ $10^{-7}$</td>
<td>0.78</td>
</tr>
<tr>
<td>log(FFA)</td>
<td>age/BMI</td>
<td>0.014</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Gene expression and phenotypes were measured in (A) adipose tissue from a lean/obese patient cohort (mean ± SD 24.2 ± 9.1 kg/m²), and (B) adipose tissue samples from BMI-matched obese patients (mean ± SD 43.2 ± 3.1 kg/m²) characterized by hyperinsulinemic euglycemic clamp. rs4684847 risk allele and nonrisk allele genotypes were determined by Sequenom-assay. $p$ values and $\beta$-estimates from linear regression analysis of PRRX1 mRNA expression levels with phenotype measures are shown. BMI, body mass index; FFA, free fatty acids; GIR, glucose infusion rate of hyperinsulinemic euglycemic clamp; HDL, high density lipoprotein; HOMA-IR, homeostasis model assessment of insulin resistance; TG, triglyceride.
gene with the strongest risk allele-dependent decrease in mRNA levels was PEPCKC (Table 2). The top scoring IPA interaction network reinforced a central role for PEPCKC (Figure 5F). PEPCK-C is the enzyme controlling the first committed step of glyceroneogenesis, a crucial metabolic process in adipocytes regulating the re-esterification of free fatty acids (FFA) to TG (Ballard et al., 1967). Glyceroneogenesis limits FFA release from adipocytes in the fasting state thereby controlling systemic FFA homeostasis and insulin sensitivity (Millward et al., 2010). In the 67 BMI- and body fat-matched obese subjects, we confirmed rs4684847 risk allele association with increased serum FFAs levels (p = 0.049) and risk allele-dependent association of PPRX1 mRNA with FFA levels (p = 0.015, Table 1). To prove that rs4684847, by determining PRRX1 binding, affects glyceroneogenesis and subsequent FFA release, we monitored pyruvate incorporation in TG (Ballard et al., 1967). We confirmed a PPRX1-dependent suppression of glyceroneogenesis in CC risk allele carriers, marked by a robust correlation with PPRX1 mRNA levels (Figure 5G) and a risk allele-dependent increase in FFA release (Figure 5H). In a parallel experiment, we also found that PPRX1 silencing was sufficient to restore cellular insulin sensitivity in risk allele carriers (Figure 5I). Importantly, the PPARγ ligand Rosi pharmacologically promotes insulin sensitivity largely via control of FFA homeostasis through glyceroneogenesis (Cadoulal et al., 2007), and Kang et al. (2005) reported impaired Rosi response in risk haplotype carriers. In our analysis of glyceroneogenesis in hASCs, we observed an impaired response to Rosi-mediated suppression of FFA release dependent on the risk allele (Figure 5J). Strikingly, PPRX1 silencing in CC risk allele patient samples was sufficient to abolish the reduced Rosi responsiveness, making PPRX1 a potential target for pharmacological T2D intervention.

In summary, by PMCA we demonstrate a clustering of specific homeobox TFBS at T2D risk SNPs. We specifically unveil a role of the homeobox TF PPRX1 as a repressor of PPARG2 via its enhanced binding at the rs4684847 C risk allele, thereby provoking dysregulation of FFA turnover and glucose homeostasis (Figure 5K).

DISCUSSION

We have developed a bioinformatics approach, PMCA, which enables the extraction of cis-regulatory variants that may mechanistically contribute to human disease by dysregulation of gene expression. In line with our approach to exploit conservation in terms of co-occurring TFBS patterns, (Visel et al., 2013) has recently shown that combination of TFBS, rather than single TFBS, via combinatorial TF binding governs spatial enhancer activity in the developing telencephalon. Further, tissue-specific enhancers were accurately detected by in vivo mapping of the enhancer-associated proteins p300, in addition to comparative genomics approaches (Visel et al., 2009a; Blow et al., 2010). Using T2D as a showcase, we demonstrate the utility of PMCA for the generic prediction of distinct homeobox TFBS at T2D risk SNPs, which is important for understanding disease regulatory circuits when we consider that interactions in a regulatory network involve numerous genes and a rather small set of TFs (Califano et al., 2012). Pursuing the results emerging from our comprehensive T2D analysis, we show that identification of the cis-regulatory variant rs4684847 at the PPARG locus enabled linking the molecular upstream factor PPRX1 to aberrant downstream mechanisms of impaired lipid handling and insulin sensitivity, explaining the GWAS association with T2D. Notably, PPRX1 was recently implicated in adipogenesis (Du et al., 2013), yet the regulated genes remain elusive.

Here, we restricted the analysis to SNPs in LD with GWAS SNPs. However, the approach could be applied to any other kind of variability, such as somatic mutations in cancer, without loss of generality. Certain issues will require consideration, e.g., analyzing genomes of closely related species to refine scoring criteria, and extending our analysis to whole genome sequencing studies, including rare variants data, should further inform on the genetic underpinnings of phenotypic diversity in humans. Our in silico scoring results predict varying numbers of regulatory SNPs per LD block. Studies have now found evidence for allelic heterogeneity (Maller et al., 2012; Schaub et al., 2012), yet the number of causal variants within a disease locus is elusive. We propose an integrative framework where computational TFBS modularity analysis may be synergistically combined with functional genomics and population genetics data.

In sum, our results demonstrate that the extension of sequence analysis to functional conservation integrates biological data with statistical signals, and our method should help to clarify the role of inherited and somatic variability in altering gene regulatory networks, in both mendelian and common human diseases.

EXPERIMENTAL PROCEDURES

See the Extended Experimental Procedures for details.

LD Block Definition

SNPs in close LD ($r^2 \geq 0.7$) to GWAS tagSNPs (references in Tables S1, S5, S7, and S13) from 1000 Genomes Project, Pilot 1, CEU data (http://www.1000genomes.org/).

PMCA

PMCA analyzes the occurrence of conserved patterns of TFBS in a CRM within the genomic region flanking a noncoding variant, to predict its cis-regulatory functionality. For each variant the PMCA method provides a classification of the region surrounding the variant as being either complex or noncomplex. Complex regions are defined as being significantly enriched in conserved co-occurring TFBS (TFBS modules) according to the scoring scheme described in Extended Experimental Procedures.

Positional Bias Analysis

Complex and noncomplex regions (SNP $\pm$ 500 bp) were scanned for presence of TFBS family matches at SNP positions. Positional bias of TFBS families was calculated using overlapping 50 bp sliding windows in steps of 10 bp. Positional bias (p) was calculated as binomial p value for each TFBS family and each window.

Correlation with Evolutionary Constraint, DHSseq, and ChIP-Seq Regions

Complex/noncomplex SNP regions (SNP $\pm$ 600 bp) were correlated to constrained regions or DHSseq and ChIP-seq peaks. From midpoint of constrained regions ($\pm$500 bp), DHSseq ($\pm$1,000 bp), or ChIP-seq peaks ($\pm$1,000 bp), the overlapping positions (correlation) with complex/noncomplex regions were counted and plotted versus position relative to anchor. For the calculation of enrichment of DHS and ChIP-seq peak overlaps to
Figure 5. Binding of PRRX1 at the rs4684847 Risk Allele in Human Adipose Cells Affects Lipid Metabolism and Insulin Sensitivity

(A) rs4684847-dependent PPARG2 and PRRX1 mRNA levels measured by qPCR (standardized to HPR7) in hASC from BMI-matched rs4684847 CT (n = 16) and CC (n = 32) risk allele carriers. siPRRX1 and siNT transfected concurrent with induction of adipogenic differentiation for 72 hr. Left: Pearson’s correlation in the siNT set. Right: box-whisker plot comparing PPARG2 mRNA in siNT- versus siPRRX1-treated cells (t test). FC, fold change.

(B and C) Global gene expression profiling by Illumina microarrays (q < 0.2) in hASCs from rs4684847 CC risk allele carriers transfected with siPRRX1 (n = 9, gray dots) and cotransfected with siPRRX1 and siPPARG (n = 4, red dots) for 72 hr after induction of adipogenic differentiation (B). Distribution of siPRRX1/siPPARG antiregulated genes among all regulated genes ranked by fold change (C).

(D and E) Biological pathways associated with siPRRX1/siPPARG antiregulated genes (D) and top scoring interaction network (E) from ingenuity pathway analysis.

(legend continued on next page)
The rs4684847 risk allele (C allele) promotes PRRX1 binding 6.5 kb upstream of the siPRRX1-treated cells; t test.

±rs4684847 CT (n = 16) and CC (n = 32) risk allele carriers after silencing of PRRX1. (G) Left: Pearson’s correlation in the siNT set. Right: box-whisker plot comparing siNT- versus siPRRX1-treated cells, t test. FFA, free fatty acids.

±rs4684847-dependent increase of [3H]-2-deoxyglucose ([3H]-2DG) uptake following insulin stimulation in hASCs. Box-whisker plot comparing siNT- versus siPRRX1-treated cells; t test. FFA, free fatty acids.

±rs4684847-dependent glyceroneogenesis rate measured by [1-14C]-pyruvate incorporation (G) and FFA release (H) in hASCs from BMI-matched heterozygous (hetero, n = 16) and homozygous (homo, n = 32) risk allele carriers (genotyped for the cis-regulatory variant rs4684847 and the tagSNP rs1801282 Pro12Ala). ADIPOQ, adiponectin, C1Q and collagen domain containing; BBOX1, Box 1 homolog; FFA, free fatty acids; GLUT4, glucose transporter type 4; LIPE, lipase, hormone-sensitive; OPG, Osteoprotegerin; p, p value from unpaired t test; PDK4, pyruvate dehydrogenase kinase, isozyme 4; PEPCKC, Phosphoenolpyruvate carboxylase cytosolic; PPAR, peroxisome proliferator-activated receptor; PPRX1, paired-related homeobox 1; THRSP, thyroid hormone responsive Spot 14 Protein; TIMP3, TIMP metalloproteinase inhibitor 3.

Table 2. Genotype-PRRX1-Dependent Regulation of PRXX1/PPARG Antiregulated Genes in hASCs

<table>
<thead>
<tr>
<th>Genotype</th>
<th>siNT</th>
<th>siPRRX1</th>
<th>siPRRX1/siNT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hetero</td>
<td>Homo</td>
<td>Hetero/Homo</td>
</tr>
<tr>
<td>PRRX1</td>
<td>Mean ± SD</td>
<td>Mean ± SD</td>
<td>FC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.52 ± 0.18</td>
<td>0.51 ± 0.19</td>
<td>1.01</td>
</tr>
<tr>
<td>PPARG2</td>
<td>4.32 ± 1.07</td>
<td>0.79 ± 0.08</td>
<td>0.18</td>
</tr>
<tr>
<td>PPARG1</td>
<td>1.07 ± 0.26</td>
<td>1.04 ± 0.33</td>
<td>1.03</td>
</tr>
<tr>
<td>PEPCKC</td>
<td>2.83 ± 0.58</td>
<td>1.03 ± 0.20</td>
<td>2.76</td>
</tr>
<tr>
<td>PDK4</td>
<td>2.01 ± 0.88</td>
<td>0.74 ± 0.18</td>
<td>2.73</td>
</tr>
<tr>
<td>LIPE</td>
<td>1.37 ± 0.64</td>
<td>0.68 ± 0.32</td>
<td>2.01</td>
</tr>
<tr>
<td>ADIPOQ</td>
<td>1.89 ± 0.32</td>
<td>0.95 ± 0.31</td>
<td>1.98</td>
</tr>
<tr>
<td>OPG</td>
<td>0.78 ± 0.36</td>
<td>1.67 ± 0.53</td>
<td>4.37</td>
</tr>
<tr>
<td>TIMP3</td>
<td>0.61 ± 0.21</td>
<td>1.50 ± 0.52</td>
<td>0.41</td>
</tr>
<tr>
<td>BBOX1</td>
<td>2.16 ± 0.48</td>
<td>0.96 ± 0.30</td>
<td>2.26</td>
</tr>
<tr>
<td>GLUT4</td>
<td>1.57 ± 0.35</td>
<td>0.99 ± 0.24</td>
<td>1.58</td>
</tr>
<tr>
<td>THRSP</td>
<td>0.99 ± 0.28</td>
<td>1.26 ± 0.39</td>
<td>0.61</td>
</tr>
</tbody>
</table>

The PPRX1/PPARG antiregulated genes were identified by Illumina microarray analysis in samples with PPRX1 knockdown and simultaneous PPRX1 and PPARG knockdown during adipogenic differentiation (Figure 5E). Confirmatory qRT-PCR was performed for these representative top regulated genes in hASC from BMI-matched heterozygous (hetero, n = 16) and homozygous (homo, n = 32) risk allele carriers (genotyped for the cis-regulatory variant rs4684847 and the tagSNP rs1801282 Pro12Ala). ADIPOQ, adiponectin, C1Q and collagen domain containing; BBOX1, Box 1 homolog; FFA, free fatty acids; GLUT4, glucose transporter type 4; LIPE, lipase, hormone-sensitive; OPG, Osteoprotegerin; p, p value from unpaired t test; PDK4, pyruvate dehydrogenase kinase, isozyme 4; PEPCKC, Phosphoenolpyruvate carboxylase cytosolic; PPAR, peroxisome proliferator-activated receptor; PPRX1, paired-related homeobox 1; THRSP, thyroid hormone responsive Spot 14 Protein; TIMP3, TIMP metalloproteinase inhibitor 3.

complex/noncomplex SNPs only those SNPs were considered where an overlap was detected within ± 20 bp from SNP positions.

Primary Human Tissue and hASC

Human islets and adipose tissue were obtained with informed consent from each subject. The studies were approved by the local ethics committees of the Technische Universität München (Germany), the Haukeland University Hospital (Norway) and the Lund University (Sweden). Primary hASCs were isolated from subcutaneous adipose tissue and differentiated in vitro. Genotyping was done by MassARRAY (Sequenom), Omni express (Illumina), or Sanger Sequencing.

RNA Preparation and Expression Analysis

Total RNA was prepared by TRIzol (Invitrogen) or RNeasy Lipid Tissue Mini Kit (QiAGEN), and gene expression was measured by qPCR or microarrays (Affymetrix, Illumina). Allele-specific primer extension was performed with SnAphosKit (ABI Prism).

Cell Culture and Reporter Assays

Huh7, INS-1, 293T, C2C12, 3T3-L1, and SGBS cells were cultured using standard protocols. Genomic sequences surrounding SNPs were synthesized (MWG), cloned in pGL4.22-TK-promoter (Promega) and transfected in cells by Lipofectamine (Invitrogen). Luciferase activity was measured by Luminoscan-Ascent (Thermo).

Gene Knockdown by Small Interfering RNA

All knockdowns were performed with ON-TARGETplus SMARTpool small interfering RNA (siRNA) (Dharmacon) and HiPerFect (QiAGEN).

CRISPR/Cas Genome Editing

HDR genome editing was performed in human SGBS preadipocytes by transfection of CRISPR/Cas9 and single guide RNA (sgRNA) expression vectors (sgRNA targeting a NGG PAM sequence 5'-CAGAGGC-3') and rs4684847 DNA donor vectors (T allele to replace endogenous allele, C allele control). Cell enrichment by MACS selected transfected cell selection kit (Miltenyi). rs4684847 genome editing was confirmed by Sanger sequencing.

EMSA

Forty-two base pairs of allelic Cy5-labeled-DNAs (MWG) and nuclear protein were used for EMSA. Supershift experiments were performed with siPRRX1 or IgG, competition with excess unlabeled probe, and protein from pCMV-PRRX1-flag transfected 293T.
DNA-Protein Affinity Chromatography LC-MS/MS

DNA-protein affinity chromatography was performed with streptavidin magnetic beads (Invitrogen) and allelic biotinylated DNA-probes (MWG and Ultimate3000nano HPLC ( Dionex) LC-MS/MS coupled to LTQ Orbitrap XL (Thermo Fisher Scientific). Data were analyzed with Progenesis software v2.5.

Statistical Analysis

Statistical analyses were done using Graph Pad Prism v5.02, R Software v2.14.2 or Perl scripts.

ACCESSION NUMBERS

Microarray data for hASC are available in ArrayExpress (E-MTAB-1906). The Gene Expression Omnibus (GEO) accession number for the adipose tissue microarray analysis reported in this paper is GSE25402.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, 4 figures, and 17 tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.10.058.

CONSORTIA


ACKNOWLEDGMENTS

This work was funded by the Else Kröner-Fresenius Foundation, Bad Homburg v.d.H, Germany; the grant Virtual Institute “Molecular basis of glucose regulation and type 2 diabetes” received from the Helmholtz Zentrum München, München-Neuherberg, Germany; the grant Clinical Cooperation Group “Nutrigenomics and type 2 diabetes” received from the Helmholtz Zentrum München, München-Neuherberg, Germany, and the Technische Universität München; the Helmholtz Graduate School for Environmental Health, HELENA; a grant from the German Federal Ministry of Education and Research to the German Centre for Diabetes Research (DZD e.V.); the Competence Network Obesity (German Obesity Biomaterial Bank; FKZ 01GI1128), and the University Duisburg-Essen (01KU1216E); the KG Jeans Center for Diabetes Research, University of Bergen, Norway and the Western Norway Regional Health Authority, Norway; by grants from the Swedish Research Council, including strategic research area grant EXODIA (2009-1039), Linneaus grant (349-2006-237), Collaborative Grant (2011-3315), and Project Grant (521-2010-3490) and by an ERC Advanced Researcher Grant GENETARGET-T2D (GA 269045); and by “Biomedical Research Program” funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation. We thank Karl-Fredrik Eriksson and Targ Elgøyr for providing human fat biopsy material and Charlotte Ling for supporting generation of microarray data (Lund University). We are grateful to Andrea Califano (Columbia University, New York, NY) for critical review and constructive scientific comments on the manuscript. We further thank Bernd Baumann, Vidar M. Steen, and Jorn V. Sagen for expert advice, and Elisabeth Hofmair, Manuela Hubersberger, Margit Solsvik, Linn Jeannette Waagbo, Tone Nygaard Flieto, Jan-Inge Bujze, Zina Fandalyuk, Vivian Veum, and Christine Haugen for excellent technical assistance. We thank Ralf Kühn (Helmholtz Zentrum München, München-Neuherberg) for providing CRISPR/Cas vectors. We thank Christine Stansberg, Rita Holdhus, and Kjell Petersen at the Norwegian Microarray Consortium (NMC) (Bergen node, Norway), Douglas P. Kiel and David Karasik (Hebrew SeniorLife Institute for Aging Research, Harvard Medical School, Boston, MA), and Michael Molla (Joslin Diabetes Center, Harvard Medical School, Boston, MA) for expert assistance. We thank surgeon Hans Jørgen Nielsen and colleagues (Voss Hospital, Norway), surgeons Barbara Auras Jaatun, Inge Glambæk, and colleagues (Høradalsplass Da coneoss Hospital, Bergen), surgeon Christian Busch (Klinik Bergen, Næsund), and all the patients for providing human adipose tissue. We thank Reiner Schroer and Dorothy Dankel for critical reading of the manuscript. The authors have filed a patent related to this work.

Received: April 6, 2013
Revised: September 5, 2013
Accepted: October 30, 2013
Published: January 16, 2014

REFERENCES

Alberobello, A.T., Congedo, V., Liu, H., Cochran, C., Skarulis, M.C., Forrest, D., and Cell, F.S. (2011), An intronic SNP in the thyroid hormone receptor 1 gene is associated with pituitary cell-specific over-expression of a mutant thyroid hormone receptor 1r (R338W) in the index case of pituitary-selective resistance to thyroid hormone. J. Transl. Med. 9, 144.


