archive ouverte UNIGE
http://archive-ouverte.unige.ch

Article

Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes

SMALL, Kerrin S, *et al*.

Abstract

Genome-wide association studies have identified many genetic variants associated with complex traits. However, at only a minority of loci have the molecular mechanisms mediating these associations been characterized. In parallel, whereas cis regulatory patterns of gene expression have been extensively explored, the identification of trans regulatory effects in humans has attracted less attention. Here we show that the type 2 diabetes and high-density lipoprotein cholesterol-associated cis-acting expression quantitative trait locus (eQTL) of the maternally expressed transcription factor KLF14 acts as a master trans regulator of adipose gene expression. Expression levels of genes regulated by this trans-eQTL are highly correlated with concurrently measured metabolic traits, and a subset of the trans-regulated genes harbor variants directly associated with metabolic phenotypes. This trans-eQTL network provides a mechanistic understanding of the effect of the KLF14 locus on metabolic disease risk and offers a potential model for other complex traits.

Reference

SMALL, Kerrin S, *et al*. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nature Genetics*, 2011, vol. 43, no. 6, p. 561-4

DOI : 10.1038/ng.833
PMID : 21572415

Available at:
http://archive-ouverte.unige.ch/unige:32204

Disclaimer: layout of this document may differ from the published version.

UNIVERSITÉ DE GENÈVE

nature genetics

Identification of an imprinted master *trans* regulator at the *KLF14* locus related to multiple metabolic phenotypes

Kerrin S Small[1,2,10], Åsa K Hedman[3,10], Elin Grundberg[1,2,10], Alexandra C Nica[4], Gudmar Thorleifsson[5], Augustine Kong[5], Unnur Thorsteindottir[5,6], So-Youn Shin[2], Hannah B Richards[7], the GIANT Consortium[8], the MAGIC Investigators[8], the DIAGRAM Consortium[8], Nicole Soranzo[1,2], Kourosh R Ahmadi[1], Cecilia M Lindgren[3], Kari Stefansson[5,6,10], Emmanouil T Dermitzakis[4,10], Panos Deloukas[2,10], Timothy D Spector[1,10] & Mark I McCarthy[3,7,9,10] for the MuTHER Consortium[8]

© 2011 Nature America, Inc. All rights reserved.

Genome-wide association studies have identified many genetic variants associated with complex traits. However, at only a minority of loci have the molecular mechanisms mediating these associations been characterized. In parallel, whereas *cis* regulatory patterns of gene expression have been extensively explored, the identification of *trans* regulatory effects in humans has attracted less attention. Here we show that the type 2 diabetes and high-density lipoprotein cholesterol–associated *cis*-acting expression quantitative trait locus (eQTL) of the maternally expressed transcription factor *KLF14* acts as a master *trans* regulator of adipose gene expression. Expression levels of genes regulated by this *trans*-eQTL are highly correlated with concurrently measured metabolic traits, and a subset of the *trans*-regulated genes harbor variants directly associated with metabolic phenotypes. This *trans*-eQTL network provides a mechanistic understanding of the effect of the *KLF14* locus on metabolic disease risk and offers a potential model for other complex traits.

Variants near the maternally expressed transcription factor *KLF14* (encoding Kruppel-like factor 14) are robustly associated with both type 2 diabetes (T2D) and high-density lipoprotein (HDL) cholesterol levels in large-scale genome-wide association studies (GWAS)[1,2]. These studies have implicated a group of highly correlated SNPs, including rs4731702 and rs972283 located ~14 kb upstream of *KLF14*. *KLF14* is the gene in the region likely to be mediating these effects, as the same SNPs show adipose-specific, maternally restricted *cis*-regulatory associations with *KLF14* expression levels, a pattern that mirrors the parent-of-origin effects for T2D susceptibility at this locus[3].

Because transcription factors such as *KLF14* typically modulate expression of other genes in *trans*, we tested for association between rs4731702 and expression levels of ~24,000 probes (16,663 genes) on the Illumina Human HT12 array in subcutaneous adipose tissue biopsies from a cohort of 776 healthy female twins[4]. The enrichment of rs4731702 *trans* associations for low *P* values (**Fig. 1** and **Supplementary Fig. 1**) suggests that *KLF14* is a master regulator of gene expression in adipose tissue. The pattern of *trans* associations at *KLF14* mirrors the GWAS associations (**Fig. 2** and **Supplementary**

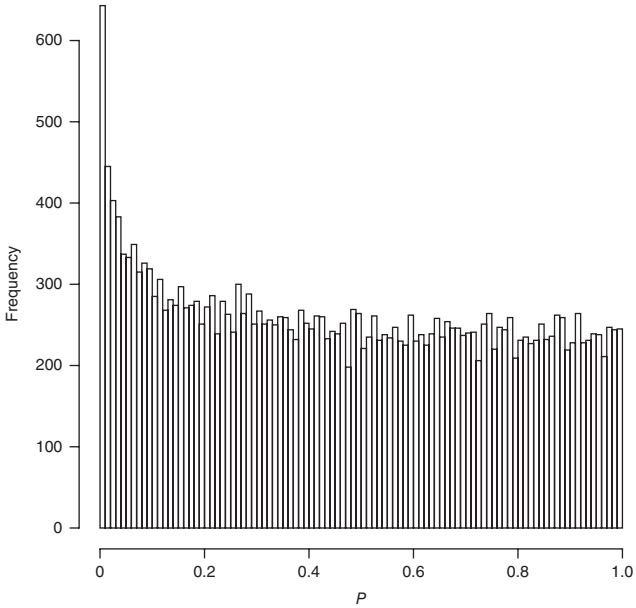

Figure 1 *KLF14* is a master regulator of gene expression in adipose tissue. Shown is the *P* value distribution of association between the *KLF14 cis*-eQTL rs4731702 and expression levels of ~24,000 probes in adipose tissue.

[1]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [2]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [4]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. [5]deCODE genetics, Reykjavik, Iceland. [6]Faculty of Medicine, University of Iceland, Reykjavík, Iceland. [7]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK. [8]A full list of members is provided in the **Supplementary Note**. [9]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK. [10]These authors contributed equally to this work. Correspondence should be addressed to T.D.S. (tim.spector@kcl.ac.uk) or M.I.M. (mark.mccarthy@drl.ox.ac.uk).

Received 17 December 2010; accepted 14 April 2011; published online 15 May 2011; corrected after print 22 September 2011; doi:10.1038/ng.833

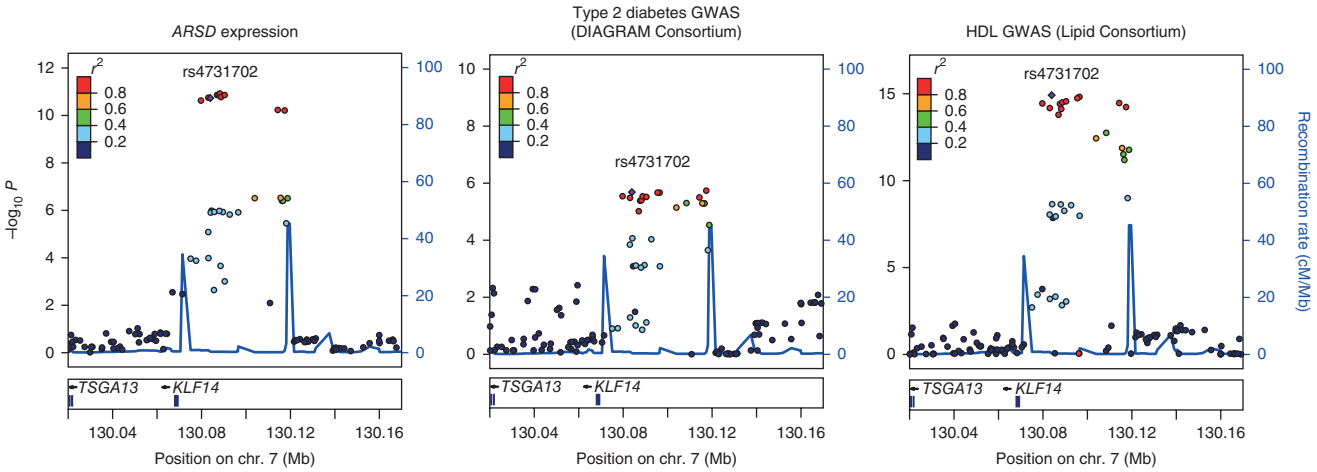

Figure 2 Regional signal plots at the *KLF14* locus. The left panel shows a signal plot of a representative *trans*-regulated gene in the MuTHER samples ($n = 776$). The middle panel shows a signal plot for the type 2 diabetes GWAS meta-analysis performed by the DIAGRAM consortium[1] (stage 1 data only, effective sample size of 22,044; rs972283 reached genome-wide significance after further replication). The right panel shows a signal plot for the HDL cholesterol GWAS meta-analysis performed by the Lipid Consortium[2] ($n = 99{,}900$). Signal plots of all ten genome-wide significant *trans* regulated genes are included in **Supplementary Figure 2**.

© 2011 Nature America, Inc. All rights reserved.

Fig. 2), and conditioning the *trans* associations on rs4731702 abolished the signal at all other SNPs. These findings indicate that the same set of SNPs (and presumably the same causal variant) underlies the *cis*, *trans* and metabolic trait associations at this locus.

We focused on the ten genes (*TPMT*, *ARSD*, *SLC7A10*, *C8orf82*, *APH1B*, *PRMT2*, *NINJ2*, *KLF13*, *GNB1* and *MYL5*) showing genome-wide significant *trans* (GWST) associations ($P < 5 \times 10^{-8}$) driven by rs4731702 (**Table 1**). First, we sought replication of the *trans* associations in an independent set of adipose tissue samples (deCODE genetics; $n = 589$)[5]. As previously reported[3], the deCODE data showed a strong maternally specific *cis* association between rs4731702 and *KLF14* expression in adipose tissue ($P = 1 \times 10^{-19}$). We did not detect this *cis* effect in the MuTHER data because of apparent problems with the *KLF14* probe represented on the Illumina HT12 array used for the MuTHER data (Online Methods). Seven of the GWST genes from the MuTHER analysis had a directionally consistent *trans* association with $P < 0.05$ in the deCODE replication set (**Table 1**), and we were able to show parent-of-origin effects for the *trans* associations consistent with the maternally specific *cis* effects for *KLF14* expression and T2D risk[3]. In the deCODE replication data, maternally inherited *trans* associations were markedly more significant than general analyses, and we saw no paternally inherited *trans* associations (**Table 1**).

The *trans* effects explain a substantial portion of the genetically regulated variation in GWST expression levels. Our heritability estimates of GWST gene expression levels ranged from 0.13 to 0.79; the rs4731702 *trans*-eQTL explained between 3% and 7.8% of the variance in expression, corresponding to 6% to 25% of the heritability (**Table 2**). Expression levels of the ten GWST genes were moderately correlated in adipose tissue, with a mean |pairwise ρ| of 0.29 (standard deviation = 0.15). *SLC7A10* is the only GWST gene downregulated by the T2D risk allele (and hence the only transcript showing anti-correlated expression levels within the GWST genes). This pattern is consistent with the known ability of the KLF family of transcription factors to act as both transcription activators and repressors[6].

We obtained further support for the hypothesis that the *trans* effects are mediated by *KLF14* expression from analysis of transcription factor binding sites in *trans*-associated genes using PSCAN[7] with the JASPAR database[8]. KLF14 itself is not represented in JASPAR, but other KLF family members have closely related binding sites (and in some cases have been shown to compete for the same binding site)[9], and KLF4 (the only KLF family member in the JASPAR database) and KLF14 share highly similar DNA-binding C-terminal regions[10]. Though we found no evidence for enrichment after correction for multiple testing when examining the ten GWST genes alone, inclusion

Table 1 Genome-wide significant ($P < 5 \times 10^{-8}$) associations of gene expression levels with rs4731702

		MuTHER		deCODE all		deCODE maternal		deCODE paternal		Combined MuTHER + deCODE maternal		
Gene	Chr.	Effect (s.e.)	*P*	Effect (s.e.)	*P*	Effect (s.e.)	*P*	Effect (s.e.)	*P*	Z score	*P*	Direction
APH1B	15	0.08 (0.013)	1.2×10^{-9}	0.11 (0.059)	0.08	0.17 (0.085)	0.07	0.07 (0.083)	0.44	6.1	9.7×10^{-10}	+ +
ARSD	X	0.08 (0.012)	1.9×10^{-11}	0.24 (0.059)	2.2×10^{-4}	0.51 (0.083)	2.6×10^{-8}	−0.004 (0.083)	0.96	8.6	5.4×10^{-18}	+ +
C8orf82	8	0.09 (0.014)	4.8×10^{-10}	0.28 (0.058)	8.9×10^{-6}	0.69 (0.080)	2.1×10^{-14}	−0.09 (0.082)	0.28	9.3	1.1×10^{-20}	+ +
GNB1	1	0.05 (0.009)	4.0×10^{-8}	0.23 (0.059)	1.8×10^{-4}	0.42 (0.085)	1.6×10^{-6}	0.06 (0.084)	0.51	7.2	6.1×10^{-13}	+ +
KLF13	15	0.10 (0.017)	2.2×10^{-8}	−0.01 (0.060)	0.94	0.01 (0.086)	0.89	−0.02 (0.084)	0.80	4.8	1.4×10^{-6}	+ +
MYL5	4	0.09 (0.017)	4.5×10^{-8}	0.20 (0.059)	1.3×10^{-3}	0.45 (0.083)	1.3×10^{-7}	−0.04 (0.083)	0.60	7.4	1.1×10^{-13}	+ +
NINJ2	12	0.08 (0.013)	8.4×10^{-9}	0.14 (0.060)	0.03	0.24 (0.087)	0.01	0.05 (0.085)	0.59	6.3	4.1×10^{-10}	+ +
PRMT2	21	0.06 (0.010)	6.9×10^{-9}	0.18 (0.060)	0.01	0.27 (0.087)	6.7×10^{-3}	0.09 (0.085)	0.33	6.4	2.1×10^{-10}	+ +
SLC7A10	19	−0.27 (0.042)	2.7×10^{-10}	−0.21 (0.057)	7.4×10^{-4}	−0.31 (0.082)	3.3×10^{-4}	−0.11 (0.081)	0.18	−7.3	3.8×10^{-13}	− −
TPMT	6	0.10 (0.013)	1.6×10^{-14}	−0.04 (0.060)	0.49	−0.03 (0.087)	0.78	−0.06 (0.084)	0.49	6.4	1.8×10^{-10}	+ −

The effect allele is the type 2 diabetes risk allele C, which has a frequency of 55% in the HapMap CEU population. Chr., chromosome; s.e., standard error.

Table 2 Heritability and *trans*-eQTL variance of GWST gene expression

Probe ID	Gene	Chr.	Transcription start site (build 36)	Percent variance explained in *trans*	h^2	Percent h^2 explained in *trans*
ILMN_1767816	*APH1B*	15	61,356,801	3.1	0.13	23.2
ILMN_1684873	*ARSD*	X	2,832,010	6.6	0.52	12.6
ILMN_1693862	*C8orf82*	8	145,722,410	6.9	0.27	25.4
ILMN_1760320	*GNB1*	1	1,706,588	4.1	0.19	21.3
ILMN_1679929	*KLF13*	15	29,406,374	4.3	0.70	6.2
ILMN_1746948	*MYL5*	4	661,710	4.0	0.40	10.0
ILMN_1731745	*NINJ2*	12	543,722	4.7	0.48	9.8
ILMN_1675038	*PRMT2*	21	46,879,954	4.9	0.55	8.9
ILMN_1681087	*SLC7A10*	19	38,391,409	5.4	0.79	6.8
ILMN_1740185	*TPMT*	6	18,236,523	7.8	0.48	16.4

Chr., chromosome.

of a larger number of *trans*-associated genes (46 with *trans* $P < 10^{-4}$ or 121 with *trans* $P < 10^{-3}$) showed strong evidence of enrichment for KLF4 binding sites. KLF4 was the most over-represented binding site in the former set (Bonferroni corrected $P = 0.01$) and the second most over-represented site in the latter set (Bonferroni corrected $P = 1.3 \times 10^{-7}$) after EGR1. These data indicate that one feature of the transcripts showing *trans* associations with the *KLF14* SNPs is enrichment for KLF binding sites.

Having shown that the same set of SNPs influences *cis* expression of *KLF14, trans* expression of members of the GWST gene network and a variety of metabolic traits including T2D and HDL cholesterol, we sought to clarify the causal connections between these effects and, in particular, to establish whether or not the *trans* effects were likely to be mediating the metabolic associations at *KLF14*. First, we examined the correlations between *trans* gene expression and concurrently measured metabolic phenotypes. At an array-wide Bonferroni threshold of $P < 1.9 \times 10^{-6}$, expression levels of six of the ten GWST genes were associated with body mass index (BMI) and HDL cholesterol, five with triglycerides, fasting insulin levels and HOMA-IR (an index of insulin sensitivity), three with adiponectin levels and two with fasting glucose (**Table 3**). Compared to all genes on the array, this represents an enrichment for expression and metabolic phenotype associations, with significance ranging from $P = 0.001$ to $P = 3.3 \times 10^{-5}$. The strength of these associations is consistent with a causal link between *trans* gene expression and metabolic phenotypes and provides clues to the biological processes in which these genes may participate.

Next, we examined large-scale association data made available by trait-specific GWAS meta-analysis consortia, focusing on SNPs in the 250-kb region surrounding each GWST gene. The rs4731702 T2D risk allele is associated with higher fasting insulin[1], indicating that the primary effect on diabetes risk is mediated by decreased peripheral insulin sensitivity. Accordingly, we focused on a set of insulin-resistance–related traits, including fasting insulin[11], fasting glucose[11], HOMA-IR[11], T2D[1], lipids (HDL, low-density lipoprotein (LDL) and triglycerides)[2], body fat distribution (BMI-adjusted waist-hip ratio (WHR))[12] and BMI[13]. In GWAS datasets ranging in size from 22,044 to 123,865 individuals, we found eight associations in five genes at a study-wide significance threshold of $P < 1.03 \times 10^{-4}$ (**Table 4**; see Online Methods for the threshold determination). For example, SNPs near *APH1B* are associated with HDL (rs2729787; $P = 9.8 \times 10^{-9}$) and triglycerides (rs17184382; $P = 1.5 \times 10^{-5}$), and SNPs near *KLF13* are associated with BMI-adjusted WHR (rs4779526; $P = 1.8 \times 10^{-5}$) and LDL (rs8034505; $P = 5.8 \times 10^{-5}$). In addition, SNPs in *MSRA* (expression levels of which marginally failed to reach genome-wide significance; *trans* association $P = 9.8 \times 10^{-8}$) have been previously associated with waist circumference[10] and are here associated with triglycerides (rs615171; $P = 7.5 \times 10^{-7}$). This pattern of association signals shows that variation involving GWST genes has the potential to affect insulin-resistance–related traits and thereby supports the notion that a subset of these genes are directly implicated in mediating the effects of *KLF14* variation on disease susceptibility.

Table 3 Association between expression of GWST genes and concurrently measured metabolic phenotypes

Gene	Value	Adiponectin	HDL	LDL	Trigylcerides	BMI	Fasting insulin	Fasting glucose	HOMA–IR
APH1B	*P* value	1.1×10^{-8}	1.9×10^{-8}	0.01	5.2×10^{-17}	5.2×10^{-15}	6.7×10^{-18}	0.001	3.6×10^{-17}
	Beta	–0.02	–0.11	0.02	0.13	0.01	0.002	0.05	0.05
ARSD	*P* value	0.002	6.8×10^{-9}	0.02	3.6×10^{-11}	5.3×10^{-11}	4.3×10^{-16}	4.2×10^{-10}	1.3×10^{-16}
	Beta	–0.008	–0.10	0.02	0.09	0.01	0.002	0.07	0.04
C8orf82	*P* value	3.1×10^{-8}	4.4×10^{-8}	0.008	4.1×10^{-5}	2.4×10^{-14}	6.0×10^{-5}	0.52	3.0×10^{-4}
	Beta	0.02	0.11	–0.03	–0.07	–0.01	–0.001	–0.009	–0.02
GNB1	*P* value	9.1×10^{-5}	6.1×10^{-10}	0.03	2.1×10^{-12}	2.9×10^{-21}	8.4×10^{-8}	5.3×10^{-4}	1.3×10^{-7}
	Beta	–0.008	–0.08	0.01	0.08	0.01	0.001	0.03	0.02
KLF13	*P* value	0.23	0.11	0.86	0.33	0.93	0.87	0.02	0.98
	Beta	0.005	–0.04	–0.002	0.02	0	0	0.04	–0.0002
MYL5	*P* value	3.6×10^{-5}	5.3×10^{-4}	0.19	0.8	8.4×10^{-5}	0.007	0.47	0.02
	Beta	0.02	0.10	–0.02	–0.005	–0.009	–0.001	–0.01	–0.02
NINJ2	*P* value	0.49	0.03	0.03	1.2×10^{-4}	0.003	6.0×10^{-4}	0.51	0.002
	Beta	–0.002	–0.04	0.02	0.06	0.005	0.001	0.009	0.02
PRMT2	*P* value	0.04	0.06	0.56	0.21	1.1×10^{-5}	0.008	0.3	0.006
	Beta	0.005	0.03	–0.004	–0.015	–0.006	–0.0004	–0.01	–0.01
SLC7A10	*P* value	1.7×10^{-14}	7.8×10^{-30}	2.7×10^{-5}	1.7×10^{-34}	3.6×10^{-48}	1.2×10^{-51}	6.7×10^{-7}	7.1×10^{-48}
	Beta	0.07	0.65	–0.12	–0.55	–0.08	–0.01	–0.21	–0.27
TPMT	*P* value	3.4×10^{-6}	2.3×10^{-9}	0.01	4.9×10^{-16}	7.5×10^{-24}	7.3×10^{-21}	5.7×10^{-4}	1.2×10^{-19}
	Beta	–0.01	–0.11	0.023	0.12	0.02	0.002	0.05	0.06

HDL, high density lipoprotein; LDL, low-density lipoprotein; BMI, body mass index.

© 2011 Nature America, Inc. All rights reserved.

Table 4 Genome-wide association meta-analysis signals within 250 kb of GWST genes

Gene	Trait	SNP	Effect allele	Z score	*P*
APH1B	HDL	rs2729787	T	5.73	9.81×10^{-9}
APH1B	Triglycerides	rs17184382	A	4.32	1.58×10^{-5}
C8orf82	Type 2 diabetes	rs2294120	A	1.14 (OR)	8.43×10^{-5}
NINJ2	LDL	rs2302408	T	3.91	9.06×10^{-5}
SLC7A10	HDL	rs8182584	T	–5.11	3.19×10^{-7}
SLC7A10	Waist-hip ratio	rs7251505	A	–4.85	3.21×10^{-6}
KLF13	LDL	rs8034505	A	4.02	5.85×10^{-5}
KLF13	Waist-hip ratio	rs4779526	A	4.50	1.79×10^{-5}
MSRA	Triglycerides	rs615171	T	–4.95	7.50×10^{-7}
MSRA	Waist circumference	rs7826222	G	5.75	8.89×10^{-9}

This table also includes the results for *MSRA*, for which the *trans* association marginally failed to reach genome-wide significance ($P = 9.8 \times 10^{-8}$). OR, odds ratio.

One of the more notable transcripts identified by these analyses is *SLC7A10*, a member of the solute carrier family that mediates transport of neutral amino acids. Adipose expression of *SLC7A10* is highly heritable ($h^2 = 0.79$) and is downregulated by the *KLF14* T2D risk allele. *SLC7A10* expression is strongly associated with diverse metabolic phenotypes, is negatively correlated with BMI ($P = 3 \times 10^{-48}$), insulin ($P = 1.1 \times 10^{-51}$), HOMA-IR ($P = 7 \times 10^{-48}$), glucose ($P = 6 \times 10^{-7}$) and triglycerides ($P = 1 \times 10^{-34}$) and is positively correlated with HDL ($P = 7 \times 10^{-30}$) and adiponectin ($P = 1 \times 10^{-12}$) levels (**Table 3**). The *SLC7A10* locus contains independent ($r^2 = 0.03$) SNPs associated with HDL (rs8182584; $P = 3.2 \times 10^{-7}$) and BMI-adjusted WHR (rs7251505; $P = 3.2 \times 10^{-6}$). The former SNP (rs8182584) is weakly associated with insulin levels ($P = 0.002$) and BMI ($P = 1.4 \times 10^{-3}$), suggesting that this gene has a wide-ranging role in metabolism.

Our data provide convincing evidence of a *bona fide* adipose *trans*-eQTL and implicate this *trans* expression network in the link between *KLF14* variation and risk of metabolic disease. This *trans* regulation uncovers new biological links between previously identified genome-wide significant associations at *KLF14* (with HDL and T2D), *APH1B* (with HDL) and *MSRA* (with waist circumference) and additional signals where metabolic trait associations have not yet been established to genome-wide significance (*SLC7A10*, *KLF13*, *C8orf82* and *NINJ2*). These links provide a framework for hypothesis-directed investigation of genetic interactions among GWAS loci and provide an example of the power of 'integrative genomics' to leverage 'omics' data from multiple sources to discover new biological and functional insights.

URLs. R, http://www.r-project.org/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession numbers. Expression data has been deposited in Array Express under accession number E-TABM-1140.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The MuTHER study was funded by the Wellcome Trust Program grant # 081917. Genotyping of TwinsUK samples was provided by the Wellcome Trust Sanger Institute and the National Eye Institute via a US National Institutes of Health (NIH)/Center for Inherited Disease Research (CIDR) genotyping project. TwinsUK also receives support from the ENGAGE project grant agreement HEALTH-F4-2007-201413 and from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St. Thomas' National Health Service Foundation Trust in partnership with King's College London. T.D.S. is an NIHR senior investigator and European Research Council (ERC) senior investigator. M.I.M. is supported by the Oxford NIHR Biomedical Research Centre. Additional support was provided by the Louis-Jeantet Foundation to E.T.D. and A.C.N. and via NIH-NIMH grant R01 MH090941 to E.T.D. and M.I.M.

AUTHOR CONTRIBUTIONS

K.S.S., Å.K.H., E.G., G.T. and A.C.N. analyzed data. G.T., A.K., S.-Y.S., H.B.R., N.S. and C.M.L. contributed reagents, materials and analysis tools. U.T., K.R.A., K.S., E.T.D., P.D., M.I.M. and T.D.S. conceived and designed the experiments. K.S.S. and M.I.M. wrote the paper with contributions from Å.K.H. and E.G. All authors read and approved the manuscript before submission.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
2. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
3. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
4. Nica, A.C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
5. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
6. Dang, D.T., Pevsner, J. & Yang, V.W. The biology of the mammalian Kruppel-like family of transcription factors. *Int. J. Biochem. Cell Biol.* **32**, 1103–1121 (2000).
7. Zambelli, F., Pesole, G. & Pavesi, G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.* **37**, W247–W252 (2009).
8. Portales-Casamar, E. *et al.* JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **38**, D105–D110 (2010).
9. Kaczynski, J., Cook, T. & Urrutia, R. Sp1- and Kruppel-like transcription factors. *Genome Biol.* **4**, 206 (2003).
10. McConnell, B.B. & Yang, V.W. Mammalian Kruppel-like factors in health and diseases. *Physiol. Rev.* **90**, 1337–1381 (2010).
11. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
12. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
13. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).

© 2011 Nature America, Inc. All rights reserved.

ONLINE METHODS

Sample collection. The MuTHER (Multiple Tissue Human Expression Resource) study includes 856 female individuals of European descent (336 monozygotic and 520 dizygotic twins) recruited from the TwinsUK Adult twin registry[14] (776 MuTHER participants had both adipose tissue expression profiles and genome-wide genotypes available). The age at inclusion ranged from 40 to 87 years, with a median age of 62 years. Punch biopsies (8 mm) were taken from a relatively photoprotected area adjacent and inferior to the umbilicus. Subcutaneous fat tissue was carefully dissected from each biopsy using forceps and a scalpel, weighed and immediately stored in liquid nitrogen. The project was approved by the local ethics committees of all institutions involved, and all samples were collected after obtaining written and signed informed consent. Metabolic phenotypes were measured at the same time point as the biopsies and were collected as previously described, including anthropometric traits (height, weight and BMI)[15], blood lipids[16], adiponectin[17], fasting glucose[18] and fasting insulin[19]. The deCODE replication samples were collected as previously described[5].

RNA extraction. RNA was extracted using TRIzol Reagent (Invitrogen) according to protocol provided by the manufacturer. Briefly, tissue samples were homogenized, and pelleted cells were lysed by repetitive pipetting. RNA quality was confirmed using the Agilent 2100 BioAnalyzer (Agilent Technologies), and the concentrations were determined using NanoDrop ND-1000 (NanoDrop Technologies).

Expression profiling. Expression profiling of the samples, each with either two or three technical replicates, were performed using the Illumina Human HT-12 V3 BeadChips (Illumina Inc.) including more than 48,000 probes, where 200 ng of total RNA was processed according to the protocol supplied by Illumina. All samples were randomized before array hybridization, and the technical replicates were always hybridized on different BeadChips. Log2-transformed expression signals were normalized with quantile normalization of the replicates of each individual followed by quantile normalization across all individuals. Post–quality-control expression profiles were obtained for 825 individuals. The Illumina probe annotations were cross-checked by mapping the probe sequence to the NCBI build 36 genome with MAQ[20]. Only the uniquely mapping probes with no mismatches and either an Ensembl or RefSeq ID were kept for analysis. Probes mapping to genes of uncertain function (LOC symbols) and those encompassing a common SNP were further excluded, leaving 23,552 probes to be used in the analysis.

Genetic association. All MuTHER eQTL associations were performed with the GenABEL and ProbABEL packages[21,22] using both twins from each pair in a polygenic model incorporating a kinship matrix in GenABEL followed by the ProbABEL mmscore option. Age and experimental batch were included as cofactors in the analysis. Seven hundred seventy six samples had both expression profiles and genotypes. Analysis of the deCODE samples, including parent-of-origin analyses, were performed as previously described[3].

Lead SNP selection. The *trans* associations did not have a single lead SNP. Each probe had barely distinguishable *P* values at five SNPs in high linkage disequilibrium ($r^2 > 0.96$, $D' = 1$). Conditioning the *trans* associations on any of the five lead SNPs (rs13234269, rs4731702, rs972283, rs11979110 and rs13233731) completely removed the signal at the other SNPs. rs13234269 had slightly lower *P* values for most probes but was not present in the deCODE replication data. The lead SNP in the T2D study[1] was rs972283, and the parent-of-origin study[3] and the HDL study[2] had rs4731702 as the lead SNP. As rs4731702 and rs972283 are in perfect linkage disequilibrium ($r^2 = 1$) and rs4731702 was directly genotyped in MuTHER and deCODE, whereas rs972283 was imputed in only some of the samples, we chose to report results with respect to rs4731702.

Isoform and probe differences. The Illumina HT12 array contains three probes for *ARSD* and *PRMT2* that are designed to assay different isoforms. Expression levels measured by the two alternate probes for *ARSD* were not associated in *trans* with rs4731702 (ILMN_1684956/NM_001669, $P = 0.53$; ILMN_1661624/NM_009589, $P = 0.13$; deCODE $P = 0.45$ and $P = 0.83$, respectively), whereas one of the two *PRMT2* alternate probes was associated in *trans* with rs4731702 (ILMN_2393544/NM_001535, $P = 6.5 \times 10^{-4}$; ILMN_2259119/NM_206962, $P = 0.59$; no alternate probes in deCODE).

The Illumina HT12 probe targeting *KLF14* (ILMN_1681168) did not detect any expression in the MuTHER samples. As expression of *KLF14* in adipose tissue has been previously reported in the deCODE data as well as three independent studies of adipose tissue in the GEO database[23], we believe the apparent lack of expression in the MuTHER data is because of a technical failure of the Illumina probe. The Illumina *KLF14* probe and the Agilent *KLF14* probe used in the deCODE data target nonoverlapping locations in the single exon of *KLF14*.

Heritability. Expression heritability was calculated with a linear mixed effects model as described[24], including age and experimental batch as covariates. All available complete twin pairs were included, corresponding to 143 monozygotic pairs and 214 dizygotic pairs.

Phenotypic association. Associations between mRNA expression and phenotypes were modeled using a linear mixed effects model, using R (see URLs) and the *lmer()* function in the lme4 package, fitted by maximum likelihood. The linear mixed effects model was adjusted for age and experimental batch (fixed effects) and family relationship (twin pairing) and zygosity (random effects). A likelihood ratio test was used to assess the significance of the phenotype effect. The *P* value of the phenotype effect in each model was calculated from the χ^2 distribution with 1 degree of freedom using −2log (likelihood ratio) as the test statistic. A Fisher exact test was used to assess enrichment of phenotype associations in the GWST gene set versus the full set of genes in adipose tissue.

GWAS lookups. The ten genome-wide significant *trans* genes and their ~250-kb flanking regions were interrogated for SNPs associated with metabolic phenotypes in large GWAS meta-analyses[1,2,11–13] (the GIANT, MAGIC, DIAGRAM and Lipids consortia). A study-wide significance of $P = 1.03 \times 10^{-4}$ was calculated by Bonferroni correcting for the 487 independents tests in the ten regions. The number of independent tests was approximated by the number of tag SNPs needed to tag all CEU HapMap SNPs with minor allele frequency (MAF) > 0.05 in the ten regions with an r^2 of 0.5 as calculated by the Tagger algorithm[25]. Alternatively, requiring an r^2 of 0.8 resulted in 880 tests for a threshold of $P = 5.7 \times 10^{-5}$. The region around *ARSD* was included despite the lack of X chromosome results in most studies. All genes within the ~250-kb flanking regions are listed in **Supplementary Table 1**.

Transcription factor binding sites. We searched 450 bp upstream and 50 bp downstream of the transcription start site of the GWST genes for over-represented JASPAR binding sites using PSCAN[7].

Genotyping. Genotyping of the TwinsUK dataset (n = ~6,000) was done with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M). Intensity data for each of the three arrays were pooled separately (with 1M-Duo and 1.2MDuo 1M pooled together), and genotypes were called with the Illuminus[26] calling algorithm, thresholding on a maximum posterior probability of 0.95. Similar exclusion criteria were applied to each of the three datasets separately. Exclusion criteria for samples were: (i) sample call rate <98%; (ii) heterozygosity across all SNPs ≥2 standard deviations from the sample mean; (iii) evidence of non-European ancestry as assessed by principal component analysis comparison with HapMap3 populations; and (iv) observed pairwise identity by descent probabilities suggestive of sample identity errors. Exclusion criteria for SNPs were: (i) Hardy-Weinberg $P < 10^{-6}$, assessed in a set of unrelated samples; (ii) MAF < 1%, assessed in a set of unrelated samples; or (iii) SNP call rate <97% (SNPs with MAF ≥ 5%) or <99% (for 1% ≤ MAF < 5%).

Prior to merging the three datasets, we performed pairwise comparison among the three datasets and further excluded SNPs and samples as follows: (i) concordance at duplicate samples <1%; (ii) concordance at duplicate SNPs <1%; (iii) visual inspection of quantile-quantile plots for logistic regression applied to all pairwise dataset comparisons; (iv) Hardy-Weinberg $P < 10^{-6}$,

© 2011 Nature America, Inc. All rights reserved.

assessed in a set of unrelated samples; or (v) observed pairwise identity by descent probabilities suggestive of sample identity errors.

Genotype imputation. Imputation was performed using the IMPUTE software package (v2)[27] using two reference panels, P0 (HapMap2, rel 22, combined CEU, YRI and, ASN panels) and P1 (610K+, including the combined HumanHap610K and 1M array). After imputation, SNPs were filtered at a MAF > 5% and an IMPUTE info value of >0.8.

14. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res. Hum. Genet.* **9**, 899–906 (2006).
15. Skidmore, P.M. *et al.* Relation of birth weight, body mass index, and change in size from birth to adulthood to insulin resistance in a female twin cohort. *J. Clin. Endocrinol. Metab.* **93**, 516–520 (2008).
16. Aulchenko, Y.S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **41**, 47–55 (2009).
17. Richards, J.B., Valdes, A.M., Burling, K., Perks, U.C. & Spector, T.D. Serum adiponectin and bone mineral density in women. *J. Clin. Endocrinol. Metab.* **92**, 1517–1523 (2007).
18. Prokopenko, I. *et al.* Variants in *MTNR1B* influence fasting glucose levels. *Nat. Genet.* **41**, 77–81 (2009).
19. Falchi, M., Wilson, S.G., Paximadas, D., Swaminathan, R. & Spector, T.D. Quantitative linkage analysis for pancreatic B-cell function and insulin resistance in a large twin cohort. *Diabetes* **57**, 1120–1124 (2008).
20. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
21. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
22. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
23. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.* **39**, D1005–D1010 (2011).
24. Visscher, P.M., Benyamin, B. & White, I. The use of linear mixed models to estimate variance components from data on twin pairs by maximum likelihood. *Twin Res.* **7**, 670–674 (2004).
25. de Bakker, P.I. *et al.* Efficiency and power in genetic association studies. *Nat. Genet.* **37**, 1217–1223 (2005).
26. Teo, Y.Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741–2746 (2007).
27. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).

© 2011 Nature America, Inc. All rights reserved.

doi:10.1038/ng.833

Corrigendum: Identification of an imprinted master *trans* regulator at the *KLF14* locus related to multiple metabolic phenotypes

Kerrin S Small, Åsa K Hedman, Elin Grundberg, Alexandra C Nica, Gudmar Thorleifsson, Augustine Kong, Unnur Thorsteindottir, So-Youn Shin, Hannah B Richards, the GIANT Consortium, the MAGIC Investigators, the DIAGRAM Consortium, Nicole Soranzo, Kourosh R Ahmadi, Cecilia M Lindgren, Kari Stefansson, Emmanouil T Dermitzakis, Panos Deloukas, Timothy D Spector & Mark I McCarthy for the MuTHER Consortium

***Nat. Genet.* 43, 561–564 (2011); published online 15 May 2011; corrected after print 22 September 2011**

In the version of this article initially published, there were several errors in the *P* values reported in the Adiponectin and HOMA-IR columns of Table 3. These errors have been corrected in the HTML and PDF versions of the article.

© 2011 Nature America, Inc. All rights reserved.