Allele-specific quantitative proteomics unravels molecular mechanisms modulated by cis-regulatory PPARG locus variation

Heekyoung Lee1,2,3,4, Kun Qian1,2,3,4,5, Christine von Toerne4,5, Lena Hoerburger2,6, Melina Claussnitzer1,2,3,4,7, Christoph Hoffmann2,8, Viktoria Glunk1,2,3,4, Simone Wahl4,9,10, Michaela Breier4,9,10, Franziska Eck5, Leili Jafari5, Sophie Molnos4,9,10, Harald Grallert3,4,9,10, Ingrid Dahlman11, Peter Arner11, Cornelia Brunner12, Hans Hauner1,2,3,4,13, Stefanie M. Hauck4,5,* and Helmut Laumen1,2,3,4,5,6,14,*

1Else Kroener-Fresenius-Center for Nutritional Medicine, Chair of Nutritional Medicine, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 2ZIEL - Institute for Food & Health, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 3Clinical Cooperation Group Nutrigenomics and Type 2 Diabetes, Helmholtz Zentrum München and Technische Universität München, 85354 Freising-Weihenstephan, Germany, 4German Center for Diabetes Research (DZD), Germany, 5Research Unit Protein Science, Helmholtz Zentrum München, 85764 Neuherberg, Germany, 6Else Kroener-Fresenius-Center for Nutritional Medicine, Paediatric Nutritional Medicine, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 7Hebrew SeniorLife Institute for Aging Research, Harvard Medical School, Boston, MA 02131, USA, 8Else Kroener-Fresenius-Center for Nutritional Medicine, Chair of Molecular Nutritional Medicine, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 9Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany, 10Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany, 11Department of Medicine, Huddinge, Karolinska Institutet, SE-141 86 Stockholm, Sweden, 12Klinik für Hals-Nasen-Ohrenheilkunde, Universitätsklinik Ulm, 89075 Ulm, Germany, 13Else Kroener-Fresenius-Center for Nutritional Medicine, Klinikum rechts der Isar, Technische Universität München, 81675 München, Germany and 14Institute of Experimental Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Received September 08, 2016; Revised January 23, 2017; Editorial Decision January 24, 2017; Accepted February 07, 2017

ABSTRACT

Genome-wide association studies identified numerous disease risk loci. Delineating molecular mechanisms influenced by cis-regulatory variants is essential to understand gene regulation and ultimately disease pathophysiology. Combining bioinformatics and public domain chromatin information with quantitative proteomics supports prediction of cis-regulatory variants and enabled identification of allele-dependent binding of both, transcription factors and coregulators at the type 2 diabetes associated PPARG locus. We found rs7647481A nonrisk allele binding of Yin Yang 1 (YY1), confirmed by allele-specific chromatin immunoprecipitation in primary adipocytes. Quantitative proteomics also found the coregulator RING1 and YY1 binding protein (RYBP) whose mRNA levels correlate with improved insulin sensitivity in primary adipose cells carrying the rs7647481A nonrisk allele. Our findings support a concept with diverse cis-regulatory variants contributing to disease pathophysiology at one locus. Proteome-wide identification of both, transcription factors and coregulators, can profoundly improve understanding of mechanisms underlying genetic associations.

INTRODUCTION

Genome-wide association studies (GWAS) identified thousands of loci associated with diverse diseases (1). The majority of variants are located in noncoding DNA regions and have been proposed to affect transcriptional regulation. Advances of the ENCODE project and novel bioinformatics approaches improved the identification of cis-regulatory
variants at complex loci (2–7). Moreover, deciphering allele-specific binding of transcription factors is essential to unravel the mechanisms ultimately affecting gene expression (6,8–10). However, the identification of allele-specific binding of transcription factors and coregulators remains elusive in most cases, despite the well-established importance of a coordinated interaction between transcription factors, coregulators, and the basal transcriptional machinery for regulation of gene expression (11). Thus, in most cases the precise molecular mechanisms underlying associations between variants and disease risk remain unknown.

Quantitative protein–DNA proteomics, coupling affinity chromatography with liquid–chromatography tandem mass spectrometry (LC–MS/MS) was reported for identification of enhancer-binding proteins and for allele-specific DNA binding transcription factors (6,12,13). In this study, we introduce label-free quantitative proteomics on salt eluted sub-fractions containing native complexes, circumventing limitations of metabolic labeling strategies such as time-consuming workflows, high-costs or inefficiencies (14), or artifacts (15). The preserved allele-specific DNA-protein binding activity in the eluted sub-fractions significantly reduced complexity and allowed high coverage of quantified proteins and thereby identification of both, transcription factors and coregulators (Figure 1A).

The Peroxisome proliferator-activated receptor gamma (PPARG) locus has been robustly associated with type 2 diabetes (16–18). PPARG is crucial for adipogenesis, lipid metabolism, and systemic insulin sensitivity (19, 20) with the PPARG2 isoform being mainly expressed in adipocytes (21–24). In a previous study, we reported that adipose tissue PPARG locus expression quantitative trait locus (eQTL) data revealed allele-specific regulation of PPARG expression (6). In the same study, using bioinformatics phylogenetic module complexity analysis (PMCA) we found the regulatory variant rs4684847. This variant overlaps with a homeobox transcription factor binding site (TFBS), a common feature inferred for type 2 diabetes risk variants. We further reported rs4684847C risk allele binding of the Paired Related Homeobox 1 (PRRX1) transcription factor. PRRX1 represses PPARG expression, and negative correlation of PRRX1 mRNA levels with insulin-sensitivity supports contribution to insulin resistance phenotype at the PPARG locus (6).

In the current study, combining PMCA (6) with public domain DNase sequencing (DNase-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) data we inferred rs7647481, a second PPARG locus cis-regulatory variant. We show that label-free proteomics can find both, transcriptional factors and public domain epigenomic marks of regulatory regions infer the cis-regulatory variant rs7647481 at the PPARG locus (related to Supplementary Figure S1). (B) PMCA analysis of cross-species TFBS pattern conservation predicted six indicated candidate cis-regulatory SNPs at complex regions (6) red out of 23 noncoding proxy SNPs (\(r^2 \geq 0.7\)) at the type 2 diabetes (T2D) associated PPARG locus (tagSNP rs1801282). (C) Overlap of six variants identified in (B) with H3K27ac (histone H3-lysine 27 acetylation), H3K4me1 and H3K4me2 (histone H3-lysine 4 mono- and di-methylation) histone modification regions at the PPARG locus during adipogenic differentiation of primary human adipocyte stem cells (36), GSE21366, genomic coordinates are given conform to hg19. (D) Localization of cis-regulatory (red) and non cis-regulatory (grey) variants subjected to workflow (A2–6) relative to transcriptional start site of the PPARG1–3 mRNA isoforms. rs7647481 over-lapping with both, day 3 and day 9, tested late stage of adipogenesis histone modification regions (Figure 1C) and with adipocyte DNase-seq regions (see Supplementary Figure S1). * rs4684847 previously identified as specifically overlapping with homeobox TFBS (6). Blue boxes = coding exons, dashed white boxes = untranslated exons, blue lines = introns, black arrows = promoters.
tion factors and cofactors at the rs7647481A nonrisk and rs4684847C risk allele, supporting the prediction of both regulatory variants. For rs7647481, we present novel experimental and human genetics data supporting the pathophysiological relevance of the transcription factor YY1 and cofactor RYBP.

MATERIALS AND METHODS

Cell culture

The SGBS human preadipocytes cell strain, primary human preadipocytes, Huh7 human hepatoma, C2C12 mouse myoblast, INS-1 rat insulinoma, 293T human embryonic kidney and the HIB1B mouse brown adipocyte cell line were cultured, induced and differentiated as previously described (25). Ethical committee approval for primary human preadipocyte cells was obtained from the Faculty of Medicine of the Technical University of Munich, Germany.

Preparation of nuclear extracts

Nuclear protein extracts from human SGBS preadipocytes, differentiated SGBS adipocytes, and human primary preadipocytes were prepared as described (26). Briefly, cells were washed with phosphate-buffered saline (PBS), lysed in pre-chilled buffer A (10 mM HEPES pH 8.0, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, protease- and phosphatase inhibitor mix (Roche, Mannheim, Germany), 0.6% Nonidet-P40). Nuclei were collected by centrifugation, lysed in buffer B (20 mM HEPES pH 7.9, 1.5 mM MgCl2, 20 mM KCl, 0.2 mM EDTA, 20 mM NaF, 0.2 mM Na2EDTA, 20 mM NaF, 0.5 mM DTT), lysed by adding high salt buffer containing 50 mM NaCl and 0.01% CHAPS using heating to 90°C for 5 min in TE buffer followed by slow cooling down at room temperature overnight. For competition experiments unlabeled oligonucleotides (CdxA 5'-G CATTATTACACCAGCTGCAGTTGTGA-3'; MyoD 5'-CCCCCAACAGCTGGTCCTGA-3'; Scramble 5'-AGCAACCCCTGACTTAGTATAGTAGTCAAG ACCGCCCACTTT-3'; YY1 5'-CGCTTCCGGCCGCTAC TGGGCGCTGTTG-3') were used. Double-stranded oligonucleotides were separated from remaining single strand oligonucleotides on a 12% polyacrylamide gel.

EMSA

Non-radioactive EMSA was performed with Cy5-labeled oligonucleotide as previously described (6) with some modifications. 5 μg of nuclear protein and 0.35 μg of poly (dI-dC) (Roche, Mannheim, Germany) were used for each reaction. The reaction conditions for affinity chromatography are in general similar to those used for EMSA. To optimize DNA-protein binding in EMSAs, several conditions were tested, such as amount of proteins, concentration of the nonspecific competitor poly (dI-dC) and salt concentration in the gel binding buffer (3% (v/v) glycerol, 0.7 mM MgCl2, 0.4 mM EDTA, 0.4 mM DTT, 37 mM NaCl, 0.7 mM Tris–HCl pH 7.5). In competition EMSA, 33-fold molar excess of unlabeled allele-specific probes and probes with perfect binding sites for YY1, MyoD, CdxA and a scramble oligonucleotides was added to the reaction, prior to addition of Cy5-labeled probes. Binding reactions were incubated for 20 min at 4°C. In supershift experiments, nuclear extracts were pre-incubated with 0.8 μg of anti-YY1 (sc-281x, Santa Cruz Biotechnology, Heidelberg, Germany) or isotype control IgG (sc-2027, Santa Cruz Biotechnology, Heidelberg, Germany) antibodies, respectively, for 30 min at 4°C. All EMSA experiments were replicated at least three times.

Affinity chromatography using magnetic beads

To isolate and enrich allele-specific binding proteins, we performed magnetic beads-based affinity chromatography. According to the manufacturer’s instructions, streptavidin coupled magnetic beads (Dynabeads M-280, Invitrogen, Darmstadt, Germany) were washed and collected using Bind&Wash buffer (Dynabeads M-280, Invitrogen, Darmstadt, Germany) and Magnetic particle separator (Magnetic-1™, Dynal, Invitrogen, Darmstadt, Germany) or isotype control IgG (sc-2027, Santa Cruz Biotechnology, Heidelberg, Germany) antibodies, respectively, for 30 min at 4°C. All EMSA experiments were replicated at least three times.

Oligonucleotides probes

40 bp oligonucleotides rs4684847 5'-TTAAATATCATCTCAATTCGAAAAAGATAAG-3'; rs7647481 5’-CAACTCCCACCTTTATCC[A/G]GATGTGGACTGACCAGCCA-3'; rs17036342 5’-GCTTCCCAAAGAATGTA[A/G]TCCAGAG GTAGGACACA-3'; rs2881479 5’GCAAGACTCTGTCTCAAATAAATATAATAAA-3' with Cy5- (for EMSA) or biotin-label (for affinity chromatography) and with the respective SNP alleles at midposition were synthesized (Eurofins, Ebersberg, Germany) and annealed with unmodified complementary oligonucleotides to obtain double stranded oligonucleotides, by
a rotator. We used 7 mg of nuclear extracts, which showed the most significant enrichment of proteins of interest. To reduce non-specific DNA binding, poly (di-dC) was subsequently added to the mixture and incubated further for 10 min. Competition with poly (di-dC) supports the specific binding of proteins to the oligonucleotides. Subsequently, the beads were washed three times with binding buffer containing 10 mM NaCl and the bound proteins were eluted by 1× binding buffer with increasing concentration of NaCl (200 and 300 mM) in a volume of 100 μl (eluates E200 and E300). All steps were performed at 4°C. Finally, 5–10 μl of protein from supernatants, washes and eluates were used for EMSA. If eluates showed enrichment of proteins of interest, the remaining volumes were subjected to LC–MS/MS analysis. All affinity chromatography experiments were replicated at least three times.

**Filter-aided sample preparation (FASP) and non-targeted LC–MS/MS**

Salt eluted fractions were processed as described before (10,27) in an adaptation of the FASP approach (28) using Microcon devices YM-30 (Merck Millipore, Darmstadt, Germany). The LC–MS/MS analyses were performed as described previously on a LTQ-Orbitrap XL (Thermo Scientific, Dreieich, Germany) with the following adjustments: A nano trap column was used (300 μm inner diameter × 5 mm, packed with Acclaim PepMap100 C18, 5 μm, 100 Å; LC Packings) before separation by reversed phase chromatography (PepMap, 25 cm, 75 μm ID, 2 μm/100 Å pore size, LC Packings, Thermo Scientific, Dreieich, Germany) operated on a RSLC (Ultimate 3000, Thermo Scientific, Dreieich, Germany) using a nonlinear 170 min LC-gradient from 5 to 31% of buffer B (98% acetonitrile at 300 nl/min flow rate followed by a short gradient from 31 to 95% buffer B in 5 min and an equilibration for 15 min to starting conditions. From the MS pre-scan (acquired in profile mode), the 10 most abundant peptide ions were selected for fragmentation in the linear ion trap if they exceeded an intensity of at least 200 counts and if they were at least doubly charged. Dynamic exclusion was set to 30 s. During fragment analysis, a high-resolution (60 000 full-width half maximum) MS spectrum was acquired in the Orbitrap with a mass range from 300 to 1500 Da.

**Protein identification and label-free relative quantification**

The RAW files (Thermo Scientific, Dreieich, Germany) were analyzed using the Progenesis LC–MS software (version 4.0, Nonlinear Dynamics, Waters, Eschborn, Germany), as described previously (29,30), with the following changes: Spectra were searched using the search engine Mascot (Matrix Science, London, UK) against the Ensembl mouse database (Release 69; 50512 sequences). False discovery rates were stringently kept below 1% as calculated by a Mascot-integrated decoy database search using the percolator algorithm (cut-off score 13, significance threshold of P < 0.05). Peptide assignments were re-imported into Progenesis LC–MS. Normalized abundances of all unique peptides were summed up and allocated to the respective protein.

**Allele-specific YY1 Chromatin Immunoprecipitation assay (ChIP)**

For each experiment, 1 × 10⁶ primary human preadipocytes as well as preadipocytes in vitro differentiated to adipocytes for 14 days were cultured in six wells of a six-well plate as previously described (6). ChIP experiments were performed using the ChIP-IT® Express Enzymatic Chromatin Immunoprecipitation Kit from Active Motif (La Hulpe, Belgium) according to the manufacturer’s protocol with slight modifications as described elsewhere (31). Briefly, after enzymatic digestion for 15 min, 10 mM EDTA was added and chromatin was sheared using the EpiShear™ Probe Sonicator (Active Motif, La Hulpe, Belgium; 20 pulses consisting of 20 s sonication followed by 30 s rest at 25% amplitude) in the same buffer. Chromatin was then incubated for 30 min with protein G magnetic beads (Active Motif, La Hulpe, Belgium) and 2 μg of rabbit polyclonal αYY1 antibody (sc-281x, Santa Cruz Biotechnology, Heidelberg, Germany). Incubation with 2 μg of rabbit IgG (sc-2027x) served as internal negative controls. The amount of precipitated DNA was evaluated by allele-specific quantitative PCR (AS-qPCR) using the Eppendorf Mastercycler (Eppendorf, Hamburg, Germany). To quantify allele-specific protein binding we performed SYBR-green qPCR (Maxima SYBR Green/ROX qPCR Master Mix, Thermo Scientific, Dreieich, Germany) using rs7647481 forward primer (5′-AAGATGTTTTGGGGCTTAATGG-3′) with the allele-specific reverse primers rs7647481A nonrisk (5′-G CTGGGTCTGAACATCATA-3′) and G-risk (5′-CTG GGTCGAAACATCACA-3′), respectively (bold: SNP position, underlined: additional mutation). Allele-specific reverse primers were designed (TIB Molbiol, Berlin, Germany) with the respective rs7647481-allele and an additional mutation to increase allele-specificity as previously described (32). Allele-specificity was tested using a 611 bp DNA fragment containing rs7647481A nonrisk and G risk allele, respectively; primer efficiencies were calculated using REST 2009 software (www.gene-quantification.de/rest-2009.html). The rs7647481 allele-specific protein-chromatin interaction at the A nonrisk/G risk allele (Figure 4E) was determined by calculating ΔCt (A) and ΔCt (G) for A- and G-allele by subtracting the input-chromatin Ct-values from respective ChIP-chromatin Ct-values for both anti-YY1 and IgG experiments, the allele-specific ratio for each antibody based on ΔΔCt method (here (primer efficiency A-allele)/ΔCt (G) / (primer efficiency G-allele)/ΔCt (A)) and finally the ratio to the respective IgG control for each experiment. Pairs of anti-YY1 and IgG ChIP experiments were performed from fixed chromatin of preadipocytes and in vitro differentiated adipocytes from three donors, which were previously genotyped with a concordance rate of >99.5% using the MassARRAY system with iPLEX™ chemistry (Sequenom, Hamburg, Germany) as described (33).

**siRNA knock down and PCR**

SGBS cells were cultured in 6-well plates and transfected with 25 nM siRNA targeting YY1, RYBP or as control non-targeting (NT) siRNA (ON-TARGETplus human siRNA SMARTpool, Dharmacon, Freiburg, Germany)
using HiPerFect (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. Seventy two hours after transfection, confluent cells were harvested using the RNasy-Minikit (Qiagen, Hilden, Germany) to extract total RNA and stored at −80°C. The high capacity cDNA Reverse Transcription kit (Applied Biosystems, Darmstadt, Germany) was used for transcription of 1 μg total RNA into cDNA according to the manufacturer’s instructions and stored at −20°C. RNA/cDNA concentration and purity (A260/280) was assessed using NanoQuant Plate™ (TECAN, Männedorf, Switzerland). qPCR analysis of human PPARG1, PPARG2 isoforms transcripts (6), GAPDH housekeeping gene, YY1 and RYBP to control for knockdown efficiency, was performed using the qPCR SYBR-Green ROX Mix (ABgene, Hamburg, Germany) with an initial activation of 15 min according to the manufacturer’s instructions, using a Mastercycler Realplex 4 and Realplex Software (Eppendorf, Hamburg, Germany) with an initial activation of 15 min at 95°C followed by 40 cycles of 15 s at 95°C, 30 s at 60°C and 30 s at 72°C. Amplification of specific transcripts was confirmed by initial sequencing, and melting curve profiles (cooling samples to 68°C and heating slowly to 95°C with measurement of fluorescence) and by agarose gel electrophoresis to assess the size of PCR products. PPARG1 (NM_138712.3): forw5'-CGTGCCCAGATTTGAGA-3' + rev5'-AGTGGGAATTCCCTGAACTAC-3' = 177 bp; PPARG2 (NM_015869.4): forw5'-GAAAGCGATCTCTCAGGTGAT-3' + rev5'-TAAAGGAGTGGGAGTGTC-3' = 146 bp; GAPDH (NM_002046.5): forw5'-GATGATCGAATGCCTCCTGC-3' + rev5'-ACAGTCTTGTTGGCGATGA-3' = 128 bp; RYBP (NM_012234.6): forw5'-CTGCACCTTTCACTGATGTT-3' + rev5'-GTGCCACCGCTGGAAGG-3' = 181 bp (specificity of primer tested by BLAST). Mean target mRNA levels, standardized to gene expression levels of the housekeeping gene, from five independent biological replicates of knockdown experiments, each PCR measured separately, were calculated by the formula 2^(-ΔΔCt) method relative to the siNT control experiment. P-values were calculated using one-sample t-test.

Luciferase reporter gene assays

To assess transcriptional activity mediated by SNP-adjacent regions, luciferase reporter gene assay was performed as described previously (6) with the following modifications. C2C12 cells cultured in 48-well plates to approximately 80% confluence were differentiated for 7 days as described above. C2C12 cells (undifferentiated myoblasts, differentiated myocytes), 293T, INS-1 β and Huh7 cells at 80–90% confluence were transfected in 48-well plates by Lipofectamine 2000 transfection reagent (Invitrogen, Darmstadt, Germany). 293T, Huh7, and C2C12 (undifferentiated and differentiated) were transfected with 0.3 μg of the respective luciferase reporter vector, 0.04 μg of the ubiquitin promoter vector and 1 μl of Lipofectamine 2000 reagent. INS-1 β-cells were transfected with 1.2 μg of the respective luciferase reporter vector, 0.16 μg of the ubiquitin promoter vector and 2 μl of Lipofectamine 2000 reagent. 3–4 h after transfection the medium was replaced by fresh medium followed by incubation at 37°C. Twenty four hours after transfection, the cells were washed once with PBS and lysed in 1× passive lysis buffer (Promega, Mannheim, Germany) on rocking for 20 min at room temperature. Firefly and renilla luciferase activity were measured using luminometer (Berthold, Pforzheim, Deutschland), respectively. The ratios of firefly luciferase expression to renilla luciferase expression were calculated and normalized to the thymidine kinase promoter control vector. All experiments were performed repeatedly as indicated. P-values were calculated using paired t-test.

GO-term, signaling pathway and transcription factor/transcriptional coregulator co-citation analysis

The Genomix GePS-tool (Genomix, Munich, Germany) was used to assess the enrichment of molecular function GO-terms (Supplementary Table S2) and signaling pathways (Supplementary Table S3) in the protein/gene data sets identified by LC-MS/MS, using all identified proteins. Next, to calculate the enrichment of transcriptional coregulators, we used the Genomix GePS-tool to build a co-citation based network. All proteins annotated as co-factors (Supplementary Table S4) and the respective candidate transcription factors PRRX1, YY1 and NFATC4 were used as input gene list to create networks using the settings co-citation level: Function word level; co-citation filter: (i) Network generation: simple network; and additional interactions per gene: (ii) Genes with interactions are shown in the Figure 5A and Supplementary Figure S5A-C.

Regression analysis of human adipose tissue samples

The insulin-resistance measure HOMA-IR (homeostasis model assessment of insulin resistance) and mRNA expression levels were measured in a cohort comprising 30 obese (BMI > 30 kg/m²) otherwise healthy and 26 non-obese (BMI < 30 kg/m²) healthy women (34), all pre-menopausal and free of continuous medication and investigated in the morning after an overnight fast. Venous blood sample was obtained for measurements of glucose, insulin and for preparation of DNA. HOMA-IR was calculated by the formula FPGlucose (mmol/L) × (fS-Insulin (microU/ml)/ 22.5) (35). Following blood sampling abdominal subcutaneous adipose tissue biopsy was obtained by needle aspiration and adipose microarray analysis was performed exactly as described (34) using the Affymetrix GeneChip Array protocol with 1 μg of total adipose RNA from each subject (gene expression deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO; https://ncbi.nlm.nih.gov/geo/) and accessible using GEO series accession number GSE25402). Linear regression analyses were performed with R, version 3.0.2 (R: A Language and Environment for Statistical Computing: from the R Development Core Team of the R Foundation for Statistical Computing, Vienna, Austria, 2014, http://www.R-project.org/) to assess correlation of YY1, RYBP and PPARG adipose tissue mRNA levels with HOMA-IR (adjusted for age, age/BMI and without


**RESULTS**

Risk and nonrisk allele-specific protein–DNA interaction at the cis-regulatory variants rs4684847 and rs7647481

At the *PPARG* locus, with a total of 23 non-coding variants in high linkage disequilibrium (*r*² ≥ 0.7 with *PPARG* tagSNP rs1801282 (6)), six variants were predicted to be cis-regulatory by conserved binding site modularity analysis (Figure 1B). We showed previously that rs4684847 specifically overlaps with a homeobox TFBS and modulates regulation of *PPARG* expression (6). Further assessing proximity of the remaining five variants to marks of regulatory regions in adipocytes, we find rs7647481 to overlap with all histone modification regions (36) at both tested late stages (day 3 and day 9) of human adipogenesis (Figure 1C), and with DNase-seq regions (2) in differentiated primary human adipocytes (Supplementary Figure S1); supporting a contribution of rs7647481 to adipocyte specific *PPARG* expression.

---

**Figure 2.** Enrichment of risk and nonrisk allele-specific binding proteins at predicted cis-regulatory SNPs. (A) Representative EMSA experiments with allele-specific Cy5-labeled probes on nuclear extracts from HIB 1B cells (triangle = allele-specific band) demonstrated allele-specific differential binding affinity of proteins at the risk/nonrisk allele of predicted cis-regulatory rs4684847/rs7647481 variants (red), respectively, and no binding at predicted non-cis-regulatory SNPs (grey). Bar charts illustrate the allelic fold change of protein–DNA complexes signal intensity (allele with highest binding/allele with lowest binding; allele with lowest binding set to one), mean ± SD of five experiments. *P < 0.05. P-value by paired t-test. (B) Risk and nonrisk allele protein–DNA interaction at predicted cis-regulatory and non cis-regulatory SNPs in human preadipocytes and adipocytes. EMSA with allelic Cy5-labeled probes for the indicated predicted cis-regulatory (red) and non cis-regulatory (grey) SNPs using nuclear extracts from undifferentiated primary human preadipocytes (left panel), the human SGBS preadipocyte cell line (mid panel) and SGBS cells differentiated to adipocytes for 14 days (right panel). (C) Enrichment of allele-specific differential binding proteins; EMSA with binding-allele specific Cy5-labeled probes of predicted cis-regulatory SNPs using protein from affinity chromatography with the respective biotin-labeled risk/nonrisk allelic-probes. Triangle = allele-specific band; input: nuclear protein used for affinity chromatography; Sn: supernatant after incubation with biotin-labeled allelic-probe-magnetic beads conjugates; Wash: low NaCl concentration wash eluates; E200/E300: 200 and 300 mM NaCl protein eluate used for LC-MS/MS. Protein eluates E200 and E300 with differential protein–DNA binding contain the prioritized transcription factors YY1 at rs7647481A nonrisk and PRRX1 at rs4684847C risk allele (Table 1). All experiments were performed in triplicates. For enrichments at predicted non cis-regulatory SNPs, see Supplementary Figure S2.
expression (Figure 1D). Both, rs7647481 and the previously reported rs4684847 (6) are in perfect linkage disequilibrium ($r^2 = 1.0$ in 1000 Genomes (37)), confirmed by sequencing in this study for all analysed tissues and cells (data not shown). To experimentally evaluate these cis-regulatory predictions, we further analysed allele-specific protein–DNA interactions at both predicted cis-regulatory variants and additionally at two variants predicted as non cis-regulatory (Figure 1D).

Electrophoretic mobility shift assays (EMSA) using nuclear extracts of mouse brown adipocytes, revealed allele-specific protein–DNA interaction for both predicted cis-regulatory SNPs, not observed with the predicted non cis-regulatory SNPs (Figure 2A upper panel). Quantification of protein–DNA complexes confirmed allele-specific binding at both predicted cis-regulatory SNPs ($P = 0.03$) in contrast to non cis-regulatory SNPs ($P = 0.82$ for rs17036342 and $P = 0.80$ for rs2881479, respectively) (Figure 2A lower panel). PPARG regulates adipogenesis and mature adipocyte metabolism. In EMSA experiments using nuclear protein extract from primary human preadipocytes, human SGBS cell strain preadipocytes and in vitro differentiated SGBS adipocytes, we find differential protein–DNA interaction patterns (Figure 2B), indicative for different transcription factors and regulatory protein complexes contributing to PPARG expression in different stages of adipogenesis. Allele-specific differential binding was observed consistently at predicted cis-regulatory variants, with increased binding at rs7647481A nonrisk and rs4684847C risk allele, respectively.

**Identification of allele-specific binding proteins at predicted cis-regulatory variants by label-free quantitative proteomics**

Next, we aimed to identify allele-specific binding proteins, i.e. regulators and coregulators at those variants by an unbiased, quantitative label-free protein–DNA proteomics (Figure 1A-5). To enrich allele-specific binding proteins for i.e. regulators and coregulators at those variants by an LC–MS/MS found 41–108 and 142–165 proteins (Figures 3 and Supplementary Figure S2). The comparison of predicted cis- versus non cis-regulatory variants, we found a significant enrichment of differentially binding proteins at cis-regulatory variants ($1.87 \times 10^{-3} \leq P \leq 1.72 \times 10^{-6}$, two-sided, two-group binomial test for pairwise comparison of differentially binding proteins, Supplementary Table S1). The comparison of predicted cis- versus cis-regulatory and non cis- versus non cis-regulatory SNPs revealed no significant enrichment for most pairs ($4.26 \times 10^{-4} \leq P \leq 0.28$, Supplementary Table S1). Thus, the highest numbers of allele-specific binding proteins were found at predicted cis-regulatory SNPs supporting specific protein–DNA interaction (Figure 2A). Moreover, when assessing GO-terms for allele-specific binding proteins (fold change $\geq 2$, $P \leq 0.05$), we found a significant enrichment in the GO-terms for allele-specific binding proteins ($P = 1.36 \times 10^{-6}$, $P = 1.44 \times 10^{-7}$) and structure-specific DNA binding proteins ($P = 2.11 \times 10^{-5}$, $P = 3.69 \times 10^{-8}$) at the predicted cis-regulatory variants rs4684847 and rs7647481 in contrast to low respective GO-term enrichment at predicted non cis-regulatory vari-
Table 1. Allele-specific differentially binding transcription factors

| SNP   | Gene symbol | Allelic ratio | Allelic fold change | P-value (fold change) | Quantified peptides |
|-------|-------------|---------------|---------------------|-----------------------|---------------------|
| rs7647481 | YY1          | A/G           | 6.6                 | 2.94×10^{-3}          | 9                   |
|        | NFATC4      | C/T           | 2.6                 | 0.01                  | 2                   |
| rs4684847 | PRRX1       | C/T           | 2.6                 | 0.01                  | 5                   |
|        | ILF3         |               | 4.2                 | 0.01                  | 4                   |

Proteins annotated as transcription factors and identified by LC–MS/MS in the fractions with the highest allelic protein–DNA binding EMSA signal intensity after affinity-chromatography enrichment at rs7647481A nonrisk and rs4684847C risk allele (Figures 2C, 200 and 300 mM elution, respectively) and significant differential binding (allelic fold change ≥ 2.0, P-value ≤ 0.01, illustrated in Figure 3 and Supplementary Figure S3) are shown.

YY1 drives transcriptional activity at the rs7647481A nonrisk allele of the PPARG locus

The common rs7647481G risk allele abrogates the core of a YY1 consensus TFBS, confirming the mass spectrometric identifications and GO-term analyses (Figure 4A). The allele-specific protein–DNA interaction at the rs7647481-adjacent region was efficiently blocked in competition and supershift EMSA experiments by 33-fold molar excess of unlabeled YY1 consensus binding sequence or by pre-incubation with a YY1 specific antibody (Figure 4B left panel), while protein binding was not affected at all other tested SNP-adjacent regions, including the cis-regulatory variant rs4684847 (Figure 4B). Competition with unspecific competitor oligonucleotides (consensus MyoD myogenic regulatory factors, consensus CdxA chicken homeodomain protein, and scrambled control sequence) did not affect the allele-specific protein binding (Supplementary Figure S4A), further confirming the specificity of YY1 binding at the rs7647481-adjacent region. PPARG expression is essential for preadipocyte differentiation and metabolic function of mature adipocytes. By YY1 competition and supershift experiments we confirmed YY1 binding at the rs7647481A nonrisk allele in both, 3T3-L1 preadipocytes and adipocytes (Supplementary Figure S4B). In reporter gene assays with luciferase constructs of all tested variants transfected into 293T cells, overexpression of the transcription factor YY1 revealed a significant higher transcriptional activity at the rs7647481A nonrisk allele as compared to the risk allele (P = 0.003, Figure 4C). Although we found an overall increase of all tested reporters, no allele-specific effects of the cis-regulatory variant rs4684847 and both non cis-regulatory variants were observed. The rs7647481A nonrisk variant also increased transcriptional activity in different cell types significantly, i.e. by 1.2-fold in 293T cells, 1.4-fold in INS1 β-cells, 2.7-fold in C2C12 myoblasts, 2.2-fold in C2C12 myocytes, 1.3-fold in Huh7 hepatocytes (P < 0.01, Figure 4D) and 1.5-fold in 3T3-L1 adipocytes (6).

Next, we aimed to confirm in vivo allele-specific binding of YY1 at the rs7647481A nonrisk allele, as results from reporter assays are limited by missing chromosome context or length of analysed sequences. We performed chromatin immunoprecipitation (ChIP) combined with allele-specific quantitative PCR (AS-qPCR) to test for allelic imbalance in primary human adipose tissue cells heterozygous for rs7647481G/A. ChIP experiments were performed using a YY1-specific antibody to pull down cross-linked YY1-protein / chromatin complexes from primary human preadipocyte and adipocyte cells of three donors (Figure 4E). In both, input chromatin and immuno-precipitated chromatin, the nonrisk and risk allele, respectively, were analyzed using AS-qPCR. In adipocytes we observed a significant mean 31.8-fold (P = 0.02, Wilcoxon signed rank test) increased binding of YY1 at the rs7647481A nonrisk allele as compared to the risk allele in chromatin immunoprecipitated with anti-YY1. In preadipocytes, we observed a large variation of YY1 binding at the nonrisk allele but no significant allele-specific binding. No allele-specific differences were observed when using an IgG control antibody. Overall, YY1 supershift (Figure 4B, Supplementary Figure S4B) and ChIP (Figure 4E) experiments in preadipocytes and adipocytes establish binding of the transcription factor YY1 at the rs7647481A nonrisk allele, and support its role in transcriptional regulation of the PPARG gene in both stages of adipogenesis.

Identification of transcription factor related transcriptional coregulators

Metabolic homeostasis is largely regulated at the transcriptional level through the coordinated interaction between transcription factors, coregulators, and the basal transcriptional machinery (11). Our pull down of protein–DNA binding complexes offers the opportunity to identify functional protein–DNA interactions as we found both, transcription factors (Table 1) and co-eluting transcriptional coregulators (Supplementary Table S4). We assessed relevant protein-protein interaction using literature co-citation...
Figure 4. rs7647481A nonrisk allele-specific binding and transcriptional activity of the transcription factor YY1 inferred from proteomics analysis. (A) The rs7647481G risk allele abrogates the core of a YY1 consensus binding site (Matbase Matrix Library 9.1, Genomatix, Munich, Germany). (B) Competition and supershift EMSA experiments using risk (R) and nonrisk (NR) allele-specific Cy5-labeled probes of the predicted cis-regulatory (red) and non cis-regulatory (grey) variants reveal a specific binding of YY1 at the rs7647481A nonrisk allele. Competition (comp.) assays using 33-fold excess of unlabeled YY1 probe and supershift assays by adding anti-YY1 or lgG isotype control antibody, respectively. (C) Reporter assays in 293T cells with constructs harbouring the risk and nonrisk allele of predicted cis-regulatory (red) and non cis-regulatory (grey) variants reveal allele-specific activation from the rs7647481A nonrisk allele upon YY1 overexpression. Mean ± SD from five experiments. (D) rs7647481A nonrisk allele-specific activation of reporter gene activity in 293T-cells, INS-1 β-cells, C2C12 cells (undifferentiated myoblasts, differentiated myocytes) and Huh7 hepatocytes assessing the effect of endogenous transcriptional regulators. Reporter assays with luciferase constructs containing the respective allele at midposition as indicated; for each cell line the TK-promoter control vector was co-transfected separately and set to one. Mean ± SD from seven experiments. (E) Increased in vivo YY1 binding at the rs7647481A nonrisk allele. The result shows allele-specific binding at the A-risk/G-risk allele, determined by ChIP experiments in primary human adipose tissue cells,preadipocytes and in vitro differentiated adipocytes, heterozygous for rs7647481G/A using αYY1 or IgG isotype control, respectively, followed by allele-specific qPCR detecting the rs7647481A nonrisk and G risk allele for each ChIP experiment (see also Materials and Methods). Mean ± SD from ChIP experiments using chromatin–DNA from three donors, P-values from Wilcoxon’s signed rank test.
Allele-specific correlation of transcription factor and cofactor expression levels in adipose tissue with insulin resistance

Next, we assessed if the risk and nonrisk allele-specific proteomics findings can be related to disease pathophysiology. The minor nonrisk allele of the PPARG locus (tagSNP rs1801282 Pro12Ala) was repeatedly associated with improved insulin-sensitivity in diverse studies. Thus, allele-specific PPARG expression may contribute to this phenotype, given the essential role of PPARG in maintaining insulin-sensitivity. In fact, testing allele-specific PPARG mRNA levels in adipose tissue samples, we observed a confirmatory age- and age/BMI-independent negative correlation of total PPARG mRNA levels with the insulin resistance measure HOMA-IR in nonrisk allele carriers ($\beta = -6.25, P = 2.18 \times 10^{-4}; \beta = -3.26, P = 0.05$, respectively) as compared to risk allele carriers ($\beta = 0.23, P = 0.89; \beta = 0.11; P = 0.92$, respectively, Table 2). A coordinated regulation of PPARG expression by YY1 and co-identified coregulators at the rs7647481A nonrisk allele may contribute to phenotypes—such as insulin-resistance—associated with the PPARG locus. Here, for adipose tissue mRNA expression levels of the coregulator RYBP, identified by proteomics and reported to interact with the YY1 transcription factor (39), we found a negative age- and age/BMI-independent correlation with HOMA-IR in individuals carrying the nonrisk ($\beta = -5.71, P = 1.15 \times 10^{-3}; \beta = -3.38, P = 7.04 \times 10^{-7}$; respectively) as compared to risk ($\beta = 0.57, P = 0.67; \beta = 0.16, P = 0.85$, respectively) allele. For none of the other proteins co-cited with YY1 or NFATC4 (Supplementary Table S5) an allele-dependent and BMI-independent correlation was observed. We found a BMI-dependent correlation of PHB, CALR and NPM1 expres-
sion (also found at rs4684847C risk allele with PPRRX1 and ILF3), requiring further analysis in the future. Taking into account the impact of BMI on insulin-sensitivity, our data support the literature-mining based initial focus on YY1 and RYBP. While we found no significant correlation of YY1 mRNA levels with HOMA-IR in the small available data set, a confirmative direction of \( \beta \)-values was observed (Table 2). Finally, we also found a positive correlation for adipose mRNA expression levels of both, YY1 and RYBP, with total mRNA levels of the insulin-sensitizing transcription factor PPARG from both alleles (Supplementary Table S5). YY1 and RYBP can function as repressor or activator depending on genomic context and availability of interaction partners (38,40). Assessing the effect on endogenous mRNA expression levels in SGBS preadipocytes, we found that knockdown of YY1 and its cofactor RYBP alone (Figure 5B, 63% or 60% knockdown efficiency, \( P = 7 \times 10^{-3} \) or \( 4 \times 10^{-4} \), respectively) resulted in a slight but not significant 1.6-fold increase of PPARG2 expression (Figure 5C) suggesting inhibitor-activity. However, in 293T cells YY1 over-expression significantly increased rs7647481A nonrisk allele reporter activity (Figure 4C) suggesting activator-activity, possibly by recruitment of endogenously expressed RYBP (Supplementary Figure S5D). Therefore, we performed simultaneous knockdown of YY1 and RYBP (Figure 5B, 68% or 73% knockdown efficiency, \( P = 3 \times 10^{-3} \) or \( 2 \times 10^{-3} \), respectively) in the PPARG1 and PPARG2 expressing adipose tissue cell line SGBS. We found a significant 2-fold reduction of endogenous mRNA expression levels for the insulin-sensitizing PPARG2 isoform (\( P = 0.027 \), Figure 5C) suggesting that YY1 and RYBP jointly activate PPARG2 expression and supporting the importance to account for both, transcription factors and related cofactors. How further cofactors identified by our unbiased proteomics approach (Supplementary Table S4) contribute to activation or inhibition of PPARG gene expression requires future experiments. Overall, our findings suggest that PPARG expression, and thereby insulin sensitivity, may be increased by rs7647481A nonrisk allele-specific binding of the transcription factor YY1 and its cofactor RYBP (Figure 5D).

**DISCUSSION**

GWAS revealed numerous risk loci associated with common traits (1). Advances of the ENCODE project and novel bioinformatics approaches facilitate identification of cis-regulatory, potentially diseases-causing variants within complex loci (2–7). However, the spatial and temporal varied expression pattern of transcription factors and coregulators supports the need to consider cell-type specific open chromatin data (24–45) to prioritize candidate cis-regulatory variants. In our study, combining computational TFBS modularity analysis (6) and functional cell type- and differentiation-specific data from human adipocytes (36) enabled identification of a second cis-regulatory variant at the PPARG locus, additional to a previously reported (6). Our unbiased quantitative label-free proteomics approach came up with significantly more allele-specific binding proteins at predicted cis-regulatory as compared to non cis-regulatory variants, supporting the integrative framework predictions. Moreover, we find both, allele-specific transcription factors and cofactors which may contribute to the rs7647481A nonrisk allele-specific association of the PPARG locus with insulin sensitivity.

Identifying the allele-dependent binding proteins is an essential step to further decipher the molecular mechanisms underlying genetic associations with disease pathophysiology and ultimately to define personalized interventions. However, identification of allele-specific binding proteins by TFBS matrix overlap or ChIP-seq faces limitations, such as availability of TFBS matrix annotation and the complexity of TFBS modularity. So far, few proteomic studies have provided successful transcription factor identification at cis-regulatory variants (12,13,46). Here, taking into account the importance of TFBS modularity for specific protein-DNA interaction (6,47,48), we assessed protein binding directly at the genomic regions of interest, instead of e.g. concatenated oligonucleotides (13). Peptide intensity-based label-free quantification across samples was previously benchmarked against SILAC labeling and compared well with respect to reproducibility, sensitivity, and robustness (14). In contrast to metabolic (13,46) or chemical labeling strategies (12) our label-free approach does not limit future application to any tissue or cell type, including primary human tissue. Our proteomics workflow on eluted fractions containing allele-specific binding-proteins allows identification of allele-specific binding transcription factors. Moreover, our workflow enables for the first time the identification of transcriptional coregulators related to allele-specific gene regulation, which are among the lowest abundant proteins in cells (42,49).

The PPARG locus is robustly associated with type 2 diabetes and insulin-sensitivity (16–18,50,51) and mRNA levels of the insulin-sensitizing PPARG are increased in non-risk allele carriers (6). At the here identified rs7647481A nonrisk allele, proteomic analysis identified allele specific binding of two transcription factors, YY1 and NFATC4. YY1 was reported to regulate metabolic, diabetes-related phenotypes in skeletal muscle, liver and adipocytes (52–56). NFATC4 contributes to regulation of PPARG and mouse adipocyte differentiation by direct binding at the PPARG2 promoter (57), YY1 indirectly by interaction with C/EBPα (56). We focused on the allele-specific activity of human YY1 and found in vivo allele-specific binding at the rs7647481A nonrisk allele in the human PPARG promoter by ChIP experiments combined with assessment of allelic imbalance in heterozygous primary human adipose tissue cells, supporting the biological relevance of the proteins identified by our label-free proteomics approach. At the rs4684847C risk allele, we previously reported an overlapping homeobox TFBS, risk allele specific binding of the transcriptional inhibitor PPRRX1 repressing PPARG expression and promoting insulin-resistance (6). Our unbiased proteomics confirmed risk allele binding of PPRRX1 and also found ILF3, a transcription factor so far not related to adipose PPARG gene regulation or diabetes. The regulation of PPARG expression is complex, and future experiments are needed to support contribution of further factors, such as NFATC4 and ILF3 at rs7647481 and rs4684847, respectively, to allele-specific PPARG expression in human adipose tissue, but also for possible interactions of PPRRX1 with ILF3 or NFATC4 with YY1.
In addition to transcription factors, the high sensitivity of the applied proteomics approach enabled identification of several cofactors which may be involved in modulating transcriptional regulation of PPARG, by yet to be characterized mechanisms in future studies. For proof-of-concept that our allele-specific proteomics can find both, allele-specific transcription factors and transcriptional cofactors, we selected the transcription factor YY1 with the highest allele-specific binding and the co-citation enriched RYBP cofactor at rs7647481. The coregulator RYBP—known to interact with YY1 and here co-identified at the rs7647481 variant—was recently shown to associate with skeletal myogenesis (38), in addition to its function as transcription repressor in cancer (58,59), embryogenesis (39) and central nervous system development (60). While YY1 has been reported previously to promote adipocyte differentiation (61), we found that the combined action of the transcription factor YY1 and the cofactor RYBP is necessary for full activation of PPARG2 isofrom expression in an adipocyte cell line. PPARG2 has been reported to be crucial for maintaining insulin-sensitivity (20). In human samples we demonstrated a nonrisk allele-specific association of RYBP expression levels with the PPARG nonrisk allele associated phenotype insulin-sensitivity. Thus, in carriers of the protective PPARG allele, binding of YY1 and RYBP at the rs7647481A nonrisk allele may increase PPARG2 expression and thereby insulin sensitivity (Figure 5D, upper panel); and in PPARG risk allele carriers the previously reported rs4684847C risk allele binding of PRRX1 inhibits PPARG expression thereby promoting insulin-resistance (6) (Figure 5D, lower panel).

In conclusion, our results demonstrate the importance to consider allele-specific protein-complexes for delineation of the molecular mechanisms affected by cis-regulatory variants. We present an approach enabling an unbiased identification of allele-specific protein–DNA interactions including both, transcription factors and transcriptional cofactors. Additionally, we provide data supporting that different cis-regulatory variants at the PPARG locus may contribute to disease pathophysiology (Figure 5D), in line with the ‘multiple enhancer hypothesis’ suggesting multiple regulatory SNPs per LD block (62). Integrative approaches combining computational and cell type specific histone / chromatin mark based cis-regulatory prediction in combination with highly sensitive proteome-wide identification of allele-specific binding proteins can help to clarify the phenotypic effects of inherited and somatic genetic variability.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We thank M. Arnold and C. Fuchs for statistics, G. Leboulle for advice with allele-specific primer design, E. Hofmair, C. Herrmann, J. Behler, S. Becker and M. Hubersberger for technical assistance. We acknowledge the
ENCODE Consortium and the John Stamatoyanopoulos (UW) ENCODE production laboratory generating the datasets ENCF013JN and ENCF020FZ0.

Authors Contributions: Conceived and designed the experiments: H. Lee, H.-L. S.M.H.; performed the experiments: H. Lee, K.Q., L.H., F.E., L.J., C.H., Cv.T., V.G., M.B.; analyzed the data: H. Lee, H.L., C.B., S.M.H., Cv.T., S.W., S.M., H.G., M.C.; contributed reagents/materials/analysis tools: I.D., C.B., P.A., H.H., H.G.; wrote the paper: H. Lee, H.L., S.M.H., Cv.T.; revised the manuscript: C.H., K.Q., C.B., L.H., V.G., S.W., M.B., M.C., S.M., H.G., I.D., P.A., H.H. All authors approved the final version.

FUNDING

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, Munich, Germany; a grant from the German Federal Ministry of Education and Research to the German Centre for Diabetes Research (DZD e.V.); and the grant LA2595/3-1 from the German Research Foundation (DFG). Funding for open access charge: Grant Clinical Cooperation Group ‘Nutrigenomics and type 2 diabetes’ received from the Helmholtz Zentrum München and the Technische Universität München.

Conflict of interest statement. None declared.

REFERENCES

1. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U.S.A., 106, 9362-9367.

2. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 56-74.

3. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. and Snyder, M. (2012) Linking disease associations with regulatory information in the human genome. Genome Res., 22, 1748-1759.

4. Edwards, S.L., Beesley, J., French, J.D. and Dunning, A.M. (2013) Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet., 93, 779-797.

5. Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E. et al. (2013) Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res., 41, 827-841.

6. Claussnitzer, M., Danzel, S.N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, O., Soskolne, N., Fadista, J., Ehlers, K. et al. (2014) Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell, 156, 343-338.

7. Ritchie, G.R., Dunham, I., Zeggini, E. and Flieck, P. (2014) Functional annotation of noncoding sequence variants. Nat. Methods, 11, 294-296.

8. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M. et al. (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature, 466, 714-719.

9. Praetorius, C., Grill, C., Stacey, S.N., Metcalf, A.M., Gorkin, D.U., Robinson, K.C., Van Otterloo, E., Kim, R.S., Bergstein-Dottir, K., Ogmundsdottir, M.H. et al. (2013) A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. Cell, 155, 1022–1033.

10. Spieler, D., Kaffe, M., Knauf, A., Bessa, J., Tenza, J.J., Giesert, F., Schormair, B., Tich, E., Lee, H., Horsch, M. et al. (2014) Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. Genome Res., 24, 592-603.

11. Mouchiroud, L., Eichner, L.J., Shaw, R. and Auwerx, J. (2014) Transcriptional coregulators: fine-tuning metabolism. Cell Metab., 20, 26-40.

12. Himeda, C.L., Ranish, J.A., Angelico, J.C., Maira, P., Aebersold, R. and Haukscha, S.D. (2004) Quantitative proteomic identification of s9x as the trec-binding factor in the muscle creatine kinase enhancer. Mol. Cell. Biol., 24, 2123-2134.

13. Butter, F., Davison, L., Viturawong, T., Scheibl, M., Vermeulen, M., Todd, J.A., Mann, M. and Barsh, G.S. (2012) Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet., 8, e1002982.

14. Merl, J., Uefting, M., Hauck, S.M. and von Toerne, C. (2012) Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal Müller cells. Proteomics, 12, 1902–1911.

15. Park, S.-S., Wu, W.W., Zhou, Y., Shen, R.-F., Martin, B. and Moudsley, S. (2012) Effective correction of experimental errors in quantitative proteomics using stable isotope labeling by amino acids in cell culture (SILAC). J. Proteomics, 75, 3720-3732.

16. Dee, S.S., Fajas, L., Nemoto, M., Pihlajamaki, J., Mykkanen, L., Kuusisto, J., Laakso, M., Fujimoto, W. and Auwerx, J. (1998) A Pro12Ala substitution in PPARγ2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. J. Proteomics, 20, 284-287.

17. Altshuler, D., Hirschhorn, J.N., Klannemark, M., Lindgren, C.M., Vohl, M.C., Nemesh, J., Lane, C.R., Schaffner, S.F., Bolk, S., Brewer, C. et al. (2000) The common PPARγ[alpha] Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nature, 26, 76–80.

18. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet., 42, 579–589.

19. Rosen, E.D., Sarraf, P., Troy, A.E., Bradwin, G., Moore, K., Milstone, D.S., Spiegelman, B.M. and Mortensen, R.M. (1999) PPARγ is required for the differentiation of adipose tissue in vivo and in vitro. Mol. Cell, 4, 611–617.

20. Medina-Gomez, G., Virtue, S., Lelliott, C., Boiani, R., Campbell, M., Christodoulides, C., Perrin, C., Jimenez-Linan, M., Blount, M., Dixon, J. et al. (2005) The link between nutritional status and insulin sensitivity is dependent on the adipocyte-specific peroxisome proliferator—activated receptor-γ2 isoform. Diabetes, 54, 1706–1716.

21. Tontonoz, P., Hu, E. and Spiegelman, B.M. (1994) Stimulation of adipogenesis in fibroblasts by PPARα, a lipid-activated transcription factor. Cell, 79, 1147–1156.

22. Fajas, L., Auboeuf, D., Rasp, E., Schoonjans, K., Lefebvre, A.-M., Saladin, R., Najib, J., Laville, M., Frucht, J.-C., Deeb, E. et al. (1997) The organization, promoter analysis, and expression of the human PPARγ gene. J. Biol. Chem., 272, 18779-18789.

23. Mukherjee, R., Jow, L., Clayton, G.E. and Paterniti, J.R. (1997) Identification, characterization, and tissue distribution of human peroxisome proliferator-activated receptor (PPAR) isoforms PPARγ2 versus PPARγ1 and activation with retinoid X receptor agonists and antagonists. J. Biol. Chem., 272, 8071–8076.

24. Saladin, R., Fajas, L., Dana, S., Halvorsen, Y.D., Auwerx, J. and Briggs, M. (1999) Differential regulation of peroxisome proliferator activated receptor gamma 1 (PPARgamma1) and PPARgamma2 messenger RNA expression in the early stages of adipogenesis. Cell Growth Differ., 10, 43–48.

25. Hoffmann, C., Zimmermann, A., Hinney, A., Volckmar, A.L., Jarrett, H.W., Fromme, T., Klingenspor, M. and Ko, B.C. (2013) A novel SNP/SP3 dependent intrinsic enhancer governing transcription of the UCP3 gene in brown adipocytes. PLoS ONE, 8, e83426.

26. Schreiber, E., Matthias, P., Müller, M.M. and Schaffner, W. (1989) Rapid detection of octamer binding proteins with ‘mini extracts’, prepared from a small number of cells, Nucleic Acids Res., 17, 6419.
27. Kretschmer,A., Müller,G., Lee,H., Laumen,H., Toerner,C., von Schramm,K., Prokisch,H., Eyerich,S., Wahl,S., Baurecht,H. et al. (2014) A common atopy-associated variant in the TH2 cytokine locus control region impacts transcriptional regulation and alters SMAD3 and SP1 binding. *Allergy*, 69, 632–642.

28. Wisniewski,J.R., Zougman,A., Nagaraj,N. and Mann,M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods*, 6, 359–362.

29. Hauck,S.M., Dietter,J., Kramer,R.L., Hofmaier,F., Zipplies,J.K., Hauck,J.M., Seni,E., Zhang,Y., Takahashi,H. et al. (2013) Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. *Mol. Cell. Proteomics*, 9, 2292–2305.

30. von Toerne,C., Kahle,M., Schäfer,A., Ispiryán,R., Blindert,M., Hrade De Angelis,M., Neschen,S., Ueffing,M. and Hauck,S.M. (2013) Aopic, Mbl2, and Psp plasma protein levels correlate with diabetic phenotype in NZO mice—an optimized rapid workflow for SRM-based quantification. *J. Proteome Res.*, 12, 1331–1343.

31. Mueller,K., Quadrat,J., Marienfeld,R.B., Wehrich,P., Fiedler,K., Claussnitzer,M., Laumen,H., Varth,M., Berberich-Siebelt,F., Serfling,E. et al. (2013) Octamer-dependent transcription in T cells is mediated by NFAT and NF-kB. *Nucleic Acids Res.*, 41, 2138–2154.

32. Feldkotter,M., Schwarzer,V., Wirth,T., Wirtz,B. (2002) Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. *Am. J. Hum. Genet.*, 70, 358–368.

33. Holzapfel,C., Baumert,J., Grallert,H., Müller,A.M., Thorand,B., Khuseyinova,N., Herder,C., Meisinger,C., Hauner,H., Wichmann,H.E. et al. (2008) Genetic variants in the USF1 gene are associated with low-density lipoprotein cholesterol levels and incident type 2 diabetes mellitus in women: results from the MONICA/KORA Augsburg case–cohort study, 1984–2002. *Eur. J. Endocrinol.*, 159, 407–416.

34. Arner,E., Mejhert,N., Kyte,E., Alberiche,M., Balwierz,P.J., Pachkov,M., Targher,G., Alberiche,M., Bonadonna,R.C., Saggiani,F., Hauck,J.M., Seni,E., Zhang,Y., Takahashi,H. et al. (2013) Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. *Mol. Cell. Proteomics*, 9, 2292–2305.

35. 1000 Genomes Project Consortium. (2012) An integrated map of regulatory variation in human genome. *Nature*, 498, 56–65.

36. 1000 Genomes Project Consortium. (2017) A common atopy-associated variant in the Th2 cytokine locus control region impacts transcriptional regulation and alters SMAD3 and SP1 binding. *Allergy*, 69, 632–642.

37. Pennacchio,L.A., Ahtiyu,N., Moses,A.M., Prabhakar,S., Nobrega,M.A., Shourky,M., Minovitsky,S., Dubchak,I., Holt,A., Lewis,K.D. et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, 444, 499–502.

38. Zhou,L., Wang,L., Lu,L., Jiang,P., Sun,H. and Wang,H. (2012) A novel target of microRNA-29, Ring1 and YY1-binding protein (RYPb), negatively regulates skeletal myogenesis. *J. Biol. Chem.*, 287, 25255–25265.

39. Garcia,E., Marcos-Gutierrez,C., del Mar Lorente,M., Moreno,J.C., Kulyte,A., Balwierz,P.J., Pachkov,M., Targher,G., Alberiche,M., Bonadonna,R.C., Saggiani,F., Hauck,J.M., Seni,E., Zhang,Y., Takahashi,H. et al. (2013) Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. *Mol. Cell. Proteomics*, 9, 2292–2305.

40. Schlisio,S., Halperin,T., Vidal,M. and Nevins,J.R. (2002) Interaction protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res.*, 12, 284–293.

41. Pennacchio,L.A., Ahtiyu,N., Moses,A.M., Prabhakar,S., Nobrega,M.A., Shourky,M., Minovitsky,S., Dubchak,I., Holt,A., Lewis,K.D. et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, 444, 499–502.

42. Visel,A., Taeh,L., Girgis,I., May,D., Golonzhka,O., Hoch,R.V., McKinsey,G.L., Pattabiraman,K., Silberberg,S.N., Blow,M.J. et al. (2013) A high-resolution enhancer atlas of the developing telencephalon. *Cell*, 152, 895–908.

43. Simoncic,J., Schmid,A.W., Gilardoni,P.A., Zoller,B., RagHAV,S.K., Krier,J., Gubelmann,C., Lisac,F., Nael,F., Moniati,M. et al. (2013) Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. *Nat. Methods*, 10, 570–576.

44. Heikkinen,S., Argmann,C., Feige,J.N., Koutnikova,H., Champy,M.-F., Dalil-Yousef,C., Schadt,E.E., Saad,Laakso,M. and Auwerx,J. (2009) The Pro12Ala PPARδ2 variant determines metabolism at the gene-environment interface. *Cell Metab.*, 9, 88–98.

45. Dupuis,J., Langenberg,C., Prokopenko,I., Saxena,R., Soranzo,N., Jackson,A.U., Wheeler,E., Glazer,N.L., Boutaia-Naji,N., Glyn,A.L. et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.*, 42, 105–116.

46. He,C.Q., Ding,N.Z. and Fan,W. (2008) YY1 repressing peroxisome proliferator-activated receptor delta promoter. *Mol. Cell. Biochem.*, 308, 247–252.

47. Blättler,S.M., Cunningham,j.T., Verduguer,F., Chim,H., Hauser,C., Meisinger,C., Hauner,H., Wichmann,H.E. et al. (2008) Genetic variants in the USF1 gene are associated with low-density lipoprotein cholesterol levels and incident type 2 diabetes mellitus in women: results from the MONICA/KORA Augsburg case–cohort study, 1984–2002. *Eur. J. Endocrinol.*, 159, 407–416.

48. Arner,E., Mejhert,N., Kyte,E., Alberiche,M., Balwierz,P.J., Pachkov,M., Corman,M., Lorente-Cebrian,S., Ehrlund,A., Laurencikiene,J., Attar-Cohen,H., Ingle,C., Beazley,C., Gutierrez,A.M., Sekowska,M., Champy,M.-F., Dalil-Yousef,C., Schadt,E.E., Saad,Laakso,M. and Auwerx,J. (2009) The Pro12Ala PPARδ2 variant determines metabolism at the gene-environment interface. *Cell Metab.*, 9, 88–98.

49. Lu,Y., Xiong,X., Wang,X., Zhang,Z., Li,J., Shi,G., Yang,J., Zhang,H., Ning,G. and Li,X. (2013) Yin Yang1 promotes hepatic gluconeogenesis through upregulation of glucoorticoid receptor. *Diabetes*, 62, 1064–1073.

50. Verduguer,F., Blättler,S.M., Cunningham,j.T., Hall,J.A., Chim,H. and Puigserver,P. (2014) Decreased genetic dosage of hepatic Yin Yang1 causes diabetic-like symptoms. *Am. J. Physiol. Endocrinol. Metab.*, 208, 308–316.

51. Han,Y., Choi,Y.H., Lee,S.H., Jin,Y.H., Cheong,H. and Lee,K.Y. (2015) Yin Yang1 is a multi-functional regulator of adipocyte differentiation in 3T3-L1 cells. *Mol. Cell. Endocrinol.*, 413, 217–227.

52. Xiong,T.C., Davis,R.J. and Chim,H.C. (2002) Phosphorylation of NFT4 by p38 mitogen-activated protein kinases. *Mol. Cell. Biol.*, 20, 3892–3904.

53. Zheng,L., Schickling,O., Peter,M.E. and Lenardo,M.J. (2001) The death effector domain-associated factor plays distinct regulatory roles in the nucleus and cytoplasm. *J. Biol. Chem.*, 276, 31945–31952.

54. Danen-van,O.A.A., Voskamp,P., Seelen,M.C., van Miltenburg,M.H., Bolk,M.W., Tait,S.W., Roosen-de,C.J.G., Rohn,J.L. and Noteborn,M.H. (2004) Human death effector domain-associated factor interacts with the viral apoptosis agonist apopxin and exerts tumor-preferential cell killing. *Cell Death Differ.*, 11, 564–573.

55. Prity,M.K., Locker,J. and Schreiβer-Agus,N. (2005) Rybp/DEDAF is required for postimplantation and for central nervous system development. *Mol. Cell. Biol.*, 25, 7193–7202.

56. Huang,H.Y., Li,X., Liu,M., Song,T.J., He,Q., Ma,C.G. and Tang,Q.Q. (2008) Transcription factor YY1 promotes adipogenesis via inhibiting CHOP-10 expression, *Biochem. Biophys. Res. Commun.*, 375, 496–500.

57. Corradino,O., Saiakova,A., Akhtar-Zaidi,B., Myeroff,L., Willis,J., Cowper-Sal lari,R., Lupien,M., Markowitz,S. and Scachetti,P. (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.*, 24, 1–13.