SUPPLEMENTARY DATA

Genetic regulation of enoyl-CoA hydratase domain-containing 3 in adipose tissue determines insulin sensitivity in African Americans and Europeans

Neeraj K. Sharma[1], Chia-Chi Chuang Key[2], Mete Civelek[3], Martin Wabitsch[4] , Mary E. Comeau[5], Carl D. Langefeld[5], John S. Parks[2], and Swapan K. Das [1,*]

[1] Department of Internal Medicine, Section of Endocrinology and Metabolism, Wake Forest School of Medicine, Winston-Salem, NC 27157

[2] Department of Internal Medicine, Section of Molecular Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157

[3] Center for Public Health Genomics, Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908

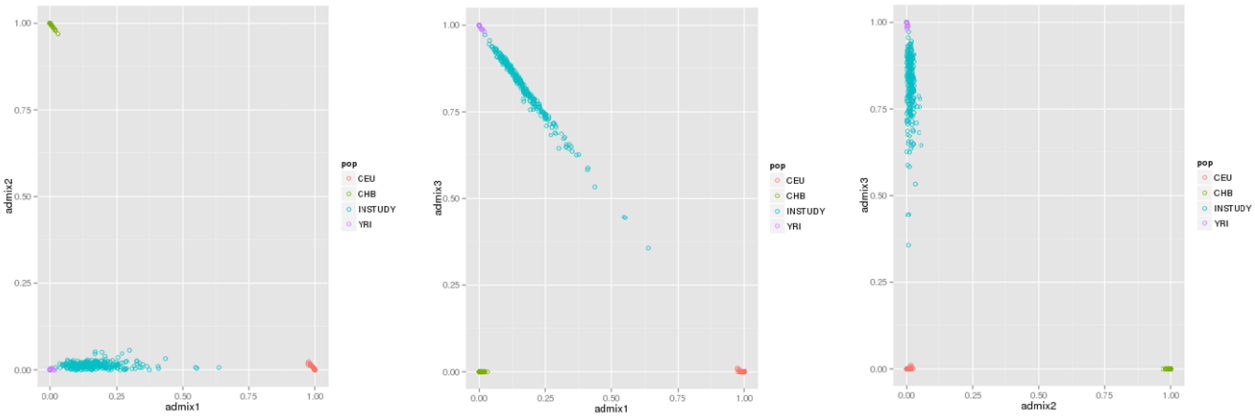
[4] Division of Pediatric Endocrinology and Diabetes, Department of Pediatrics and Adolescent Medicine, University Medical Center Ulm, Eythstr. 24, D-89075 Ulm, Germany.

[5] Department of Biostatistics and Data Science, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157

[*] Corresponding author and person to whom reprint requests should be addressed:
Swapan K. Das, Ph.D.
Section on Endocrinology and Metabolism
Department of Internal Medicine, Wake Forest School of Medicine
Medical Center Boulevard, Winston-Salem, North Carolina 27157
Email: sdas@wakehealth.edu, Telephone: 336-713-6057; Fax: 336-713-7200

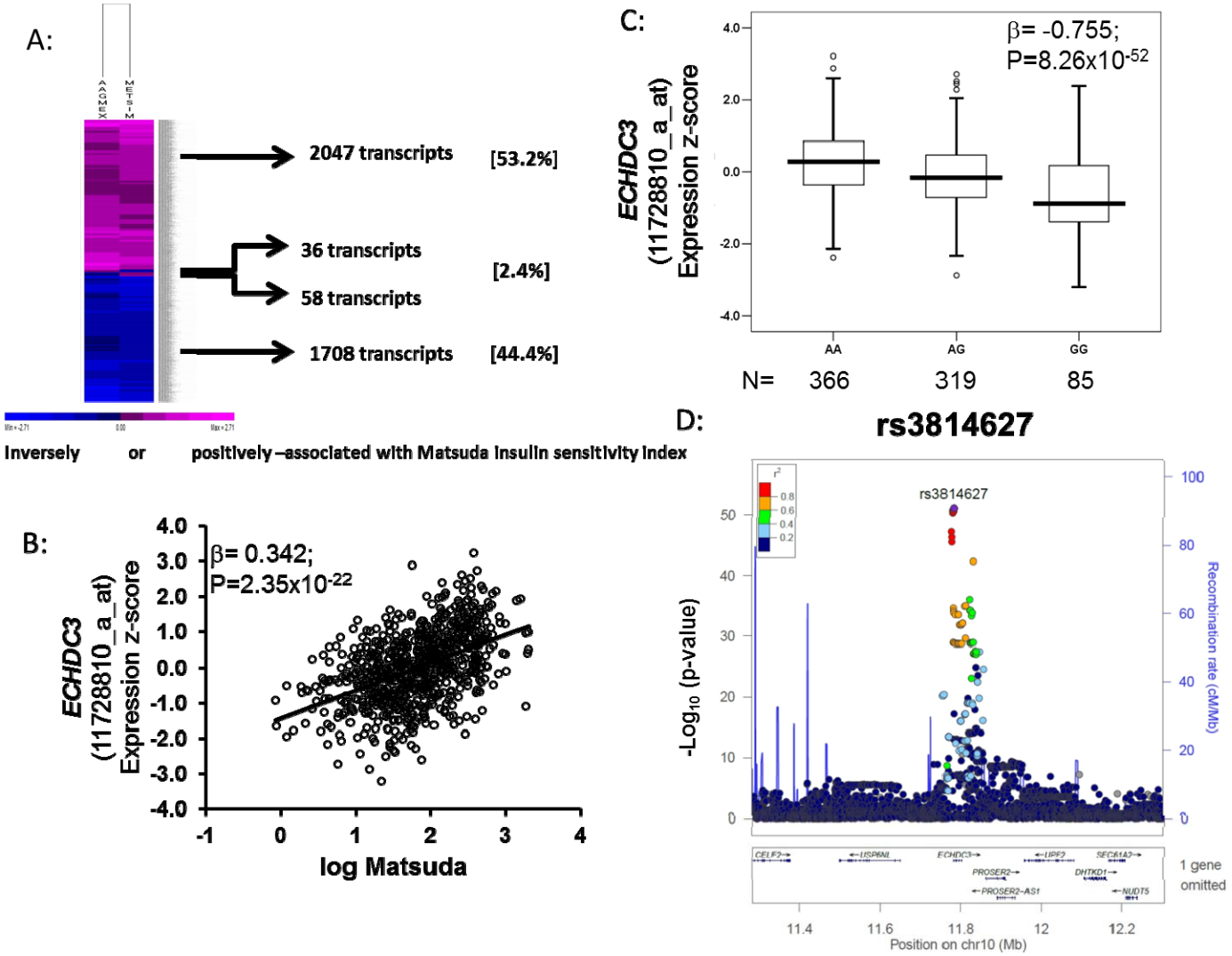
SUPPLEMENTARY DATA

Supplementary Figure S1. Plots of Admixture coefficients. HapMap Phase 3 CEU (Caucasian), YRI (African), and CHB (Chinese) samples were used as reference panels and were merged with AAGMEx study samples and admixture estimates were computed using the software ADMIXTURE (http://software.genetics.ucla.edu/admixture/; Alexander et al., Genome Research, 19:1655–1664, 2009). AAGMEx participants are labeled as INSTUDY in the plots. Three samples with >50% European ancestry proportion were excluded from further analyses.



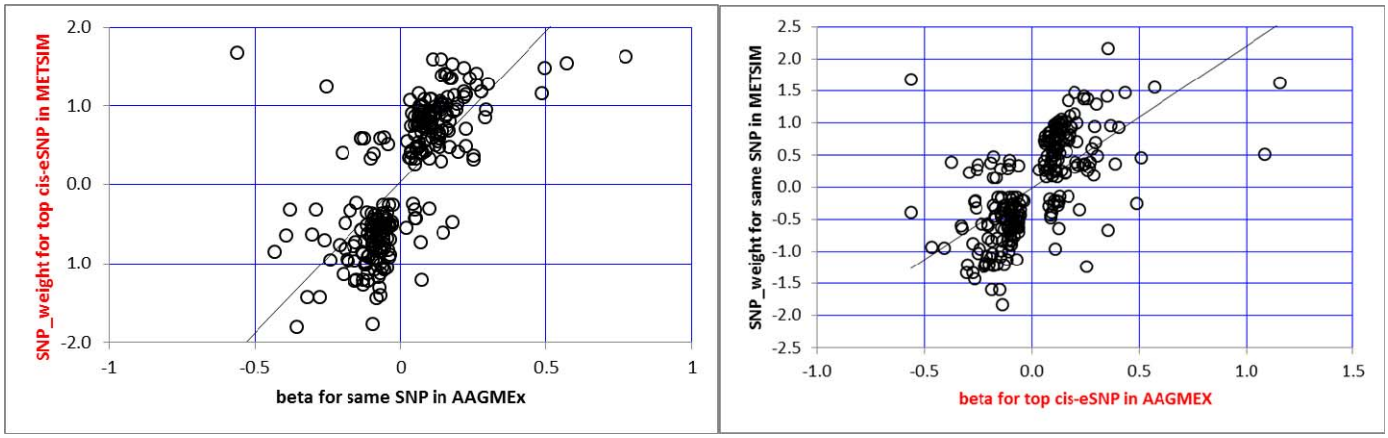
SUPPLEMENTARY DATA

Supplementary Figure S2. Correlation of adipose tissue transcript levels with Matsuda index and their genetic regulation in African Americans replicates in Metabolic Syndrome in Men (METSIM), an independent European ancestry cohort. A) Heat map from unsupervised hierarchical clustering showing high concordance in the direction of correlation of 3,849 adipose tissue transcripts correlated with Matsuda index (q<0.05) in both AAGMEx and METSIM cohorts. B) The scatter plot shows correlation of *ECHDC3* transcript expression (11728810_a_at) in adipose tissue with Matsuda index in METSIM. **C)** The box plot shows association of ECHDC3 transcript expression in adipose with genotype of the top *cis*-eSNP rs3814627 in METSIM. The box represents the interquartile range, which contains 50% of the values. The whiskers are lines that extend the box to the highest and lowest values, excluding outliers. A line across the box indicates the median. **D)** LocusZoom plots show regional association of *ECHDC3 cis*-eQTL region SNPs with transcript expression in METSIM cohort.



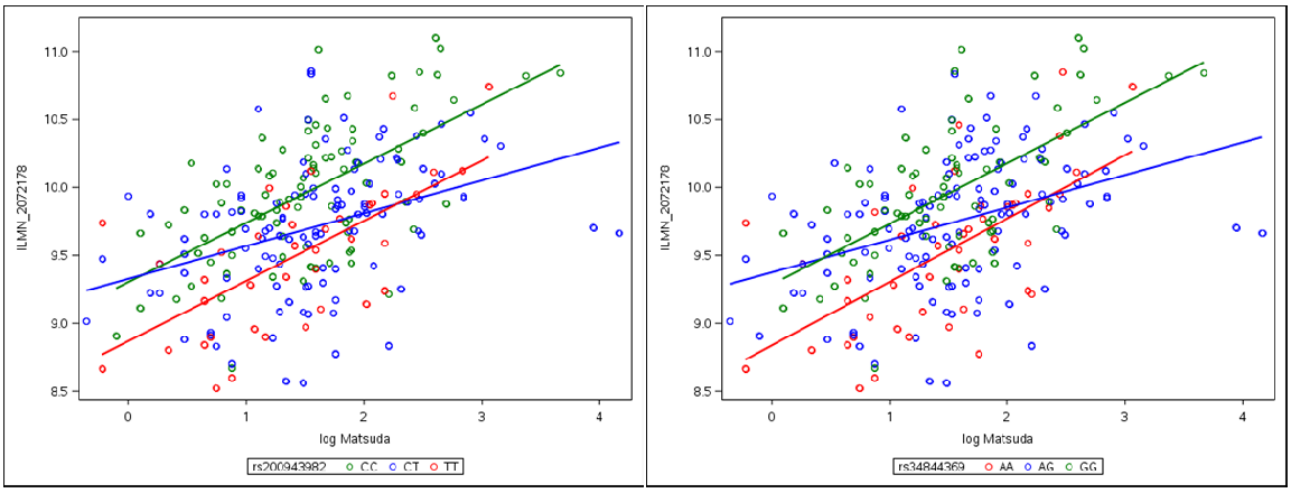
SUPPLEMENTARY DATA

Supplementary Figure S3. Comparison of the effect of top cis-eSNPs for Matsuda index-correlated eGenes in AAGMEx and METSIM cohort. A different genotyping platform was used in the two cohorts, and based on genotype data availability; of the total 587 common Matsuda index-correlated cis-eGenes, we were able to retrieve results for 498 METSIM top- cis eSNPs from the AAGMEx dataset and 363 AAGMEx top-cis-eSNPs from the METSIM dataset. After carefully matching the effect allele we observed that 52% (259/498) of the METSIM top cis-eSNPs were also significant in AAGMEx ($p<0.05$). **(A)** METSIM top cis-eSNPs replicated in AAGMEx shows 91.5% concordance in effect direction and strong correlation in effect size ($r=0.70$). Similarly, 65.8% (239/363) of the AAGMEx top cis-eSNPs were also significant in the METSIM cohort ($p<0.01$). **(B)** AAGMEx cis-eSNPs replicated in METSIM shows 86.6% concordance in effect direction and a strong correlation in effect size ($r=0.62$).



SUPPLEMENTARY DATA

Supplementary Figure S4. Enoyl-CoA hydratase domain-containing-3 (ECHDC3) transcript expression in adipose tissue is correlated with Matsuda index of insulin sensitivity in all genotype groups. The scatter plot shows correlation of ECHDC3 transcript expression (ILMN_2072178) in adipose tissue with Matsuda index in AAGMEx. Results for genotype groups based on top genotyped (rs200943982) and imputed (rs34844369) *ECHDC3* cis-eSNP are shown in different colors. Results from linear regression analysis are also shown in a tabular format.



SNP	Genotype	N	β	SE(β)	R-square full	Adjusted R-square	P-value
rs200943982	CC	89	0.96284	0.13725	0.3910	0.3620	5.44×10^{-10}
	CT	113	0.57119	0.14921	0.1646	0.1337	2.17×10^{-4}
	TT	39	0.91404	0.19380	0.4534	0.3891	3.99×10^{-5}
rs34844369	GG	75	0.93036	0.14756	0.4004	0.3662	2.25×10^{-8}
	AG	122	0.64326	0.14759	0.1559	0.1270	2.83×10^{-5}
	AA	46	0.78834	0.16290	0.4760	0.4249	1.88×10^{-5}

SUPPLEMENTARY DATA

(in Microsoft –Excel file: MATSUDA-ECHDC3 paper-Supplementary tables-final-Diabetes.xlsx; can be viewed in Google drive link:
https://drive.google.com/file/d/1ddJ8xaz4AvsZhSlraPfdIxfyuaqiScsu/view?usp=sharing)

Supplementary Table 1: Expression levels of subcutaneous adipose tissue transcripts are correlated with Matsuda Insulin sensitivity index in AAGMEx. Transcripts of genes with Entrez gene id and correlated (q<0.01) with Matsuda index, based on regression analysis (adjusted for age, gender and ancestry proportion/admixture) are shown. Results in Column L-Q are based on secondary analysis adjusted for age, gender, BMI, and admixture.

Supplementary Table 2A: Adipose tissue transcripts correlated with Matsuda insulin sensitivity index in AAGMEx are enriched for genes in biological pathways. Results from Ingenuity Pathway Analysis (IPA) for adipose tissue genes correlated with Matsuda index (q<0.01 in Supplemental Table 1) are shown. Significant pathways (two-sided Fisher's exact test p-value < 0.05) are shown. Predicted direction of activation and number of genes negatively ($\beta < 0$) or positively ($\beta > 0$) correlated with Matsuda index are also shown.

Supplementary Table 2B: Adipose tissue transcripts correlated with Matsuda insulin sensitivity index in AAGMEx are enriched for genes in biological pathways. Results from DAVID v6.9 analysis for adipose tissue genes correlated with Matsuda index (q<0.01 in Supplemental Table 1) are shown. Selected significant pathways (Benjamini-Hochberg adjusted p-value < 0.05) are shown.

Supplementary Table 3: Comparison of correlation of subcutaneous adipose tissue transcripts with Matsuda insulin sensitivity index in African Americans from AAGMEx cohort and Europeans from METSIM cohort. Most significantly correlated probes for each gene in METSIM study (from Civelek et al., 2017) are shown. Sorted list of genes significant (FDR-P /q -value <0.05) in both cohort are shown. List sorted on average ranking on significant correlation in both cohorts.

Supplementary Table 4: Matsuda Insulin sensitivity-correlated transcripts are *cis*-eQTL in adipose tissue from both African American (AAGMEx cohort) and European ancestry (METSIM) individuals. *Sorted list of 587 Matsuda index-correlated cis-eGenes based on average rank of genes on Matsuda index correlation and eQTL analysis p-values in AAGMEx and METSIM cohort.* Matsuda insulin sensitivity-correlated transcripts of Entrez_ID genes in adipose tissue associated with a SNP (Q-value<0.04) within ±500kb of the 5' and 3' end of the transcript in AAGMEx are shown. Data on Europeans ancestry males are from METSIM cohort (Civelek et al., 2017). Data for most significantly associated genotyped cis-eSNP in AAGMEx and best cis-eSNP in FaST-LMM eQTL analysis in METSIM are presented. In AAGMEx: ρ, Spearman partial correlation coefficient; p value, significance level of correlation of transcript level with Matsuda index in Spearman partial correlation analysis; A1, Minor Allele; A2, Major Allele; MAF, Minor Allele Frequency; beta, effect size of minor allele (A1); eQTL p-value, significance in additive model (in MatrixEQTL analysis). Count Of *cis*-eSNP, Number of genotyped *cis*-SNPs (MAF>0.01) associated with transcript at Q-value ≤0.04. In METSIM: β, beta value; and p-value, significance level of correlation of transcript level with Matsuda index in linear regression analysis; alt_allele, alternative allele; β, effect size for alt_allele of best *cis*-eSNP in FaST-LMM eQTL analysis.

Supplementary Table 4A: Most significantly associated cis-eSNP for Matsuda index-correlated cis-eGenes in METSIM are shown and effect of the same SNP in AAGMEx are shown for comparison.

Supplementary Table 4B: Most significantly associated genotyped cis-eSNP for Matsuda index-correlated cis-eGenes in AAGMEx are shown and effect of the same SNP in METSIM are shown for comparison.

Supplementary Table 4C: Matsuda Insulin sensitivity-correlated transcripts are cis-eQTL in adipose tissue from African American (AAGMEx cohort) individuals in BMI adjusted analyses.

Supplementary Table 5: Expression of ECHDC3 Transcripts levels in adipose tissues of African Americans from AAGMEx participants are genetically regulated. Results for significant ($p < 1X10^{-4}$) genotyped or imputed *cis*-eSNPs for ECHDC3 (ILMN_2072178) from expanded *cis*-eQTL analysis

are presented. MAF, Minor allele frequency; A1, Minor Allele; A2, Major Allele; MAF, Minor Allele Frequency; beta, effect size of minor allele (A1); eQTL p-value, significance in additive model (in MatrixEQTL analysis); TSS, transcription start site. Data in Column L-V shows association statistics of respective cis-eSNPs with MATSUDA insulin sensitivity index in AAGMEx cohort. RAF, reference allele frequency; Pvalue, significance from best genetic model; Add_Pvalue, p-value from additive genetic model.

Supplementary Table 6: *cis*-eQTL for *ECHDC3* in adipose tissue samples of European ancestry individuals from METSIM cohort. eQTL results are from Civelek M et al, Am J Hum Genet 2017;100(3):428-43. *ECHDC3 cis*-eSNPs ($p<1x10^{-3}$) from subcutaneous adipose tissue eQTL analysis in METSIM study is shown. MAF= Minor Allele Frequency; effect_size=allele effect for alt_allele. Data in Column L-N shows association statistics of respective cis-eSNPs with MATSUDA insulin sensitivity index in METSIM cohort. P-values from two linear regression models are shown.

Supplementary Table 7: Epigenetic regulatory annotation of ECHDC3 *cis*-eSNPs identified in AAGMEx, METSIM and GTEx study.

A) Annotation based on Haploreg-v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) and

B) Annotation based on RegulomeDb-v1.1 (http://regulomedb.org/)

C) Annotation based on SNP2TFBS (https://ccg.vital-it.ch/snp2tfbs/) web interface aimed at studying variations (SNPs/indels) that affect transcription factor binding (TFB) in the Human genome. ECHDCE3 cis-eSNPs with ScoreDifference for TF binding > or < 0, estimated based on position weight matrix (PWM) models are shown

Supplementary Table 8: RNA-seq analysis showing transcripts for protein coding genes differentially expressed in mature SGBS adipocytes (14^{th} day differentiation) expressing ECHDC3-shRNA compared to SGBS adipocytes expressing control-shRNA. Differentially expressed coding genes (DEGs) with Bayes posterior probability ≥70% in NOISeq and average $\log_2$ fold change (log2FC) +/-0.58 are shown.

Supplementary Table 9: Ingenuity pathway analyses (IPA) suggest enrichment of canonical pathways among genes differentially expressed in mature SGBS adipocytes (14^{th} day differentiation) expressing ECHDC3-shRNA compared to SGBS adipocytes expressing control-shRNA. Selected categories (-log p-value>1.3) are shown.

Supplementary Table 10: Ingenuity pathway analyses (IPA) suggest enrichment of disease or biological function among genes differentially expressed in mature SGBS adipocytes (14days differentiation) expressing ECHDC3-shRNA compared to SGBS adipocytes expressing control-shRNA. Selected categories (metabolism or lipid) are shown.

Supplementary Table 11: Considering DEGs in ECHDC3-shRNA expressing SGBS cells, regulator effects analysis in IPA suggest the effect of upstream regulators on expression of downstream target molecules and its impact on biological function and diseases.