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

ACKNOWLEDGMENTS

We thank all the participants and the staff of the BioBank Japan project. The project was supported by a grant from the Leading Project of Ministry of Education, Culture, Sports, Science and Technology Japan.

The Japan Cardiometabolic Genome Epidemiology (CAGE) Network Studies were supported by grants for the Program for Promotion of Fundamental Studies in Health Sciences, National Institute of Biomedical Innovation Organization (NIBIO); the Core Research for Evolutional Science and Technology (CREST) from the Japan Science Technology Agency; the Grant of National Center for Global Health and Medicine (NCGM).

We thank the Office of Population Studies Foundation research and data collection teams for the Cebu Longitudinal Health and Nutrition Survey. This work was supported by US National Institutes of Health grants DK078150, TW05596, HL085144 and TW008288 and pilot funds from grants RR20649, ES10126, and DK56350.

We acknowledge support from the Hong Kong Government Research Grants Council Central Allocation Scheme (CUHK 1/04C), Research Grants Council Earmarked Research Grant (CUHK4724/07M) and the Innovation and Technology Fund of the Government of the Hong Kong Special Administrative Region (ITS/487/09FP). We acknowledge the Chinese University of Hong Kong Information Technology Services Center for support of computing resources. We would also like to thank the dedicated medical and nursing staff at the Prince of Wales Hospital Diabetes and Endocrine Centre.

This work was supported by grants from Korea Centers for Disease Control and Prevention (4845-301, 4851-302, 4851-307) and an intramural grant from the Korea National Institute of Health (2011-N73005-00), the Republic of Korea.

The work by the National Taiwan University Hospital was supported in part by the grant (NSC99-3112-B-002-019) from the National Science Council of Taiwan. We would also like to acknowledge the National Genotyping Center of National Research Program for Genomic Medicine (NSC98-3112-B-001-037), Taiwan.

The work by the Shanghai Diabetes Genetic Study was supported in part by the US National Institutes of Health grants R01CA124558, R01CA64277, R01CA70867, R01CA90899, R01CA100374, R01CA118229, R01CA92585, UL1 RR024975, DK58845 and HG004399, the Department of Defense Idea Award BC050791, Vanderbilt Ingram professorship funds and the Allen Foundation Fund. We thank the dedicated investigators and staff members from research teams at Vanderbilt University, Shanghai Cancer Institute and the Shanghai Institute of Preventive Medicine, and especially, the study participants for their contributions in the studies.

The work of the Shanghai Diabetes Study was supported by grants from the National 973 Program (2011CB504001), the Project of National Natural Science Foundation of China (30800617) and the Shanghai Rising-Star Program (09QA1404400), China.

The work of the Shanghai Jiao Tong University Diabetes Study was supported by grants from the National 863 Program (2006AA02A409) and the major program of the Shanghai Municipality for Basic Research (08dj1400601), China.

The work of the Seoul National University Hospital was supported by grants from the Korea Health 21 R&D Project, Ministry of Health & Welfare (00-PJ3-PG6-GN07-001) and the World Class University project of the Ministry of Education, Science and Technology (MEST) and National Research Foundation (NRF) (R31-2008-000-10103-0), Korea. The Singapore Prospective Study Program (SP2) was funded through grants from the Biomedical Research Council of Singapore (BMRC05/1/36/19/413 and 03/1/27/18/216) and the National Medical Research Council of Singapore (NMRC/1174/2008). E.S.T. also receives additional support from the National Medical Research Council through a clinicians scientist award (NMRC/CSA/008/2009). The Singapore Malay Eye Study (SiMES) was funded by the National Medical Research Council (NMRC0796/2003 and NMRC/STaR/0003/2008) and Biomedical Research Council (BMRC, 09/1/35/19/616). Y.Y.T. acknowledges support from the Singapore National Research Foundation, NRF-RF-2010-05. The Genome Institute of Singapore carried out all the genotyping for the samples from Singapore also provided funding for the genotyping of the samples from SP2.

AUTHOR CONTRIBUTIONS

The study was supervised by E.S.T., B.-G.H., N.K., Y.S.C., Y.Y.T., W.Z., Q.C., X.O.S., Y.-T.C., J.-Y.W., L.S.A., K.L.M., T.K., C.H., W.J., L.-M.C., Y.M.C., K.S.P., J.-Y.L. and J.C.N.C. The experiments were conceived of and designed by Y.S.C., E.S.T., N.K., D.P.-K.N., J.J.-M.L., M.S., T.Y.W., Y.Y.T., W.Z., F.B.H., X.O.S., C.-H.C., F.-J.T., Y.-T.C., J.-Y.W., L.S.A., K.L.M., S.M., C.H., L.-M.C., K.S.P., M.J.G., M.I.M. and R.C.W.M. The experiments were performed by J.L., M.S., J.J.L., J.-Y.W., S.M., R.Z., K.Y., Y.-C.C., T.-J.C., L.-M.C. and S.H.K. Statistical analyses was performed by M.J.G., X.S., Y.J.K., R.T.H.O., W.T.T., Y.Y.T., F.T., J.L., C.-H.C., L.-C.C., Y.W., Y.L., K.H., C.H., Y.-C.C., S.H.K., A.P.M. and R.C.W.M. The data were analyzed by M.J.G., X.S., Y.J.K., R.T.H.O., W.T.T., Y.Y.T., J.L., C.-H.C., L.-C.C., Y.W., N.R.L., Y.L., L.S.A., K.L.M., T.Y., C.H., Y.-C.C., S.H.K., Y.S.C., S.K., Å.K.H. and R.C.W.M. The reagents, materials and analysis tools were contributed by E.S.T., B.-G.H., N.K., D.P.-K.N., J.J.-M.L., J.L., M.S., T.A., T.Y.W., E.N., M.Y., J.N., J.J.L., W.Z., Q.C., Y.G., W.L., F.B.H., X.O.S., F.-J.T., Y.-T.C., J.-Y.W., N.R.L., Y.L., K.O., H.I., R.T., C.W., Y.B., T.-J.C., L.-M.C., K.S.P., H.-L.K., N.H.C., J.-Y.L., W.Y.S. and J.C.N.C. The manuscript was written by Y.S.C., M.S. and E.S.T. All authors reviewed the manuscript.

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://www.nature.com/reprints/index.html.

1. Kang, H.S. *et al.* Transcription factor Glis3, a novel critical player in the regulation of pancreatic beta-cell development and insulin gene expression. *Mol. Cell. Biol.* **29**, 6366–6379 (2009).
2. Yang, Y., Chang, B.H., Samson, S.L., Li, M.V. & Chan, L. The Kruppel-like zinc finger protein Glis3 directly and indirectly activates insulin gene transcription. *Nucleic Acids Res.* **37**, 2529–2538 (2009).
3. 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).
4. Barker, A. *et al.* Association of genetic loci with glucose levels in childhood and adolescence: a meta-analysis of over 6,000 children. *Diabetes* **60**, 1805–1812 (2011).
5. Takeuchi, F. *et al.* Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* **58**, 1690–1699 (2009).
6. Barroso, I. *et al.* Population-specific risk of type 2 diabetes conferred by HNF4A P2 promoter variants: a lesson for replication studies. *Diabetes* **57**, 3161–3165 (2008).
7. Silander, K. *et al.* Genetic variation near the hepatocyte nuclear factor-4 α gene predicts susceptibility to type 2 diabetes. *Diabetes* **53**, 1141–1149 (2004).
8. Zimmet, P., Alberti, K.G. & Shaw, J. Global and societal implications of the diabetes epidemic. *Nature* **414**, 782–787 (2001).
9. Tkác, I. Metabolic syndrome in relationship to type 2 diabetes and atherosclerosis. *Diabetes Res. Clin. Pract.* **68** (suppl. 1), S2–S9 (2005).
10. Prokopenko, I., McCarthy, M.I. & Lindgren, C.M. Type 2 diabetes: new genes, new understanding. *Trends Genet.* **24**, 613–621 (2008).
11. Rung, J. *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat. Genet.* **41**, 1110–1115 (2009).
12. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
13. Yasuda, K. *et al.* Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **40**, 1092–1097 (2008).
14. Unoki, H. *et al.* SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in east Asian and European populations. *Nat. Genet.* **40**, 1098–1102 (2008).
15. Yamauchi, T. *et al.* A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at *UBE2E2* and *C2CD4A-C2CD4B*. *Nat. Genet.* **42**, 864–868 (2010).
16. Tsai, F.J. *et al.* A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet.* **6**, e1000847 (2010).
17. Shu, X.O. *et al.* Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet.* **6**, e1001127 (2010).
18. Stommel, M. & Schoenborn, C.A. Variations in BMI and prevalence of health risks in diverse racial and ethnic populations. *Obesity (Silver Spring)* **18**, 1821–1826 (2010).
19. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
20. Kadereit, B. *et al.* Evolutionarily conserved gene family important for fat storage. *Proc. Natl. Acad. Sci. USA* **105**, 94–99 (2008).
21. Nakajima, H. *et al.* Hepatocyte nuclear factor-4 α gene mutations in Japanese non-insulin dependent diabetes mellitus (NIDDM) patients. *Res. Commun. Mol. Pathol. Pharmacol.* **94**, 327–330 (1996).
22. Johansson, S. *et al.* Studies in 3,523 Norwegians and meta-analysis in 11,571 subjects indicate that variants in the hepatocyte nuclear factor 4 α (HNF4A) P2 region are associated with type 2 diabetes in Scandinavians. *Diabetes* **56**, 3112–3117 (2007).
23. Girard, C. *et al.* Genomic and functional characteristics of novel human pancreatic 2P domain K^+ channels. *Biochem. Biophys. Res. Commun.* **282**, 249–256 (2001).
24. Ashcroft, F.M. ATP-sensitive potassium channelopathies: focus on insulin secretion. *J. Clin. Invest.* **115**, 2047–2058 (2005).
25. Soni, S. *et al.* Absence of erythroblast macrophage protein (Emp) leads to failure of erythroblast nuclear extrusion. *J. Biol. Chem.* **281**, 20181–20189 (2006).
26. Nam, D., Kim, J., Kim, S.Y. & Kim, S. GSA-SNP: a general approach for gene set analysis of polymorphisms. *Nucleic Acids Res.* **38**, W749–W754 (2010).
27. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).

28. Luke, M.R., Houghton, F., Perugini, M.A. & Gleeson, P.A. The trans-Golgi network GRIP-domain proteins form α-helical homodimers. *Biochem. J.* **388**, 835–841 (2005).
29. Jo, W., Endo, M., Ishizu, K., Nakamura, A. & Tajima, T. A novel *PAX4* mutation in a Japanese patient with maturity-onset diabetes of the young. *Tohoku J. Exp. Med.* **223**, 113–118 (2011).
30. Wang, X. *et al.* Mass spectrometric characterization of the affinity-purified human 26S proteasome complex. *Biochemistry* **46**, 3553–3565 (2007).
31. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
32. Frayling, T.M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
33. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).

[1]Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Cheongwon-gun, Gangoe-myeon, Yeonje-ri, Korea. [2]Institute of Biomedical Sciences, Academia Sinica, Nankang, Taipei, Taiwan. [3]School of Chinese Medicine, China Medical University, Taichung, Taiwan. [4]Shanghai Diabetes Institute, Shanghai Key Laboratory of Diabetes Mellitus, Department of Endocrinology and Metabolism, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China. [5]Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [6]Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore. [7]Centre for Molecular Epidemiology, National University of Singapore, Singapore, Singapore. [8]Research Institute, National Center for Global Health and Medicine, Shinjuku-ku, Tokyo, Japan. [9]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [10]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [11]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. [12]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea. [13]Department of Medicine and Therapeutics, Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China. [14]Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka, Japan. [15]Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina, USA. [16]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. [17]Department of Ophthalmology, National University of Singapore, Singapore, Singapore. [18]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. [19]Department of Nutrition and Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [20]Department of Biochemistry, School of Medicine, Ewha Womans University, Seoul, Korea. [21]School of Systems Biomedical Science, Soongsil University, Dongjak-gu, Seoul, Korea. [22]Department of Epidemiology and Public Health, National University of Singapore, Singapore, Singapore. [23]Office of Population Studies Foundation Inc., University of San Carlos, Cebu City, Philippines. [24]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA. [25]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [26]Shanghai Institute of Preventive Medicine, Shanghai, China. [27]Division of Endocrinology and Diabetes, Department of Internal Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan. [28]Department of Diabetes and Endocrinology, Chubu Rosai Hospital, Nagoya, Japan. [29]Centre for Eye Research Australia, University of Melbourne, East Melbourne, Victoria, Australia. [30]Department of Genome Science, Aichi-Gakuin University, School of Dentistry, Nagoya, Japan. [31]Department of Geriatric Medicine, Graduate School of Medical Sciences, Kyushu University, Higashi-ku, Fukuoka, Japan. [32]Department of Endocrinology, Metabolism and Diabetes, Kinki University School of Medicine, Osaka-sayama, Osaka, Japan. [33]Department of Preventive Medicine, Ajou University School of Medicine, Suwon, Korea. [34]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [35]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK. [36]A full list of members is provided in the **Supplementary Note**. [37]Department of Internal Medicine and Bioregulatory Science, Graduate School of Medical Sciences, Kyushu University, Higashi-ku, Fukuoka, Japan. [38]World Class University program, Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology and College of Medicine, Seoul National University, Seoul, Korea. [39]Graduate Institute of Clinical Medicine, National Taiwan University School of Medicine, Taipei, Taiwan. [40]Laboratory for Endocrinology and Metabolism, RIKEN Center for Genomic Medicine, Yokohama, Japan. [41]Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore. [42]Department of Medicine, National University of Singapore, Singapore, Singapore. [43]Duke-National University of Singapore Graduate Medical School, Singapore, Singapore. [44]Institute for Human Genetics, University of California, San Francisco, California, USA. [45]Blood Systems Research Institute, San Francisco, California, USA. [46]These authors contributed equally to this work. [47]These authors jointly directed this work. Correspondence should be addressed to Y.S.C. (yooncho33@korea.kr), M.S. (mark.seielstad@ucsf.edu) or E.S.T. (eshyong@pacific.net.sg).

ONLINE METHODS

Study subjects. Stage 1 subjects were drawn from eight T2D GWAS participating in the AGEN consortium, which was organized to enable genetic studies on diverse complex traits in 2010. These eight studies included 6,952 cases with T2D and 11,865 controls from the Korea Association Resource Study (KARE), the Singapore Diabetes Cohort Study (SDCS), the Singapore Prospective Study Program (SP2), the Singapore Malay Eye Study (SiMES), the Japan Cardiometabolic Genome Epidemiology Network (CAGE), the Shanghai Diabetes Genetic Study (SDGS), the Taiwan T2D Study (TDS) and the Cebu Longitudinal Health and Nutritional Survey (CLHNS). Subjects in stage 2 included 5,843 cases with T2D and 4,574 controls from three independent GWAS, the BioBank Japan Study (BBJ), the Health2 T2D Study (H2T2DS) and the Shanghai Jiao Tong University Diabetes Study (SJTUDS), for *in silico* replication analysis. Stage 3 included up to 12,284 cases with T2D and 13,172 controls from five different studies, the Japan Cardiometabolic Genome Epidemiology Network (CAGE), the Shanghai Diabetes Study I/II (SDS I/II), the Chinese University of Hong Kong Diabetes Study (CUHKDS), the National Taiwan University Hospital Diabetes Study (NTUHDS) and the Seoul National University Hospital Diabetes Study (SNUHDS), for *de novo* replication analysis. The study design and T2D diagnosis criteria used in each study included in stages 1, 2, and 3 are described in **Supplementary Table 1** and the **Supplementary Note**. Each study obtained approval from the appropriate institutional review boards of each participating institution, and written informed consent was obtained from all participants. The three-stage design of the overall study is depicted in **Supplementary Figure 1**.

Genotyping and imputation. Subjects for the stage 1 and 2 analyses were genotyped with high-density SNP typing platforms that covered the entire human genome. In most of the studies, only unrelated samples with missing genotype call rates below 5% were included for subsequent GWAS analyses. For the genome-wide association meta-analysis, each study participating in stages 1 and 2 performed SNP imputation. IMPUTE, MACH or BEAGLE (see URLs) were used, together with haplotype reference panels from the JPT and CHB samples that are available in the HapMap database (JPT+CHB+CEU and/or YRI, in some studies) on the basis of HapMap build 36 (release 21, 22, 23a or 24). Only imputed SNPs with high genotype information content (proper info > 0.5 for IMPUTE and Rsq > 0.3 for MACH and BEAGLE) were used for the association analysis. Genotyping for the stage 3 analysis was carried out using TaqMan, Sequenom MassARRAY or the Beckman SNP Stream method. All SNPs included in stage 3 had a genotype success rate of >98% (**Supplementary Table 2**).

Statistical analyses, analysis tools and SNP prioritization for stages 2 and 3. Associations between SNPs and T2D were tested by logistic regression with an additive model (1 degree of freedom) after adjustment for sex. Other adjustments were permitted according to the situations in the individual studies. The meta-analysis was performed using an inverse-variance method assuming fixed effects, with a Cochran's *Q* test to assess between-study heterogeneity. METAL software (see URLs) was used for all meta-analyses. A plot of the negative log of the association results from the stage 1 meta-analysis, by chromosome, was generated using WGAViewer software (see URLs).The quantile-quantile plot was constructed by plotting the distribution of observed *P* values for the given SNPs against the theoretical distribution of the expected *P* values for T2D[34]. The genomic control inflation factor, λ, was calculated by dividing the median χ^2 statistics by 0.456 (ref. 35) for individual GWAS, as well as for the stage 1 GWAS meta-analysis. We did not correct for genomic control in the stage 1 analyses because the inflation was modest, suggesting that population structure is unlikely to cause substantial inflation of the stage 1 results (**Supplementary Table 2**). The selection criteria for the lead SNPs to take forward to stage 2 *in silico* replication analysis were as follows: (i) stage 1 meta-analysis $P < 5 \times 10^{-4}$ (based on the divergence between the observed and expected *P* values on the quantile-quantile plot; **Supplementary Fig. 2**); (ii) heterogeneity $P > 0.01$; and (iii) at least seven studies having been included in the stage 1 meta-analysis (**Supplementary Table 4**). After removing known variants associated with T2D, proxies for each lead SNP ($r^2 > 0.8$) were selected using the SNAP software (see URLs). The replication genotyping for stage 3 was performed for the new SNPs having a stage 2 combined $P < 10^{-5}$. Regional association results from genome-wide meta-analysis were plotted using LocusZoom software (see URLs) for SNPs reaching genome-wide significance from the combined meta-analysis of stages 1, 2 and 3.

Principal components analysis. A list of 76,534 common SNPs across the Illumina 550, 610 and 1M and Affymetrix 5.0 and 6.0 arrays were first selected. This set of SNPs in the Asian (CHB+JPT) HapMap II samples was then trained to generate a list of 44,524 SNPs having pairwise LD < 0.3 in a sliding window of 50 SNPs. Individuals from each component study and from HapMap II were plotted based on the first two eigenvectors produced by the principal components analysis.

eQTL analysis. Gene expression information from 776 adipose tissues, 667 skin tissues and 777 lymphoblastoid cell lines was obtained from the MuTHER consortium[36]. The eQTL data for eight of the ten T2D loci identified in this study were available in the MuTHER dataset. Most of those loci passed the filtering criteria, such as MAF > 5% and INFO > 0.8, except for rs16955379, which has MAF = 1.5% in the MuTHER data set. Two of the ten loci that were used in the eQTL analysis, rs6815464 (on chromosome 4) and rs17797882 (on chromosome 16), are not included in the MuTHER data set. Association between each SNP with a significant association to T2D and the normalized mRNA expression values of genes within 1 Mb of the lead SNP were performed with the GenABEL and ProbABEL package (see URLs) using the polygenic linear model incorporating a kinship matrix in GenABEL followed by the ProbABEL mmscore test with imputed genotypes. A multiple-testing correction was applied to the *cis* association results. *P* value thresholds of $P = 5.06 \times 10^{-5}$ in adipose tissue, $P = 3.81 \times 10^{-5}$ in skin and $P = 7.80 \times 10^{-5}$ in lymphoblastoid cell lines correspond to an estimated genome-wide false discovery rate of 1%.

Gene relationships among implicated loci (GRAIL) analysis. A GRAIL analysis was performed as described previously[31,33]. A total of 38 genes within T2D-associated regions were selected for the analysis. Among these genes, 28 were from the previously implicated set (**Supplementary Table 3**), and the other 10 genes were newly implicated in this study (**Table 1**). PubMed abstracts published after December 2006 were omitted from the analysis to reduce confounding by results from T2D GWAS.

34. Hyndman, R.J. & Fan, Y. Sample quantiles in statistical packages. *Am. Stat.* **50**, 361–365 (1996).
35. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* **60**, 155–166 (2001).
36. Nica, A.C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).

doi:10.1038/ng.1019

Expression-based genome-wide association study links the receptor *CD44* in adipose tissue with type 2 diabetes

Keiichi Kodama[a,b], Momoko Horikoshi[c], Kyoko Toda[d,1], Satoru Yamada[e,1], Kazuo Hara[c,1], Junichiro Irie[e,f,1], Marina Sirota[a,b], Alexander A. Morgan[a,b], Rong Chen[a,b], Hiroshi Ohtsu[g], Shiro Maeda[h], Takashi Kadowaki[c], and Atul J. Butte[a,b,2]

[a]Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305; [b]Lucile Packard Children's Hospital, Palo Alto, CA 94304; Departments of [c]Metabolic Diseases and [g]Clinical Trial Data Management, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan; [d]Division of Basic Research, Biomedical Laboratory, Kitasato Institute Hospital, Kitasato University, Tokyo 108-8642, Japan; [e]Diabetes Center, Kitasato Institute Hospital, Tokyo 108-8642, Japan; [f]Department of Internal Medicine, Keio University School of Medicine, Tokyo 160-8582, Japan; and [h]Laboratory for Endocrinology and Metabolism, Center for Genomic Medicine, RIKEN, Yokohama City, Kanagawa 230-0045, Japan

Edited by Alan D. Attie, University of Wisconsin, Madison, WI, and accepted by the Editorial Board March 19, 2012 (received for review September 7, 2011)

Type 2 diabetes (T2D) is a complex, polygenic disease affecting nearly 300 million people worldwide. T2D is primarily characterized by insulin resistance, and growing evidence has indicated the causative link between adipose tissue inflammation and the development of insulin resistance. Genetic association studies have successfully revealed a number of important genes consistently associated with T2D to date. However, these robust T2D-associated genes do not fully elucidate the mechanisms underlying the development and progression of the disease. Here, we report an alternative approach, gene expression-based genome-wide association study (eGWAS): searching for genes repeatedly implicated in functional microarray experiments (often publicly available). We performed an eGWAS across 130 independent experiments (totally 1,175 T2D case-control microarrays) to find additional genes implicated in the molecular pathogenesis of T2D and identified the immune-cell receptor *CD44* as our top candidate ($P = 8.5 \times 10^{-20}$). We found *CD44* deficiency in a diabetic mouse model ameliorates insulin resistance and adipose tissue inflammation and also found that anti-CD44 antibody treatment decreases blood glucose levels and adipose tissue macrophage accumulation in a high-fat, diet-fed mouse model. Further, in humans, we observed CD44 is expressed in inflammatory cells in obese adipose tissue and discovered serum CD44 levels were positively correlated with insulin resistance and glycemic control. *CD44* likely plays a causative role in the development of adipose tissue inflammation and insulin resistance in rodents and humans. Genes repeatedly implicated in publicly available experimental data may have unique functionally important roles in T2D and other complex diseases.

bioinformatics | meta-analysis | integration | obesity | hyperglycemia

Type 2 diabetes (T2D) is a common multifactorial disease characterized by hyperglycemia primarily resulting from peripheral insulin resistance, and growing functional evidence has indicated the causative link between adipose tissue inflammation and the development of insulin resistance (1, 2). In the past decade, a number of genetic genome-wide association studies (GWAS) have revealed 40 loci consistently associated with susceptibility to T2D and have rapidly expanded our knowledge of the genetic architecture of this disease (3–13). However, the genes located in or near these loci do not fully elucidate the tissue-specific molecular mechanisms underlying the development of T2D.

A large number of experiments using genome-wide gene-expression microarray measurements have been also performed over the past decade; however, there has been little success in fully identifying functionally important genes in the pathogenesis of T2D. Because a large number of genes are often detected as significant in each microarray experiment, it may be hard to subselect optimal candidates from individual studies for further verification. A combination of genome-wide data from two or more experiments has been performed for obesity (14–16), T2D (14), and other multifactorial disorders (17–20). Several of these methods have used microarray technology to focus on candidate genes already implicated in a region of a congenic or model animal (accelerated positional candidate identification). More recently, investigators have applied microarrays to genetics by considering gene expression levels as quantitative traits (expression quantitiative trait loci, eQTLs) and finding relations between gene variants and transcripts, with successful application to the identification of genes and targets for T2D (21, 22). The need for large numbers of simultaneously acquired genetic and gene-expression measurements within a single study makes this approach less scalable.

We suspected that the large number of molecular measurements from experimental results that are now publicly available, because of requirements from journals and funding agencies (23), could be used as an alternative scalable source of data. The strategy of finding commonly implicated genes across related—but deliberately varied—experimental conditions has been theorized to yield less overfit, potentially more generalizable causal factors (24), and publicly available data could be used as a source of these varied experimental vantage points for a condition. In this report, we propose the application of a gene expression-based genome-wide association study (eGWAS), a meta-analysis method for computing the likelihood of finding repeated differential expression for every gene across a large number of case and control microarray experiments, compared with expected. Our hypothesis is that those genes most repeatedly implicated across a large set of experimental representations of T2D can serve as data-driven causal T2D genes and candidates for validation. This approach is only feasible because many of these source raw experimental results are publicly available; here, we integrated 130 independent microarray experiments for T2D. In this case, our T2D candidates were found independent of any knowledge about insulin signaling, glucose, or lipid metabolism. For our top candidate gene, identification was followed by confirmatory functional studies using mouse models and samples from human subjects. (Fig. S1).

Author contributions: K.K., M.H., K.H., T.K., and A.J.B. designed research; K.K., M.H., K.T., S.Y., K.H., J.I., S.M., T.K., and A.J.B. performed research; M.S., A.A.M., and R.C. contributed new analytic tools; K.K., M.H., K.H., H.O., S.M., T.K., and A.J.B. analyzed data; and K.K., M. H., K.H., T.K., and A.J.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.D.A. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]K.T., S.Y., K.H., and J.I. contribute equally to this work.

[2]To whom correspondence should be addressed. E-mail: abutte@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1114513109/-/DCSupplemental.

Results

eGWAS Identifies *CD44* as a Functional Candidate Gene for T2D. We carried out an eGWAS for T2D by using 130 independent microarray experiments, totaling 1,175 samples collected from public repositories (Fig. 1, Tables S1 and S2, and *SI Materials and Methods*). We ranked all 24,898 genes by the likelihood that repeated differential expression for that gene was due to chance, then controlled for multiple-hypothesis testing. To overview which molecular functions are most shared in the highest-ranked genes in our T2D eGWAS, we took the top 127 genes (Table S3; Bonferroni threshold, $P < 2.0 \times 10^{-6}$) from our eGWAS and then estimated the enrichment of Gene Ontology (GO) terms. Interestingly, "receptor activity" and "receptor binding" functions were the most implicated of the top-ranked genes (receptor activity, 38%; receptor binding, 19%) (Fig. S2). These activities suggested that a number of top-ranked genes on our list are involved in intra- and intertissue signaling cascades in the development of T2D (1).

Our top-most T2D candidate gene was *CD44* (Fig. 1: χ^2 analysis, $P = 8.5 \times 10^{-20}$; Fig. S3: Fisher's exact test, $P = 6.1 \times 10^{-17}$; Fig. S4: weighted Z-method), markedly differentially expressed in experiments studying diabetes in adipose tissue compared with other tissues (Fig. S5). *CD44* is located on chromosome 11p13 and codes for a cell-surface glycoprotein, an immunological cell (macrophage/T-cell) receptor, involved in inflammatory cell migration and activation. Interestingly, one of the known ligands for *CD44*, secreted phosphoprotein 1 [SPP1; also known as osteopontin (OPN)], a Th1 cytokine secreted by immunological cells (macrophages), was also included in the top-ranked genes (Fig. 1: χ^2 analysis, $P = 1.3 \times 10^{-11}$; Fig. S3: Fisher's exact test, $P = 3.8 \times 10^{-10}$). Recent studies have indicated that obese adipose tissue is hallmarked with chronic, low-grade inflammation, and that inflammation plays a central role in the development of insulin resistance (1, 2). Although the contributions of the *CD44* encoded protein to the molecular pathogenesis of T2D have not yet been reported, SPP1 was reported as a link between adipose tissue inflammation (stromal infiltration by inflammatory cells) and the development of insulin resistance in a murine model of diet-induced obesity (25). Furthermore, the expression profile of *CD44* and *SPP1* are coordinately dysregulated, especially in adipose tissue (Fig. S6; coordinate dysregulation rate = 0.90). These findings suggest that CD44 might have a key role in mediating obesity-induced adipose tissue inflammation and insulin resistance.

***CD44* Expression Increases in Obese Adipose Tissue.** High-fat feeding in C57BL/6J mice leads to the development of obesity, adipose inflammation, and insulin resistance (25, 26). To examine whether the *CD44* mRNA transcript is expressed in adipose tissue and modulated by obesity, C57BL/6J mice were maintained either on a normal-fat diet (NFD; 12% of total calories from fat) or high-fat diet (HFD; 60% of total calories from fat) for 16 wk ($n = 8$ per group). Compared with the NFD group, mice fed a HFD gained 37% more weight after a 16-wk feeding period (29.9 ± 0.5 g versus 40.9 ± 1.6 g; $P = 5.8 \times 10^{-8}$). Epididymal white adipose tissue (EWAT) was removed from these mice to analyze *CD44* mRNA expression levels. Feeding a HFD resulted in a significant 11.3-fold increase of *CD44* mRNA levels in adipose tissue compared with NFD (Fig. 2*A*). To establish the presence of CD44 protein in adipose tissue, an immunohistochemical localization of CD44 was performed on EWAT isolated from HFD mice. We found that CD44 was abundantly expressed in inflammatory cells within adipose tissue (Fig. 2*B*). These results clearly indicate that $CD44^+$ cells accumulated in EWAT of diet-induced obese mice. In addition, we confirmed that the expression levels of *SPP1* mRNA in adipose tissue in the HFD group was significantly higher than that in the NFD mice (0.005 ± 0.003 versus 0.06 ± 0.02; $P < 0.05$), similar to previous reports (25). Interestingly, we also found that *CD44* mRNA expression level was positively correlated with the *SPP1* mRNA expression level in the HFD mice (Fig. 2*C*; $r = 0.78$, $P = 0.02$), suggesting that CD44 and SPP1 may be closely related in obese adipose tissue.

***CD44* Deficiency Ameliorates Adipose Tissue Inflammation and Insulin Resistance.** To next determine the contribution of CD44 to the development of adipose tissue inflammation and insulin resistance, we fed male $CD44^{-/-}$ and diabetes-prone C57BL/6J ($CD44^{+/+}$) mice with either a HFD ($n = 16$ per group) or a NFD ($n = 10$ per group) for 12 wk and performed immunohistochemical analysis and metabolic measurements on these mice. There were no significant differences in body weights between $CD44^{-/-}$ and $CD44^{+/+}$ mice after feeding a NFD or a HFD (NFD: $CD44^{-/-}$ 28.5 ± 0.5 g versus $CD44^{+/+}$ 29.5 ± 0.5 g; HFD: 38.1 ± 1.5 g versus 39.5 ± 1.2 g). In $CD44^{+/+}$ mice that were fed a HFD, we frequently observed the accumulation of inflammatory cells (macrophages) forming crown-like structures (CLSs) surrounding adipocytes in obese visceral adipose tissue. However, $CD44^{-/-}$ mice fed a HFD exhibited strikingly less macrophage infiltration into the stroma of adipose tissue compared with $CD44^{+/+}$ mice fed a HFD (Fig. 3*A*). Fasting blood glucose levels were significantly lower in $CD44^{-/-}$ mice fed a HFD compared with the diabetes-prone $CD44^{+/+}$ mice fed a HFD (Fig. 3*B*). Glucose tolerance tests also indicated that $CD44^{-/-}$ mice fed a HFD were significantly more efficient in their ability to clear intraperitoneally injected glucose than $CD44^{+/+}$ mice fed a HFD (Fig. 3*C*, solid lines) despite similar insulin secretory responses after the injection of glucose. Furthermore, insulin sensitivity, as measured by insulin tolerance test, showed that $CD44^{-/-}$ mice fed a HFD were significantly more efficient at insulin-mediated suppression of blood glucose than $CD44^{+/+}$ mice fed a HFD (Fig. 3*D*, solid lines).

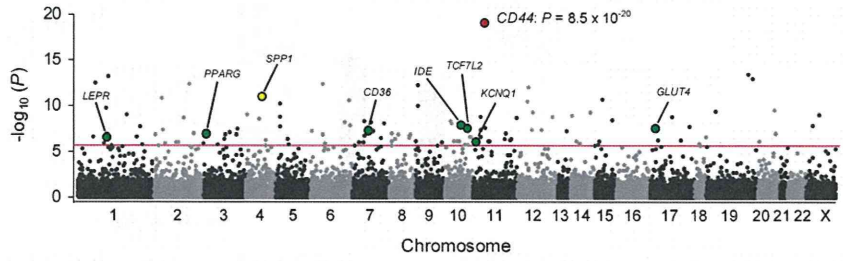

Fig. 1. eGWAS for T2D using a χ^2 analysis. Plot of -$\log_{10}$ (*P* value) (*y* axis) by chromosomal position (*x* axis). *P* values for each gene were calculated from our eGWAS across 130 microarray experiments with 1,175 T2D case-control microarray samples (591 T2D cases and 584 controls) as the likelihood of finding repeated differential expression compared with expected using a χ^2 analysis, or a Fisher's exact test (Fig. S3). Our top gene, *CD44*, showed a significant differential expression in 78 experiments ($P = 8.5 \times 10^{-20}$). The red line indicates the Bonferroni threshold ($P = 2.0 \times 10^{-6}$). The green dots indicate several well known T2D-susceptibility genes.

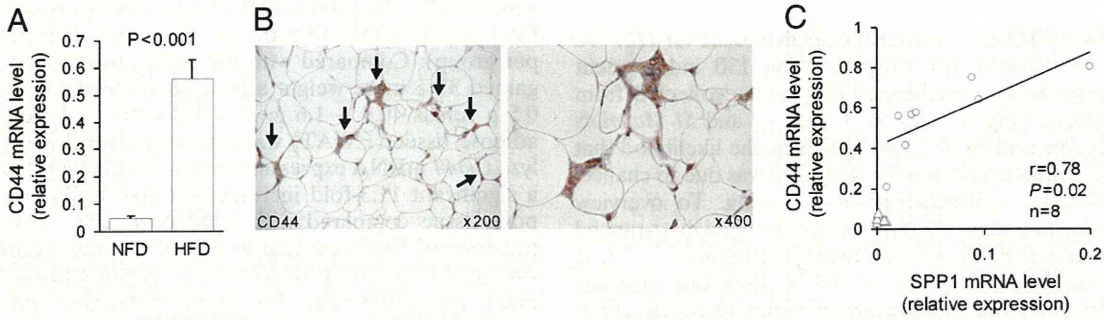

Fig. 2. CD44 expression in adipose tissue of obese mice. (*A*) *CD44* mRNA expression levels in epididymal adipose tissues in C57BL/6J mice fed either a NFD or HFD (n = 8 per group). (*B*) CD44 immunoreactivity (arrows, DAB chromogen; brown) in epididymal adipose tissues from C57BL/6J mice fed a HFD. Sections were counterstained with hematoxylin (blue). (*C*) *CD44* and *SPP1* gene expression profiles in the HFD (n = 8; circles) and NFD (n = 8; triangles) groups. Correlation between *CD44* and *SPP1* mRNA expression in the HFD group (circles) was analyzed by using the Pearson's correlation test. Gene expression was monitored by using real-time RT-PCR and normalized to expression of *GAPDH* mRNA.

Interestingly, even in the mice fed a NFD, the ameliorative effects of *CD44* deficiency on insulin sensitivity was observed in both glucose and insulin tolerance tests (Fig. 3 *C* and *D*, dashed lines). These results that we obtained with CD44-deficient mice confirm that CD44 molecules are essential for macrophage recruitment and inflammation in adipose tissue and the development of insulin resistance in diet-induced obese mice.

CD44 Blockade Decreases Blood Glucose Levels and Adipose Macrophage Infiltration. Several in vivo studies have shown that anti-CD44 monoclonal antibody (CD44 mAb) treatment exhibits robust antiinflammatory effects in animal models of immune-mediated diseases (27–30). We therefore sought to investigate whether CD44 blockade might demonstrate a therapeutic effect on T2D. We performed daily i.p. injections of CD44 mAb in diabetic model mice for 1 wk and found that blood glucose levels and adipose macrophage accumulation were significantly reduced in CD44 mAb treated mice compared with isotype-control treated mice, despite continuing on the HFD and similar body weight increase during the treatment (Fig. 4 *A* and *B*). When adipose inflammation was quantified as the average number of CLSs per low power field, EWAT from CD44 mAb treated mice contained significantly fewer inflammatory cells compared with control treated mice (2.4 ± 0.5 versus 5.4 ± 0.7; P = 0.0005). Collectively, these effects of CD44 mAb clearly show that CD44 molecules are required for the recruitment of macrophages into obese adipose tissue and the maintenance of inflammatory reactions there, and the CD44 receptor may be useful as a therapeutic target for T2D.

To gain additional insight into the clinical importance of CD44 in obese fat, we performed an immunohistochemical analysis of CD44 in omental adipose tissue in human obesity. Consistent with our mouse model observation, we discovered that $CD44^+$ cells infiltrated into the stroma of adipose tissue in human obese subjects, suggesting that CD44 molecules may mediate macrophage migration into obese adipose tissue in humans (Fig. 5*A*).

Soluble CD44 shed from cell surfaces exists in normal human serum. To estimate the relevance between CD44 protein and glucose homeostasis in human subjects, we evaluated the relationship between serum levels of CD44 and metabolic traits in human and found that serum CD44 was positively correlated with glycemic control and insulin resistance as estimated through HbA1c (n = 55, r = 0.49, P < 0.001) and HOMA-IR (n = 55, r = 0.29, P = 0.03) (Fig. 5 *B* and *C*). We then classified the 55 subjects into two groups according to the WHO criteria (31): hyperglycemia (n = 21: "diabetes mellitus" + "impaired glucose regulation") and normoglycemia (n = 34: "Normal Glucose Tolerance"), and found that the serum levels of CD44 were significantly higher in the hyperglycemic group than the normoglycemic group (246.9 ± 15.0 versus 209.5 ± 9.8 ng/mL; P = 0.02). These results suggest that CD44 protein may be released from insulin-resistant and diabetic tissues into circulation in humans.

Discussion

Adipose tissue inflammation is thought to be a pivotal event leading to the metabolic syndrome, insulin resistance, and T2D.

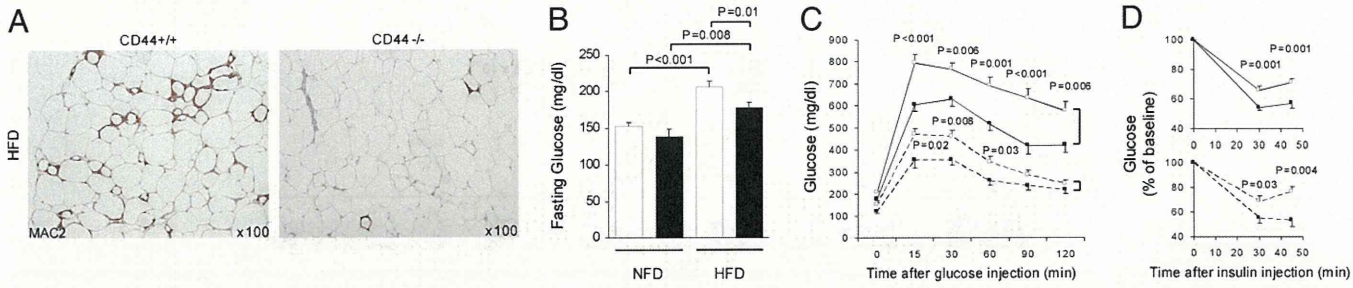

Fig. 3. Histological and metabolic analyses of wild-type $CD44^{+/+}$ and $CD44^{-/-}$ mice. (*A*) Inflammatory cell (macrophage) content determined by immunohistochemical staining for Mac-2 (DAB, brown; hematoxylin, blue) in epididymal adipose tissues from $CD44^{-/-}$ and $CD44^{+/+}$ mice fed a HFD. (*B–D*) Metabolic measurements on $CD44^{+/+}$ (open bars and symbols; diabetes-prone) and $CD44^{-/-}$ (filled bars and symbols) mice fed either a HFD (n = 16 per group; solid lines) or a NFD (n = 10 per group; dashed lines). (*B*) Fasting blood glucose. (*C*) Glucose tolerance tests [i.p. glucose (2 g/kg body weight)] after a 14-h overnight fast. Venous blood was obtained for measurement of blood glucose at 0, 15, 30, 60, 90, and 120 min after the injection. (*D*) Insulin tolerance tests [i.p. insulin (1.0 unit/kg body weight)] after a 4-h fast. Venous blood was obtained for measurement of blood glucose at 0, 30, and 45 min after the injection.

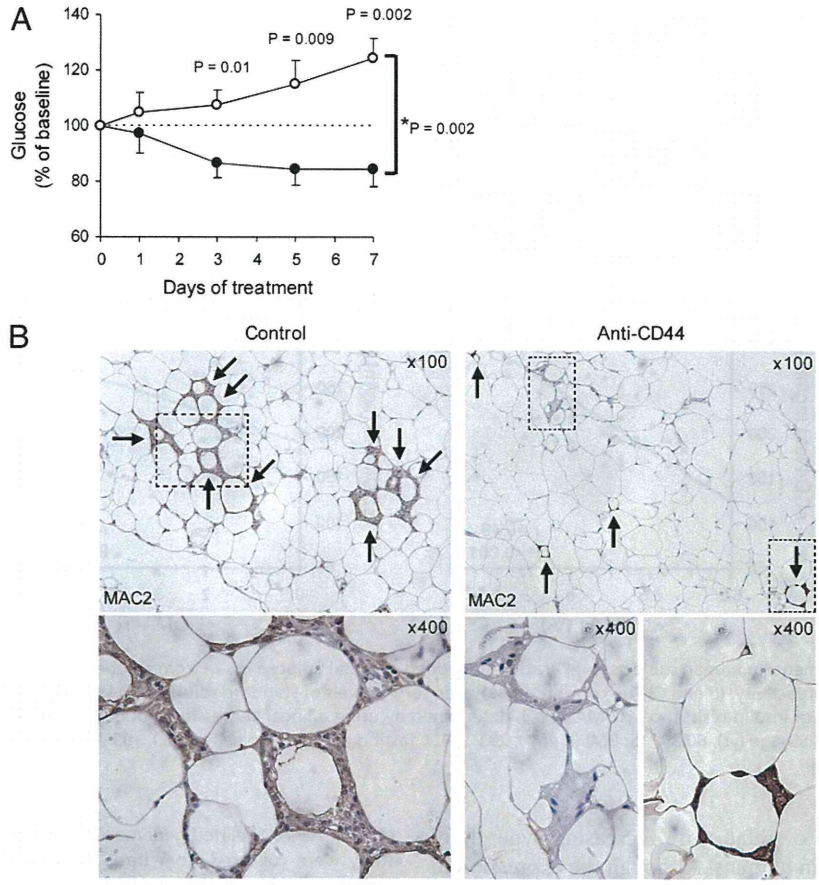

Fig. 4. Anti-CD44 antibody treatment for diabetic mice. (*A*) HFD-fed C57BL/6J mice were injected intraperitoneally with purified rat anti-mouse CD44 (IM7; 553131, BD Pharmingen) (n = 6; filled circles) or purified rat IgG2b, κ isotype control (A95-1; 559478, BD Pharmingen) (n = 8; open circles) for 1 wk (100 μg at day 0 and 50 μg at day 1–7). Morning blood glucose was measured at day 0, 1, 3, 5, and 7 during the treatment. The effect of anti-CD44 treatment on blood glucose levels was evaluated with two-way repeated measures ANOVA (*P; treatment × time). Comparisons between two groups were performed by using the two-tailed Welch's t test. Data are represented as mean ± SE. (*B*) Epididymal adipose tissues from control and anti-CD44 antibody-treated mice were analyzed for inflammatory cell (macrophage) content by using a Mac-2 antibody (magnified as indicated). Arrows indicate crown-like structures (CLSs) surrounding individual adipocytes.

Genome-wide experimental methods to identify disease genes, such as association studies (GWAS), linkage studies (GWL), and eQTL analyses, have been performed for T2D by many researchers to date, and these methods have revealed a number of loci to be linked with T2D (3–14, 21, 32–35). However, these genetically mapped loci do not fully account for the tissue-specific mechanisms underlying the development of T2D. In this report, we proposed an alternative methodology, eGWAS, computing the likelihood of finding repeated differential expression of a gene in disease-related tissues by using thousands of case-control microarray samples. To detect additional genes functionally implicated in the molecular pathogenesis of T2D, we successfully performed an eGWAS to identify T2D candidate genes and verified our top candidate gene, *CD44*, an immunological cell receptor, plays a significant role in the development of adipose tissue inflammation and insulin resistance in mouse models and human subjects.

In our T2D eGWAS, we identified in total 127 genes as significantly repeatedly dysregulated with P values $<2.0 \times 10^{-6}$ (under the Bonferroni-corrected threshold) and rediscovered several genes that have been shown to be important in T2D pathogenesis, including *TCF7L2*, *PPARG*, *KCNQ1*, *IDE*, *CD36*, *GLUT4*, and *LEPR*, supporting the validity of our methodology. Of the 127 genes, we found that more than one-half were implicated in the "receptor" or "ligand" activity by using the GO term enrichment analysis (Fig. S2). Interestingly, *SPP1*, encoding a ligand for the CD44 receptor, was shown to be included also in our top-ranked gene list (Fig. 1 and Table S3). Furthermore, our eGWAS provided the prediction of tissue specificity of gene expression by calculating a distribution of scores for each gene across tissues, indicating that *CD44* mRNA expression was more highly up-regulated in adipose tissue in diabetes than other tissues (Fig. S5). These analysis results led us to the speculation that our top-most candidate gene, *CD44*, encoding an immune-cell surface receptor, may be implicated in adipose tissue inflammation causing insulin resistance in obesity, given the fact that the CD44 receptor is known to regulate immune-cell migration and activation and its ligand SPP1 has been reported to mediate macrophage infiltration into obese adipose tissue (25). When interpreting this hypothesis, however, we need to consider the possibility that the dysregulation of *CD44* expression in relevant tissues can result secondarily from hyperglycemia and diabetes. We therefore performed verification experiments to see whether the dysregulated expression of *CD44* gene in obese adipose tissue can be accepted as the cause of insulin resistance.

In our functional tests for the *CD44* gene products, we showed that knocking out the receptor *CD44* leads to striking reduction in immune-cell infiltration into visceral adipose tissue and improvements in insulin sensitivity in mouse models, and anti-CD44 monoclonal antibody treatment decrease the blood glucose levels and visceral adipose tissue macrophages in diabetic obese mice. In addition, we showed that higher serum levels of

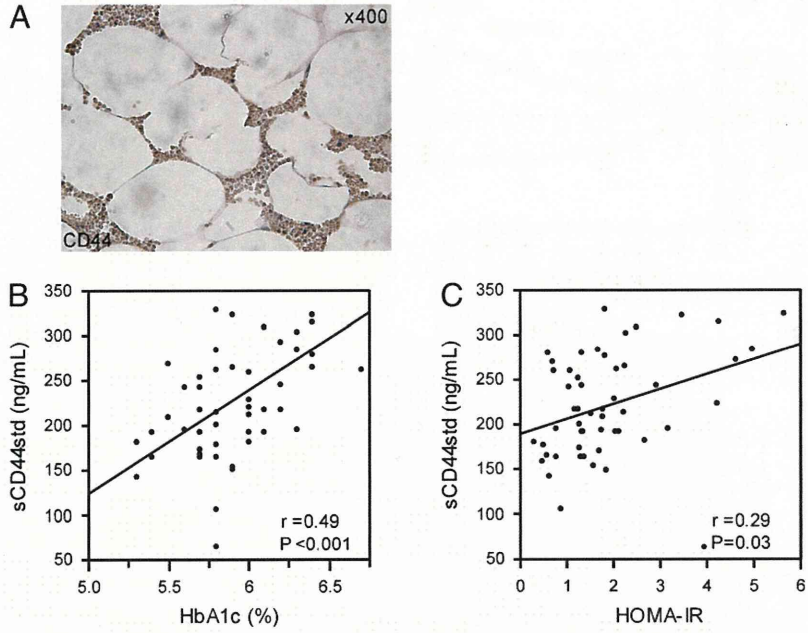

Fig. 5. CD44 functional experiments in human subjects. (*A*) Paraffin-embedded omental adipose tissue from an obese woman [age (yr); 57, BMI (kg/m^2); 36.9] was analyzed for CD44 immunoreactivity. (*B* and *C*) The correlation between serum levels of standard soluble CD44 (sCD44std) and either an index of glycemic control, HbA1c (*B*), or an index of insulin resistance, HOMA-IR (*C*), determined by using a linear regression model estimated with minimal square method in human subjects [n = 55: mean ± SD age (yr), 60.3 ± 15; BMI (kg/m^2), 23.2 ± 4.3; fasting plasma glucose (mg/dL), 109 ± 13; fasting plasma insulin (μU/mL), 6.22 ± 3.84; HbA1c (%), 5.9 ± 0.34].

the soluble form of CD44 correlate with increasing hyperglycemia and insulin resistance in humans. Although the expression of CD44 in macrophages and T cells in obese adipose tissue has been reported (36, 37), this study directly addresses the functional role of CD44 in adipose tissue inflammation. The findings that systemic glucose intolerance is ameliorated by CD44 depletion and blockade by using CD44 mAb strongly suggest that CD44-dependent adipose inflammation has an impact on systemic metabolism. However, further studies are needed to determine which immune-cell population primarily expresses CD44 receptor (macrophages or T cells?) and how the CD44 molecules initiate and maintain immune-cell infiltration into adipose tissue (through the SPP1 signals?). Even so, these results indicate the significance of CD44 immune-receptor as a possible therapeutic target for T2D and a unique biomarker for insulin resistance.

In conclusion, we discovered that an immune-cell receptor gene, *CD44*, is pathogenetically implicated in the development of adipose tissue inflammation and insulin resistance, using a data-driven candidate gene approach by using an eGWAS method of integrating >1,000 publicly available genome-wide functional microarrays related to T2D. Application of the eGWAS methodology to publicly available data can yield promising candidate genes that are differentially expressed in T2D-relevant tissues, independently of knowledge about insulin signaling, glucose, or lipid metabolism. We suggest that a data-driven approach can enable investigators to consider glucose homeostasis phenotypes from a different point of view and notice new pathways that could be involved in the development of T2D. Although GWAS and other genetic analyses will continue as the method of choice for the next few years, an eGWAS approach could complement these studies to yield additional pathogenetically important genes for many other complex diseases by using this integrated data-driven approach.

Materials and Methods

See *SI Materials and Methods* for further descriptions.

eGWAS. All T2D-related genome-wide microarray experiments used for this study were collected from three public data sources: the NCBI Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo), the Diabetes Genome Anatomy Project (DGAP; www.diabetesgenome.org), and the Nuclear Receptor Signaling Atlas (NURSA; www.nursa.org). There were a total of 1,175 samples (591 T2D cases and 584 controls) in 130 independent datasets. To estimate differences between groups of samples from diabetic subjects and groups representing control, raw postquantitation microarray data were reanalyzed by using Significance Analysis of Microarrays software (SAM) (38). For each gene in every microarray experiment with three or more samples in each group, we calculated a d score (d_i), which denotes the standardized change in gene expression:

$$d_i = \frac{\bar{x}_{i-t2d} - \bar{x}_{i-control}}{S_i + S_0},$$

where $\bar{x}_{i-t2d}$ is the mean expression level of gene i in group T2D, $\bar{x}_{i-control}$ is the mean expression level of gene i in group control, S_i is the SD for the numerator calculation, and S_0 is a small positive constant. We considered genes to be significantly dysregulated with either an absolute value of the d score ≥2 or a fold change ≥2 between controls and cases. We then converted all probe identifiers across the various microarray platforms for mouse, rat, and human to the latest human Entrez Gene identifiers by using our published AILUN system (39). Gene expression profiles were assigned in our eGWAS database according to the standardized (human) Entrez Gene ID. There were 24,898 genes in the database in total. For every one of the 24,898 genes, we counted the observed number of microarray experiments in which each gene was significantly dysregulated. We then calculated P values from the number of positive/negative experiments for every one of genes and sum of the number of positive/negative experiments for all other genes, using a 2 × 2 χ^2 analysis (Fig. 1) or a Fisher's exact test as an alternative (Fig. S3), and ranked all of the genes according to their P values [$-\log_{10}(P)$]. A third method, Liptak–Stouffer's weighted Z-method (40), provided additional support for *CD44* (Fig. S4).

Animal Experiments. Mice for breeding, C57BL/6J wild-type (diabetes-prone) and *CD44*-deficient mice backcrossed to C57BL/6J for at least 10 generations (B6.Cg-*Cd44*^{tm1Hbg}/J), were obtained from The Jackson Laboratory. Mice were housed in a barrier facility under specific pathogen-free conditions. The Animal Care and Use Committee of Kitasato University approved all animal experiments.

MEDICAL SCIENCES

Human Studies. Venous peripheral blood samples were collected from human subjects who went through a 75 g oral glucose tolerance test after an overnight fast [n = 55: sex (M/F); 36/19]. HbA1c was measured in Japan Diabetes Society (JDS)-HbA1c units and then converted to National Glycohemoglobin Standardization Program (NGSP) levels by the formula HbA1c (%) (NGSP) = HbA1c (JDS) (%) + 0.4% (41). We then calculated homeostasis model assessment as an index of insulin resistance [HOMA-IR = fasting plasma insulin (μU/mL) × fasting plasma glucose (mg/dL)/405] as described (42). Serum sCD44std (standard soluble CD44) and SPP1 concentrations were determined by using a quantitative ELISA technique (sCD44std ELISA, Bender MedSystems; Human Osteopontin Quantikine ELISA, R&D Systems).

Informed consent was obtained from all of the subjects enrolled in this study, and the protocol was approved by the ethics committee of the University of Tokyo.

Immunohistochemistry. For histological analysis of CD44 expression in adipose tissue, EWAT was removed from mouse models, and omental adipose tissue obtained from consented donors undergoing elective gastric bypass surgery (lot no. OM020304B) was purchased from Zen-Bio. Formalin-fixed paraffin-embedded sections were stained with mouse monoclonal antibody against CD44 at 1:50 dilution (DF1485/sc-7297; Santa Cruz Biotechnology), followed by reactions with anti-mouse immunoglobulins-HRP, and anti-fluorescein-HRP. 3,3′-Diaminobenzidine (DAB) was used as a chromogen.

Analysis of inflammatory cell (macrophage) content in EWAT was performed on tissue pads isolated from model mice. Formalin-fixed paraffin-embedded sections were incubated overnight with primary antibody: Purified Anti Mouse MAC-2 Monoclonal Antibody (CL8942AP, 1:100; Cedarlane Laboratories) and stained by using Histofine Simple Stain Mouse MAX-PO secondary antibody (Nichirei Biosciences) with a DAB solution.

Statistics. For verification studies in mouse models and human subjects, comparisons between two groups were performed by using the two-tailed Welch's t test. Two-way repeated measures ANOVA was used to examine the treatment (antibodies type) × time interaction on blood glucose changes from baseline. P values of <0.05 were considered significant. All experimental data are represented as mean ± SE unless otherwise noted.

ACKNOWLEDGMENTS. We thank Saori Ohta of Kitasato University School of Pharmacy for her support in animal experiments; Dr. Shojiro Morinaga of Department of Pathology, Kitasato Research Institute Hospital, for his assistance in histological analysis; and Dr. Damon Tojjar (Department of Clinical Sciences, Lund University, Scania University Hospital) and Dr. Purvesh Khatri (Division of Systems Medicine, Stanford University School of Medicine) for their suggestions in preparing the manuscript. This work was supported in part by the Howard Hughes Medical Institute, National Library of Medicine Grants (R01 LM009719 and K22 LM008261), and by the Lucile Packard Foundation for Children's Health.

1. Hotamisligil GS (2006) Inflammation and metabolic disorders. *Nature* 444:860–867.
2. Shoelson SE, Lee J, Goldfine AB (2006) Inflammation and insulin resistance. *J Clin Invest* 116:1793–1801.
3. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661–678.
4. Scott LJ, et al. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345.
5. Saxena R, et al.; Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336.
6. Sladek R, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885.
7. Zeggini E, et al.; Wellcome Trust Case Control Consortium (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645.
8. Unoki H, et al. (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 40:1098–1102.
9. Yasuda K, et al. (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 40:1092–1097.
10. Dupuis J, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42:105–116.
11. Saxena R, et al.; GIANT consortium; MAGIC investigators (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 42:142–148.
12. Voight BF, et al.; MAGIC investigators; GIANT Consortium (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42:579–589.
13. Yamauchi T, et al. (2010) A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 42:864–868.
14. Aitman TJ, et al. (1999) Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nat Genet* 21:76–83.
15. Schadt EE, et al. (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 37:710–717.
16. Chen Y, et al. (2008) Variations in DNA elucidate molecular networks that cause disease. *Nature* 452:429–435.
17. Karp CL, et al. (2000) Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat Immunol* 1:221–226.
18. Klein RF, et al. (2004) Regulation of bone mass in mice by the lipoxygenase gene Alox15. *Science* 303:229–232.
19. Yagil C, et al. (2005) Identification of hypertension-related genes through an integrated genomic-transcriptomic approach. *Circ Res* 96:617–625.
20. Hsu YH, et al. (2010) An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. *PLoS Genet* 6:e1000977.
21. Zhong H, et al. (2010) Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* 6:e1000932.
22. Derry JM, et al. (2010) Identification of genes and networks driving cardiovascular and metabolic phenotypes in a mouse F2 intercross. *PLoS ONE* 5:e14319.
23. Perou CM (2001) Show me the data!. *Nat Genet* 29:373.
24. Paylor R (2009) Questioning standardization in science. *Nat Methods* 6:253–254.
25. Nomiyama T, et al. (2007) Osteopontin mediates obesity-induced adipose tissue macrophage infiltration and insulin resistance in mice. *J Clin Invest* 117:2877–2888.
26. Surwit RS, Kuhn CM, Cochrane C, McCubbin JA, Feinglos MN (1988) Diet-induced type II diabetes in C57BL/6J mice. *Diabetes* 37:1163–1167.
27. Hutás G, et al. (2008) CD44-specific antibody treatment and CD44 deficiency exert distinct effects on leukocyte recruitment in experimental arthritis. *Blood* 112: 4999–5006.
28. Mikecz K, Brennan FR, Kim JH, Glant TT (1995) Anti-CD44 treatment abrogates tissue oedema and leukocyte infiltration in murine arthritis. *Nat Med* 1:558–563.
29. Katoh S, et al. (2003) A role for CD44 in an antigen-induced murine model of pulmonary eosinophilia. *J Clin Invest* 111:1563–1570.
30. Brocke S, Piercy C, Steinman L, Weissman IL, Veromaa T (1999) Antibodies to CD44 and integrin alpha4, but not L-selectin, prevent central nervous system inflammation and experimental encephalomyelitis by blocking secondary leukocyte recruitment. *Proc Natl Acad Sci USA* 96:6896–6901.
31. Alberti KG, Zimmet PZ (1998) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 15:539–553.
32. Hanis CL, et al. (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat Genet* 13: 161–166.
33. Mahtani MM, et al. (1996) Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat Genet* 14:90–94.
34. Ghosh S, et al. (1999) Type 2 diabetes: Evidence for linkage on chromosome 20 in 716 Finnish affected sib pairs. *Proc Natl Acad Sci USA* 96:2198–2203.
35. Watanabe RM, et al. (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. *Am J Hum Genet* 67:1186–1200.
36. Nishimura S, et al. (2009) CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nat Med* 15:914–920.
37. Zeyda M, et al. (2011) Osteopontin is an activator of human adipose tissue macrophages and directly affects adipocyte function. *Endocrinology* 152:2219–2227.
38. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98:5116–5121.
39. Chen R, Li L, Butte AJ (2007) AILUN: Reannotating gene expression data automatically. *Nat Methods* 4:879.
40. Whitlock MC (2005) Combining probability from independent tests: The weighted Z-method is superior to Fisher's approach. *J Evol Biol* 18:1368–1373.
41. Seino Y, et al. (2010) Report of the Committee on the Classification and Diagnostic Criteria of Diabetes Mellitus. *J Diabetes Invest* 1:212–228.
42. Matthews DR, et al. (1985) Homeostasis model assessment: Insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28:412–419.

OPEN ACCESS Freely available online PLoS GENETICS

Global Mapping of Cell Type–Specific Open Chromatin by FAIRE-seq Reveals the Regulatory Role of the NFI Family in Adipocyte Differentiation

Hironori Waki[1,2☯], Masahiro Nakamura[1☯], Toshimasa Yamauchi[1☯*], Ken-ichi Wakabayashi[3☯], Jing Yu[1], Lisa Hirose-Yotsuya[1], Kazumi Take[1], Wei Sun[1], Masato Iwabu[1,4], Miki Okada-Iwabu[1,5], Takanori Fujita[3], Tomohisa Aoyama[1], Shuichi Tsutsumi[3], Kohjiro Ueki[1], Tatsuhiko Kodama[6], Juro Sakai[7*], Hiroyuki Aburatani[3*], Takashi Kadowaki[1*]

1 Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Tokyo, Japan, **2** Functional Regulation of Adipocytes, Graduate School of Medicine, University of Tokyo, Tokyo, Japan, **3** Genome Science Division, Laboratory of Systems Biology and Medicine, Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan, **4** Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, University of Tokyo, Tokyo, Japan, **5** Molecular Medicinal Sciences on Metabolic Regulation, Graduate School of Medicine, University of Tokyo, Tokyo, Japan, **6** Systems Biology and Medicine Division, Laboratory of Systems Biology and Medicine, Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan, **7** Metabolic Medicine Division, Laboratory of Systems Biology and Medicine, Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan

Abstract

Identification of regulatory elements within the genome is crucial for understanding the mechanisms that govern cell type–specific gene expression. We generated genome-wide maps of open chromatin sites in 3T3-L1 adipocytes (on day 0 and day 8 of differentiation) and NIH-3T3 fibroblasts using formaldehyde-assisted isolation of regulatory elements coupled with high-throughput sequencing (FAIRE-seq). FAIRE peaks at the promoter were associated with active transcription and histone modifications of H3K4me3 and H3K27ac. Non-promoter FAIRE peaks were characterized by H3K4me1+/me3-, the signature of enhancers, and were largely located in distal regions. The non-promoter FAIRE peaks showed dynamic change during differentiation, while the promoter FAIRE peaks were relatively constant. Functionally, the adipocyte- and preadipocyte-specific non-promoter FAIRE peaks were, respectively, associated with genes up-regulated and down-regulated by differentiation. Genes highly up-regulated during differentiation were associated with multiple clustered adipocyte-specific FAIRE peaks. Among the adipocyte-specific FAIRE peaks, 45.3% and 11.7% overlapped binding sites for, respectively, PPARγ and C/EBPα, the master regulators of adipocyte differentiation. Computational motif analyses of the adipocyte-specific FAIRE peaks revealed enrichment of a binding motif for nuclear family I (NFI) transcription factors. Indeed, ChIP assay showed that NFI occupy the adipocyte-specific FAIRE peaks and/or the PPARγ binding sites near PPARγ, C/EBPα, and aP2 genes. Overexpression of NFIA in 3T3-L1 cells resulted in robust induction of these genes and lipid droplet formation without differentiation stimulus. Overexpression of dominant-negative NFIA or siRNA-mediated knockdown of NFIA or NFIB significantly suppressed both induction of genes and lipid accumulation during differentiation, suggesting a physiological function of these factors in the adipogenic program. Together, our study demonstrates the utility of FAIRE-seq in providing a global view of cell type–specific regulatory elements in the genome and in identifying transcriptional regulators of adipocyte differentiation.

Citation: Waki H, Nakamura M, Yamauchi T, Wakabayashi K-i, Yu J, et al. (2011) Global Mapping of Cell Type–Specific Open Chromatin by FAIRE-seq Reveals the Regulatory Role of the NFI Family in Adipocyte Differentiation. PLoS Genet 7(10): e1002311. doi:10.1371/journal.pgen.1002311

Editor: Jason D. Lieb, The University of North Carolina at Chapel Hill, United States of America

Received April 17, 2011; **Accepted** August 9, 2011; **Published** October 20, 2011

Funding: This work was supported by a Grant-in-Aid for Young Scientists (B) from the Japan Society for the Promotion of Science (JSPS) (#20890055, #21790864, H Waki, http://www.jsps.go.jp/english/), funds from the Kanae Foundation for the Promotion of Medical Science (H Waki, http://kanae.sanofi-aventis.co.jp/en/index.html), Senri Life Science Foundation (H Waki, http://www.senri-life.or.jp/), Takeda Science Foundation (H Waki, http://www.takeda-sci.or.jp/), Japan Foundation for Applied Enzymology (H Waki, http://www.mt-pharma.co.jp/jfae/index.html), Sankyo Foundation of Life Science (H Waki, http://www.sankyo-fdn.or.jp/), Banyu Life Science Foundation International (H Waki, http://www.banyu-zaidan.or.jp/index.html), and by a Grants-in-Aid for Scientific Research (S) from the Ministry of Education, Culture, Sports, Science, and Technology (#20221009, H Aburatani, http://www.jsps.go.jp/english/), and Grant-in-Aid for Scientific Research on Innovative Areas (Research in a proposed research area) "Molecular Basis and Disorders of Control of Appetite and Fat Accumulation (to T Yamauchi), Funding Program for Next Generation World-Leading Researchers (NEXT Program) (to T Yamauchi). These institutions had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kadowaki-3im@h.u-tokyo.ac.jp (T Kadowaki); haburata-tky@umin.ac.jp (H Aburatani); jmsakai-tky@umin.ac.jp (J Sakai); tyamau-tky@umin.net (T Yamauchi)

☯ These authors contributed equally to this work.

Introduction

Sequencing allowed identification and mapping of the human genome [1]. Transcriptional regulation of genes is essential for manifesting cellular phenotypes and complex biological processes. Coordinated actions of transcription factors and cofactors on regulatory DNA sequences produce transcriptional activation of the eukaryotic gene. Therefore, identification and mapping of the genome's regulatory elements is critical for understanding how cell-type-selective regulation of genes in the genome is achieved.

Author Summary

Humans consist of a few hundred types of specialized-function cells. Spatial and temporal transcriptional regulation of genes is essential for manifestation of cellular phenotypes. Identification of regulatory regions in the genome is central to understanding the mechanism of cell type–specific gene regulation. Recently developed high-throughput sequencing technology and computational analyses allow genome-wide investigation of the genome's chromatin structure. Using the FAIRE-seq technique, we identified the genome's open chromatin regions, which harbor regulatory elements in adipocytes. Open chromatin regions distal to genes' transcription start sites significantly differ among cell types. Multiple cell type–specific open chromatin regions exist near genes regulated during adipocyte differentiation. Computational motif analysis of adipocyte-specific open chromatin regions revealed enrichment of a binding motif for the NFI transcription factor family. These factors bind to the regulatory elements near adipogenic PPARγ, C/EBPα, and aP2 genes and regulate their expression. Overexpression of NFIA in 3T3-L1 cells resulted in robust induction of these genes and lipid droplet formation without differentiation stimulus and knockdown of NFIA or NFIB significantly suppressed both induction of genes and lipid accumulation during differentiation. Our study demonstrates the utility of FAIRE-seq in providing a global view of regulatory elements and in identifying transcriptional regulators of cellular functions.

Traditionally, regulatory elements have been identified by DNase I hypersensitivity assay combined with Southern blot analysis [2]. That assay coupled with microarray or high-throughput sequencing (DNase-Chip or DNase-seq) were effectively applied in genome-wide identification of open chromatin regions [3,4,5,6]. Lieb and his colleagues recently developed formaldehyde-assisted isolation of regulatory elements (FAIRE) as a simple procedure to isolate nucleosome-depleted DNA from chromatin [7,8]. FAIRE detects open chromatin structure much the way the DNase I hypersensitivity assay does [8,9]—but with advantages, like obviating the need for clean nuclei preparation and laborious enzyme titrations [7,8]. Coupled with high-throughput sequencing (FAIRE-seq), FAIRE allows unbiased identification of potential regulatory elements without requiring prior knowledge of (or about) binding factors. FAIRE-seq's genome-wide detection of open chromatin genomic regions in human pancreatic islets was successfully used to determine a causal single nucleotide polymorphism in loci associated with type 2 diabetes development in genome-wide association studies [10].

The adipocyte is central in controlling energy balance and whole-body glucose and lipid homeostasis [11]. Advances in adipocyte research have shown that adipose tissue stores excess energy *and* secretes hormones and metabolites to communicate with other organs, maintaining systemic metabolic homeostasis [12]. Peroxisome proliferator-activated receptor gamma (PPARγ; NR1C3) is both necessary [13,14,15] and sufficient [16] for adipocyte differentiation. Necessary for both development and maintenance of mature adipocytes, PPARγ is crucial in systemic glucose and lipid homeostasis [13,14,15,17], and, importantly, is the molecular target of thiazolidinediones, widely prescribed for obese diabetics [18]. C/EBPα-β-δ act with PPARγ, forming the adipogenic transcription cascade [19]. C/EBPβ and δ are induced by adipogenic stimulus, inducing PPARγ, which activates expression of C/EBPα, which binds and further activates expression of PPARγ, providing a positive regulatory loop [11,20]. Genome-wide approaches now dissect the transcriptional mechanisms of adipocyte differentiation. ChIP-chip or ChIP-seq studies of adipogenic regulators [21,22,23,24,25,26,27,28,29] have provided valuable mechanistic insights into adipogenic transcription never before gained by conventional experiments: New concepts include co-localization of PPARγ and cell type–specific transcription factors [27], low conservation rate of PPARγ binding sites between murine and human adipocytes [28] and the role of C/EBPβ as a pioneer factor that establishes "hot spots" where multiple adipogenic regulators cooperatively work in the very early stage of differentiation [6].

Our study took an unbiased approach to mapping adipocyte-specific regulatory elements in the genome by using FAIRE in 3T3-L1 adipocytes (on day 0 and day 8 of differentiation) and NIH-3T3 fibroblasts. We show that the FAIRE peaks contain regulatory elements such as promoters, enhancers and insulators, and that adipocyte-specific non-promoter FAIRE peaks are functionally linked to genes regulated during differentiation—about half these peaks being overlapped by PPARγ. We show that highly regulated genes in adipocyte differentiation are associated with clusters of multiple adipocyte-specific non-promoter FAIRE peaks. Furthermore, because FAIRE does not require a prioi knowledge of bound transcription factors, we could employ computational motif analyses of DNA sequences from the adipocyte-specific FAIRE peaks in an unbiased manner and identify a motif for nuclear family I (NFI) transcription factors in addition to motifs for PPAR and C/EBPs. We show the functional role of NFIA and NFIB in adipocyte differentiation. We demonstrate the utility of FAIRE-seq both in providing a global view of cell type–specific *cis*-regulatory elements in the genome and identifying transcriptional regulators of adipocyte differentiation.

Results

Genome-Wide Profiling of Open Chromatin Regions in 3T3-L1 Adipocytes by FAIRE-seq

Regulatory elements in the genome are characterized by open chromatin structures accessible to regulatory factors [30]. To explore genome-wide changes in open chromatin conformation during adipocyte differentiation, we used FAIRE—a method of isolating genomic regions depleted of nucleosomes [7]—combined with high-throughput sequencing (FAIRE-seq) to identify open chromatin sites in the adipogenic cell line 3T3-L1 before (day 0) and after (day 8) differentiation and in NIH-3T3 fibroblasts, which cannot differentiate into adipocytes. This approach identified in the genome 37,781 FAIRE peaks in 3T3-L1 on day 0 and 26,611 on day 8, plus 36,111 in NIH-3T3 cells—all, with a false discovery rate of $<10^{-4}$. By using ChIP-seq analyses, we also generated genome-wide maps of binding sites for PPARγ, the master regulator of adipocyte differentiation, for RXRα, its heterodimer partner, for histone H3 lysine 4 trimethylation (H3K4me3), and for CCCTC-binding factor (CTCF) [31].

Figure 1 shows a representative map of results generated near *Klf15* and *Pparg*, both transcription factors up-regulated by differentiation, and both important in adipocyte differentiation [16,32]. Consistent with previous observations [10], 28% of the FAIRE peaks were detected near the transcription start sites (TSSs ±500 bp) of RefSeq genes [33] and are referred to as promoter FAIRE peaks (Figure S1A), while 72% were located outside known TSSs, and are referred to as non-promoter FAIRE peaks. Notably, only 8% of the non-promoter FAIRE peaks were located in a −5 kb proximal promoter region while the majority of non-promoter FAIRE peaks were located in introns and distal regions (Figure S1A). Average profiling revealed that a FAIRE signal, H3K4me3

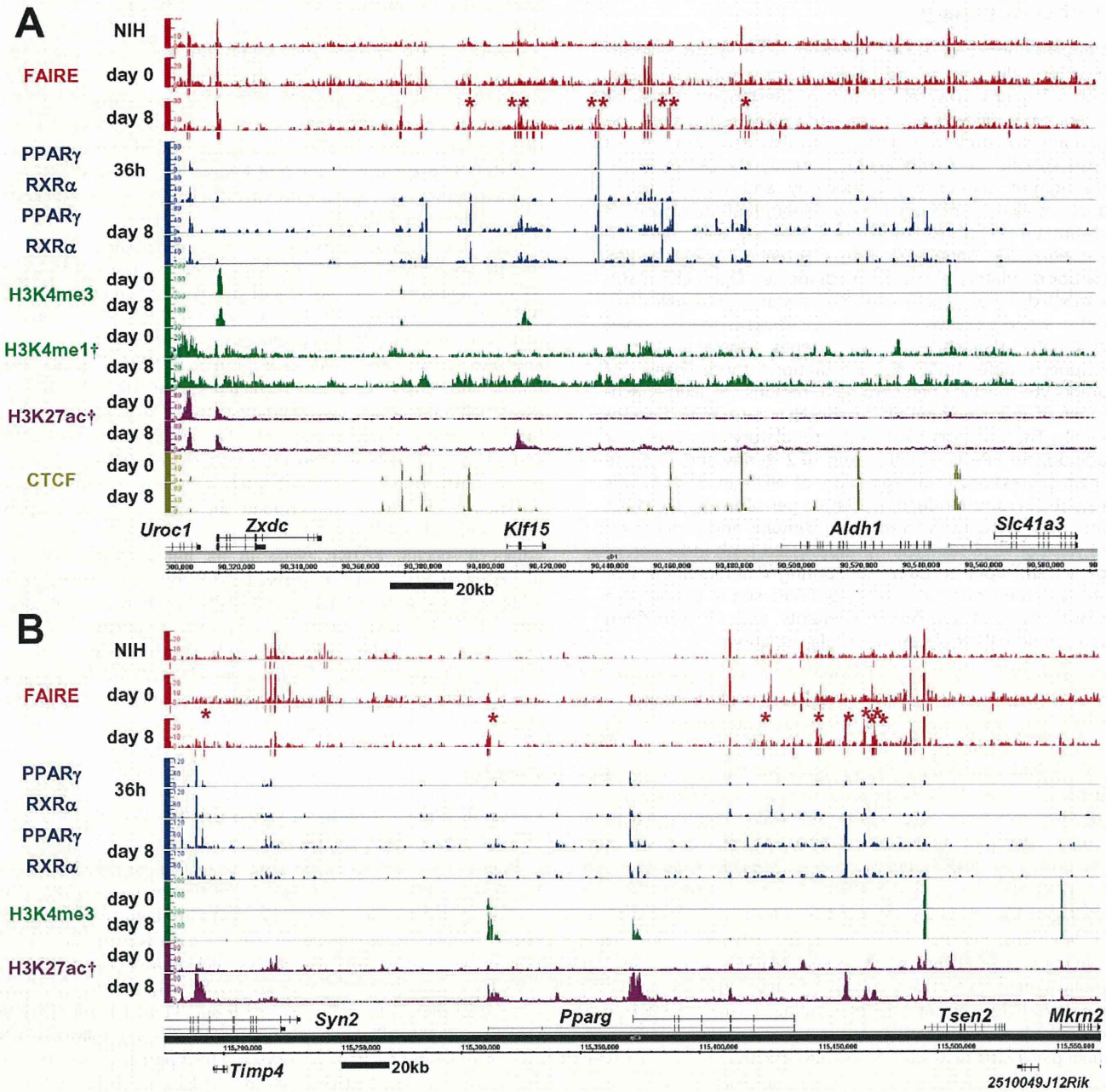

Figure 1. Genome-wide profiling of open chromatin regions by FAIRE-seq in 3T3-L1 adipocyte differentiation. Open chromatin regions detected by FAIRE-seq were observed in both promoter and non-promoter regions. The non-promoter FAIRE peaks were associated with the binding of PPARγ/RXRα or CTCF, and with the enhancer signature H3K4me1(+)/me3(−) and H3K27ac modification—while the promoter FAIRE peaks were associated with H3K4me3 and H3K27ac modification. Bars below the FAIRE peaks data represent statistically significant FAIRE positive peaks (FDR<10^{-4}). Red asterisks indicate the adipocyte-specific FAIRE peaks on day 8 (see Figure 2B for definition). Multiple adipocyte-specific FAIRE peaks were located within genomic regions near *Klf15* (A) and *Pparg* (B) in 3T3-L1 adipocytes. Data marked (†) were obtained from Mikkelsen et al. [28] (GSE20752).
doi:10.1371/journal.pgen.1002311.g001

and histone H3 lysine 27 acetylation (H3K27ac) were observed at TSSs of actively transcribed genes (Figure S1B and S1D). On the other hand, non-promoter FAIRE peaks were accompanied by monomodal enrichment of H3K4me1 and were devoid of H3K4me3 enrichment, a condition described as the signature of enhancers [34,35] (Figure S1D). CTCF binding sites are important in insulator function and high-order chromatin structure [31]. The CTCF binding sites in our study (day 0 or day 8) were largely overlapped by those in a study by Mikkelsen (day 0 or day 7) [28] (86.3% and 88.5%, respectively). CTCF binding accounted for about one fifth of either the promoter or non-promoter FAIRE peaks (Figure 1 and Figure S1C). Collectively, these data suggest that the open chromatin sites identified by FAIRE-seq show characteristics of regulatory elements such as promoter, enhancer and insulator.

Analysis of Differentiation-Dependent Non-Promoter FAIRE Peaks

We next compared the FAIRE peaks in 3T3-L1 cells on day 0 and day 8 and in NIH-3T3 cells. The promoter FAIRE peaks were relatively constant among the three groups. Over 70% of

those peaks on day 0 and day 8 3T3-L1 cells and in NIH-3T3 cells were shared by all three groups (Figure 2A). In contrast, non-promoter FAIRE peaks showed dynamic change. The three groups shared only 25%, 45%, and 26% of non-promoter FAIRE peaks in, respectively, day 0 and day 8 3T3-L1 cells and NIH-3T3 cells. This contrasts with an invariable biding pattern of CTCF in

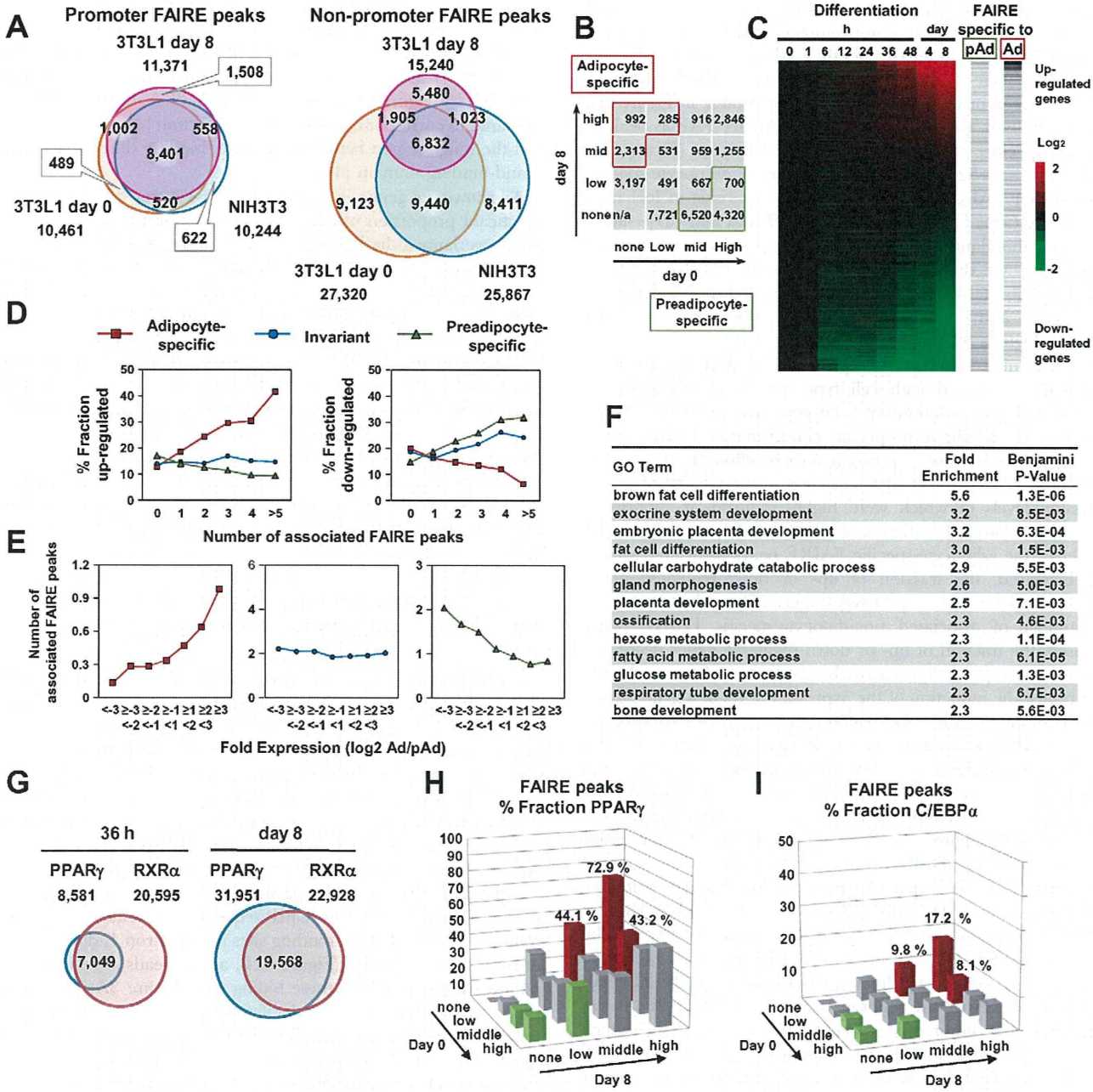

Figure 2. Cell type– and differentiation-dependent FAIRE peaks. (A) Venn diagrams comparing the FAIRE peaks among 3T3-L1 (day 0), 3T3-L1 (day 8) and NIH-3T3 at promoter (+/−500 bp from RefSeq TSS) and non-promoter regions. The promoter FAIRE peaks were relatively constant among the three cell groups while the non-promoter FAIRE peaks were highly variable. (B) The FAIRE peaks in 3T3-L1 (day 0 or day 8) were divided into tertiles by peak height and adipocyte- (red boxes) and preadipocyte-specific (green boxes) FAIRE peaks, and were defined as indicated. (C) A heat map showing enrichment of the adipocyte- and preadipocyte-specific FAIRE peaks in the vicinity (+/−25 kb from TSS) of genes up-regulated or down-regulated during differentiation. The horizontal bars in the two right panels indicate each gene with Ad or pAd FAIRE peaks in the vicinity (+/−25 kb from TSS). (D) Fractions of genes that were up-regulated (left) or down-regulated (right) more than two-fold during differentiation among genes that had the indicated number of adipocyte- (red), preadipocyte-specific (green) or invariant (blue) FAIRE peaks. (E) The number of the adipocyte- (red), preadipocyte-specific (green) or invariant (blue) FAIRE peaks associated with genes that were stratified by the ratio of the expression levels between preadipocytes and adipocytes. Each FAIRE peak was defined as associated with the nearest gene in analyses (D) and (E). (F) Ontology analysis by DAVID of genes associated (+/−25 kb from TSS) with adipocyte-specific FAIRE peaks [13]. (G) Venn diagrams showing the numbers and overlap of the binding sites for PPARγ and RXRα in 3T3-L1, day 0 and day 8. (H, I) Fractions of the non-promoter FAIRE peaks that overlap PPARγ binding sites (day 8) (H) or C/EBPα binding sites (Schmidt et al., GSE27450 [86]) (I). PPARγ and C/EBPα represented 45.3% and 11.7% of the adipocyte-specific FAIRE peaks (average of red bars).

doi:10.1371/journal.pgen.1002311.g002

the non-promoter regions; in 3T3-L1 cells, 89.5% of the non-promoter CTCF binding sites on day 0 overlapped those on day 8. What's more, a significant proportion of the non-promoter FAIRE peaks were cell type–specific (Figure 2A), implying the role of non-promoter regulatory elements in cell type–specific transcriptional regulation. We divided the non-promoter FAIRE peaks in day 0 and day 8 3T3-L1 cells into tertiles by FAIRE signal intensity, and defined adipocyte- or preadipocyte-specific FAIRE peaks as indicated by red or green boxes in the 4-by-4 table in Figure 2B. By this definition, we judged each non-promoter FAIRE peak as adipocyte-specific, preadipocyte-specific or invariant (Figure 2B). Figure 1, Figures S2 and S3 show examples of adipocyte-specific non-promoter FAIRE peaks (indicated by asterisks) in loci near *Klf15*, *Pparg*, *Cebpa* [16,20], *Mgll* [36], *Srebf1* and *cidec* [37]—all of which are abundantly expressed in adipose tissue and induced during adipocyte differentiation (data not shown). Remarkably, multiple adipocyte-specific FAIRE peaks existed in the vicinity of these genes and included introns and downstream regions (Figure 1, Figures S2 and S3).

To determine whether non-promoter FAIRE peaks were functionally associated with cell type–specific gene expression, we analyzed the relationship between the presence of the adipocyte- or preadipocyte-specific non-promoter FAIRE peaks and the change in gene expression during adipocyte differentiation. Those FAIRE peaks were enriched in the vicinity of genes, expression levels of which were highly induced or suppressed during adipocyte differentiation (Figure 2C). Importantly, as the number of the adipocyte-specific FAIRE peaks associated with a gene increased, the fraction of up- or down-regulated genes increased or decreased, respectively (Figure 2D, red lines), while as the number of associated preadipocyte-specific FAIRE peaks increased, the fraction of up- or down-regulated genes decreased or increased, respectively (Figure 2D, green lines). Conversely, the more robust the induction of the expression level of a gene during adipocyte differentiation, the greater the numbers of adipocyte-specific FAIRE peaks associated with the gene (Figure 2E, red line). In contrast, the more robust the reduction of the expression levels of a gene during adipocyte differentiation, the greater the numbers of associated preadipocyte-specific FAIRE peaks that were associated (Figure 2E, green line). Invariant FAIRE peaks were associated specifically with neither up- nor down-regulated genes (Figure 2E, blue line). We next employed a gene ontology (GO) analysis tool (DAVID) [38] to determine what kind of biological processes were associated with genes bound by the adipocyte-specific FAIRE peaks. We found that biological processes (e.g., adipocyte differentiation) were significantly enriched compared with the genomic background (Figure 2F). It was of interest that embryonic placenta development—for which PPARγ is critical [13,14,15]—was enriched (Figure 2F). Together, these data highlight the role of the cell type–specific non-promoter open chromatin sites detected by FAIRE-seq in differentiation-dependent transcriptional regulation.

Analysis of Binding Sites for PPARγ and RXRα in 3T3-L1 Adipocytes

PPARγ is key regulator of adipocyte development [16,20]. To elucidate the contribution of PPARγ to adipocyte-specific transcriptional regulation, we conducted ChIP-seq analyses using antibodies specific for either PPARγ or RXRα [24] in 3T3-L1 adipocytes at 36 hours and day 8 after induction of differentiation. The number of PPARγ binding sites increased during differentiation while that of RXRα binding sites remained virtually constant (Figure 2G). Significant overlap between the PPARγ and RXRα binding sites was consistent with the heterodimer formation of PPARγ and RXRα [21,39] (Figure 2G). Microarray and GO analysis revealed that the PPARγ binding sites were enriched in the vicinity of genes up-regulated by adipocyte differentiation (Figure S4B) and the bound genes were associated with adipocyte differentiation and lipid metabolism (Figure S4C). Using MEME [40], we performed de novo motif analysis of genomic regions bound by PPARγ, and found that the AGGTCA-n-AGGTCA (called DR-1) shown was the most over-represented one (E-value 1.3×10^{-055}) (Figure S4A). An extension AGT 5′ outside of DR-1 appeared to correspond to the direct interaction between the DNA and the hinge region between the DNA-binding domain and the ligand-binding domain [41].

As shown in genomic loci (Figure 1, Figures S2 and S3), a significant proportion of adipocyte-specific non-promoter FAIRE peaks overlapped the PPARγ/RXRα binding sites. To determine the contribution of PPARγ to the adipocyte-specific open chromatin regions, we calculated percent fractions of the FAIRE peaks that were stratified by FAIRE signal in 3T3-L1 on day 0 and day 8 (Figure 2B) —and that overlapped either the PPARγ binding sites (Figure 2G) or C/EBPα binding sites in 3T3-L1 reported by Schmidt et al. [42]. Both PPARγ and C/EBPα binding sites were enriched in the fractions of adipocyte-specific FAIRE peaks (Figure 2H and 2I), and they respectively accounted for 45.3% and 11.7% of the adipocyte-specific FAIRE peaks (averages of the red bars in Figure 2H and 2I). These data support the role of PPARγ and C/EBPα as primary transcription factors that drive adipocyte-specific gene expression—although they may not explain all of it.

Clustering of Multiple Adipocyte-Specific Non-Promoter FAIRE Peaks and the PPARγ Binding Sites

Genes that were highly induced by adipocyte differentiation often harbored multiple adipocyte-specific FAIRE peaks and/or PPARγ binding sites in their vicinity, as suggested by the linear correlation between the number of the associated adipocyte-specific FAIRE peaks and the robustness of the induction of the gene by adipocyte differentiation (Figure 2D and 2E). (See Figure 1, Figures S2 and S3 for representative genes.) To determine whether these multiple regions have functional regulatory elements, we selected AdipoR2 [43,44]. Although AdipoR2 was regulated by PPARγ and PPARα ([45] and data not shown), conventional −2 kb promoter studies failed to identify the response element [46]. Our ChIP-seq analysis revealed a cluster of multiple PPARγ/RXRα binding sites in the intron 1, downstream of the TSS of AdipoR2 (Figure S2B, arrow heads). We identified putative DR-1 motifs in these biding sites (Figure 3A) and tested them by gel-shift assay and luciferase assay. These motifs were indeed bound by the PPARγ/RXRα heterodimer, and were functional in the luciferase assay (Figure 3B and 3C), suggesting the functionality of these elements and the advantage of a genome-wide approach over the conventional "promoter-bashing" approach to identifying such response elements.

Recent genome-wide studies revealed clustering of open chromatin regions detected by Dnase I hypersensitivity assay or by FAIRE in the genomes of CD4+ T cells [47], pancreatic islet cells [10,48] and binding sites for certain transcription factors [49])—certainly the PPARγ binding sites and adipocyte-specific FAIRE peaks in our analyses tended to form clusters as indicated by an additional peak in distribution histograms of the distance to the nearest peak among the PPARγ binding sites or the adipocyte-specific FAIRE peaks (Figure 4A). We calculated the total number of PPARγ binding site clusters for different window sizes and compared them with a random data set comprised of the same number of sites (Figure 4B). The PPARγ binding sites had a

A

PPRE consensus DR-1

AGGTCAAAGGTCA

Fabp4 (aP2) (-5.4 kb)		
ARE7	+	CTTACTGGATCAGAGTTCACTAGTG
near ARE7	-	AGTAAGAAGGCAAATGTCAGGCATC
ARE6	-	CTCTCTGGGTGAATGTGCATTTCT
Adipor2 peak 1 (+20 k)		
PPRE #1	+	TGTGCAGGGACAATGCCAAGCCCC
PPRE #2	-	CCCATGAAGGCAGAGGTTATGGGCA
Adipor2 peak 2 (+17 k)		
PPRE #3	+	TGTGCTAGGGTAAAGGTGGCACAAA
PPRE #4	-	CATTGTAGGTCAAGGGTTTTGTGGC
Adipor2 peak 3 (+8 k)		
PPRE #5	+	TTTCCTGGCACACAGGTCTCCAGAC
PPRE #6	-	AGAATTGGGGCAGAGGGAAACTGTT
Adipor2 peak 4 (+6 k)		
PPRE #7	+	TTTGCCAGGGAAATGGTCACAGAGT
PPRE #8	-	CTATGTAGGCCAGAGGGTGATCTGC

B

Peak	Peak 1							Peak 2						Peak 2			Peak 3			Peak 4	
^{32}P-AdipoR2 PPRE	#2							#3						#4			#6			#8	
Cold aP2 PPRE	-				x20	x200		-				x20	x200	-	x200		-		x200		x200
PPARγ	-	-	+	+	+	+		-	-	+	+	+	+	-	+	+	-	+	+	+	+
RXRα	-	+	-	+	+	+		-	+	-	+	+	+	-	+	+	-	+	+	+	+

Free probe

C

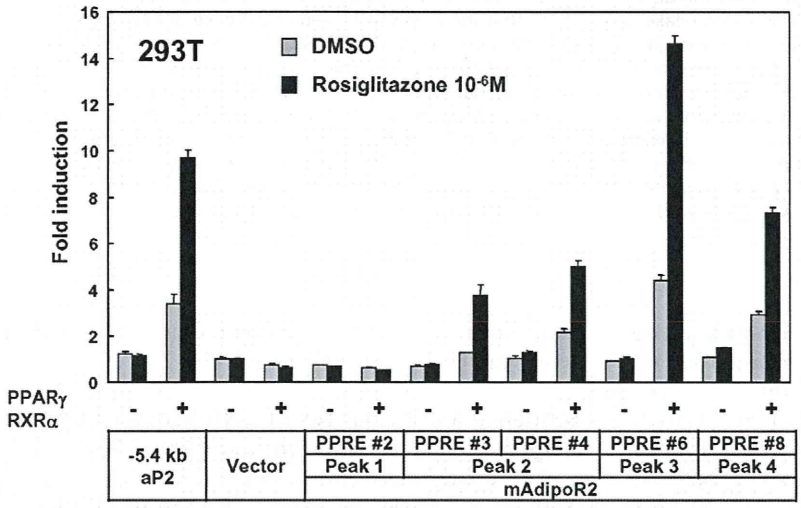

Figure 3. Identification of functional regulatory elements in the intron 1 of *Adipor2*. The PPARγ binding sites in the *Adipor2* gene locus (Figure S2B, arrow heads) were analyzed. (A) Putative DR-1 motifs (PPAR response elements or PPREs) in the regions. ARE6 and ARE7 in the −5.4 kb promoter upstream of *Fabp4*(aP2) were previously identified PPREs [87]. (B) Gel shift assay showing binding of the PPARγ/RXR heterodimer to the motifs. An arrow indicates the PPARγ/RXRα heterodimer bound by radiolabeled probe. Competition by cold oligos showed the specificity of the binding. (C) Luciferase reporter assay in HEK293T cells. Most of the motifs inserted into reporter vectors with the tk minimal promoter responded to over-expressed PPARγ/RXRα and stimulation with its synthetic ligand, rosiglitazone. The −5.4 kb promoter of PPARγ target gene *Fabp4* (aP2) [84] was included as a positive control.

doi:10.1371/journal.pgen.1002311.g003

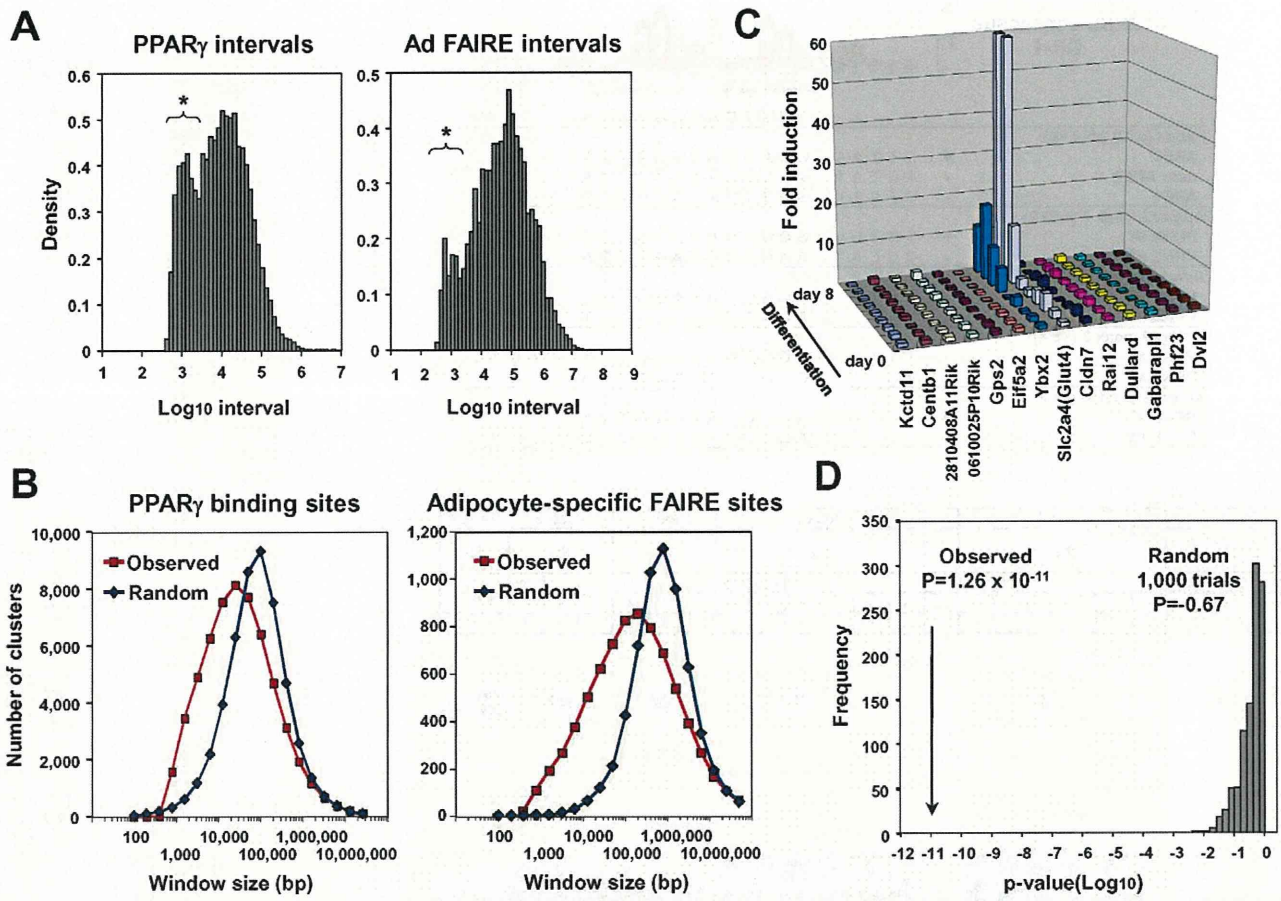

Figure 4. Statistical analyses for clustering of adipocyte-specific FAIRE peaks and PPARγ binding sites and co-regulation of neighbor genes during adipogenesis. (A) Histogram showing distribution of intervals (defined as distances to the nearest neighbor sites) among all PPARγ peaks (left) and among the adipocyte-specific FAIRE peaks (right). Note that there was increased occurrence of sites separated by short intervals (indicated by asterisks). See [48] for details of the method. (B) Clustering analysis of the PPARγ binding sites and the adipocyte-specific FAIRE peaks by counting the total number of clusters (defined as more than two peaks) determined for windows with indicated width. The PPARγ binding site or adipocyte-specific FAIRE peak clusters occurred more frequently in the observed data set than in random data with the same number of sites. The difference in the number of clusters was observed at window sizes ranging from 800 bp to 30~100 kb compared with the random sample. See reference [47] for details of the method. (C) Microarray analysis showing both *Slc2a4 (Glut4)* and *Ybx2* included in the adipocyte-specific FAIRE peak cluster (Figure S2C) co-regulated during differentiation. (D) Neighbors of highly induced genes (>10 fold) were more likely to be up-regulated over three fold (18%, or 112 of 618 neighbors) than the 2,012 of 21,343 total genes (9%) that were up-regulated over three fold ($p = 1.26 \times 10^{-12}$, one-sided Fisher test). Neighbors of randomly selected genes were not significantly up-regulated ($p = -0.67$, average of 1,000 trials). See reference [50] for method.

doi:10.1371/journal.pgen.1002311.g004

significantly higher number of clusters in a window size raging from 800 bp to ~30 kb. Similar results were obtained for the adipocyte-specific FAIRE peaks (Figure 4A and 4B).

On the other hand, multiple genes involved in adipocyte function [55,56,57] were often co-regulated in certain genomic regions that harbor clusters of adipocyte-specific regulatory elements (see Figure S2C, Figure 4C, and Figure S5). We therefore statistically tested—method in reference [50]—to see if neighboring genes tended to be co-regulated during adipocyte differentiation, and found that neighbors of highly induced genes (>10 fold) were indeed more likely to be up-regulated over three fold (18%, or 112 of 618 neighbors) than the 2,012 of 21,343 total genes (9%) that were up-regulated over three fold ($P = 1.26 \times 10^{-12}$, one-sided Fisher test). Neighbors of randomly selected genes were not significantly up-regulated ($p = -0.67$, average of 1,000 trials, Figure 4D). Together, these data suggest that the transcriptional regulation of genes during adipocyte differentiation involves multiple adipocyte-specific regulatory elements—which tend to form clusters—and that co-regulation of neighboring genes often occurs during adipocyte differentiation.

Sequence Motif Analyses of DNA Sequences of the Adipocyte-Specific Non-Promoter FAIRE Peaks

Next, we performed enrichment analyses of known motifs using AME in the MEME suite and the TRANSFAC [51] and JASPER [52] motif databases to identify motifs enriched in either adipocyte- or preadipocyte-specific FAIRE peaks compared with the background (statistical values shown as corrected p-value in Figure 5). We also determined the enrichment ratio (Ad/pAd) by calculating the ratio of occurrence of a motif in the adipocyte-specific FAIRE peaks and in the preadipocyte-specific FAIRE peaks as described in reference [28]. Using both parameters, we obtained motifs that had been significantly enriched in either kind of FAIRE peak and that occurred in significantly different number. Figure 5 shows the top of the list of TRANSFAC motifs enriched in the adipocyte- and preadipocyte-specific FAIRE

TRANSFAC motifs

Adipocyte-specific FAIRE peaks

Motif	Name	Corrected p-value	Enrich-ment Ratio (Ad/pAd)	Logo
M00193	NF-1	7.9E-27	1.60	
M01196	CTF1 (NF-1)	5.1E-22	1.55	
M01100	LRF	2.6E-20	1.65	
M00528	PPAR	2.7E-12	2.14	
M01728	EAR2	1.2E-09	1.47	
M01031	HNF4 (PPAR)	3.8E-08	2.06	
M01772	C/EBP	1.7E-07	2.69	
M00109	C/EBPbeta	3.1E-07	1.51	
M00121	USF/Tcf4 Max/c-Myc	6.3E-07	1.52	
M00491	MAZR	2.1E-05	1.29	

Preadipocyte-specific FAIRE peaks

Motif	Name	Corrected p-value	Enrich-ment Ratio (Ad/pAd)	Logo
M00925	AP-1	1.1E-221	0.07	
M00495	Bach1	1.2E-183	0.09	
M00037	NF-E2	2.3E-84	0.23	
M00769	AML	1.8E-15	0.53	
M00984	PEBP	3.1E-15	0.49	
M01305	TEF	2.7E-13	0.44	
M00284	TCF11:MafG	5.2E-12	0.30	
M00115	Tax/CREB	2.2E-06	0.47	
M01080	CBF	1.1E-05	0.50	
M01666	STAT4	3.6E-02	0.59	

Figure 5. Known motif enrichment analysis of adipocyte- or preadipocyte-specific FAIRE peaks (TRANSFAC motifs). Enrichment analysis of the adipocyte- (left) and the preadipocyte-specific (right) FAIRE peaks for known motifs in the TRANSFAC database (Release 2010.4) performed by using AME in the MEME suite. After removing repeat regions with RepeatMasker [83], DNA sequences from the center 150 bp regions of the top 2,000 cell type–specific FAIRE peaks were analyzed (p-value report threshold : 0.05). Motif enrichment ratios (Ad/pAd FAIRE) for motifs in the TRANSFAC database were also determined by a method described in reference [28]. Motifs with an enrichment ratio greater than 1.20 (for the adipocyte-specific FAIRE peaks, left) or less than 0.833 (for the preadipocyte-specific FAIRE peaks, right) are shown in the table. See "Materials and Methods" for details.
doi:10.1371/journal.pgen.1002311.g005

peaks. The motifs for PPARγ (and other DR1 motifs) and C/EBPs were among the list, consistent with their critical roles in adipogenic transcription. Motif analyses using the JASPER motif database showed enrichment of the motifs for PPARγ, C/EBPs and the motif for Zfp423, a recently identified adipogenic regulator [53] (Figure S6). Motif analyses of the preadipocyte-specific FAIRE peaks showed significant enrichment of a motif for AP-1, a downstream transcription factor complex of the growth factor/MAP kinase signaling pathways, which include epidermal growth factor and c-Jun N-terminal kinases, known inhibitors of adipogenesis [54,55] (Figure 5 and Figure S6). We also performed de novo motif analysis (MEME) [40] of the adipocyte-specific FAIRE peaks, and observed significant enrichment of motifs that corresponded to those for PPARγ and C/EBPs (Figure S7). Together, these instances of enrichment of known regulators indicate the validity of this approach.

Identification of NFI Family Transcription Factors as Novel Regulators of Adipocyte Differentiation

There were several other motifs for transcription factors, their functions not previously linked to adipocyte differentiation (Figure 5, Figures S6 and S7). We focused on a motif for the NFI family transcription factors. The murine NFI family consists of NFIA, NFIB, NFIC and NFIX, and was identified as a site-specific DNA-binding protein that bound to the adenovirus origin of replication [56]. It forms a dimer to bind to the symmetric consensus sequence TTGGC(N5)GCCAA [57]. We first examined the expression change of these factors in in vitro adipocyte differentiation and found that the expression of NFIA and NFIB were significantly induced during differentiation of 3T3-L1 and of another adipogenic cell line, 3T3-F442A (Figure 6A and 6C). Consistent with this pattern, both NFIA and NFIB were highly expressed in a variety of adipose tissue depots in addition to the brain (Figure 6B). We next examined the effect of siRNA knockdown of NFIA and NFIB on adipogenic gene regulation and adipocyte differentiation (Figure 6C). Interestingly, induction of the expression of the adipogenic transcription factors PPARγ and C/EBPα and of downstream genes was significantly suppressed by siRNA knockdown of either NFIA or NFIB (Figure 6C). Consistent with the gene expression change, we observed significant reduction of lipid accumulation as judged by oil red O staining, suggesting physiological roles for NFIA and

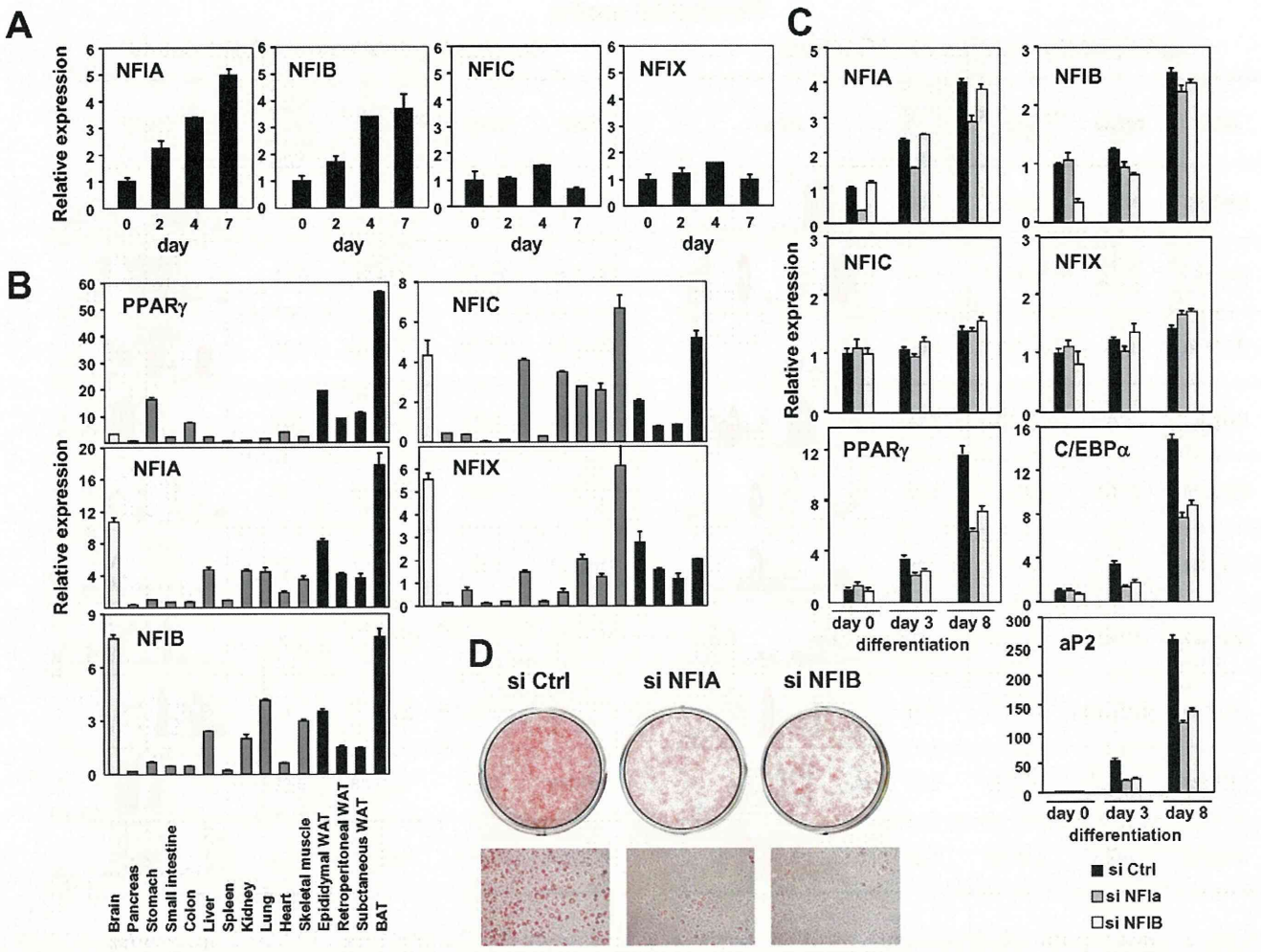

Figure 6. NFIA and NFIB are novel regulators of adipocyte differentiation. (A) Transcriptional regulation of NFI transcription factors during adipocyte differentiation (3T3-F442A). (B) Tissue distribution of the NFI family genes. Expression levels relative to 36B4 in various tissues were determined by qPCR. (C, D) Effects of siRNA-mediated knockdown of NFIA and NFIB on adipogenic gene expression (C) and lipid accumulation in 3T3-L1 adipocytes judged by oil red O staining (D). Knockdown of either NFIA or NFIB resulted in suppression of the induction of PPARγ, C/EBPα and the PPARγ target gene, aP2, as well as increase in lipid accumulation during adipocyte differentiation.
doi:10.1371/journal.pgen.1002311.g006

NFIB in adipocyte differentiation (Figure 6D). We confirmed the effect of NFIA and NFIB knockdown on adipogenesis by using independent pooled siRNA (Figure S8).

We next asked whether overexpression of these factors influence adipocyte differentiation. We amplified NFIA and NFIB coding sequences from cDNA prepared from adipocytes, and cloned them into retroviral pMXs-puro vectors. We also made a dominant negative NFIA that lacks the C-terminal transactivation/repression domain (NFIA-DN) [58]. Overexpression of NFIA—but not NFIA-DN or NFIB—resulted in robust induction of PPARγ, C/EBPα and aP2 (Figure 7A) at a basal state. Surprisingly, the induction of these factors was robust enough to make the cells to form lipid droplets visible and stainable by oil red O even before initiation of differentiation by the DMI (dexamethasone, IBMX and insulin) treatment (Figure 7B and 7C). However, after the DMI treatment, NFIA-expressing cells were overtaken by control cells, and on day 7, NFIA and NFIB overexpressing cells showed attenuated differentiation (Figure 7D and 7E). We speculate that this was caused by secondary effects of overly strong overexpression levels (>30 fold, Figure 7A). Almost complete suppression of adipogenesis by NFIA-DN overexpression was consistent with the results of knockdown experiments (Figure 6, Figure 7D and 7E). Nevertheless, the robust induction of PPARγ, C/EBPα and aP2 by NFIA overexpression at the basal state implies direct action of NFIA on transcriptional control of these factors.

To dissect the mechanism by which NFIs regulate PPARγ, C/EBPα and aP2, we examined DNA sequences of the adipocyte-specific FAIRE peaks and/or the PPARγ binding sites in the vicinity of these factors and found that some of them have NFI binding motifs as listed in Figure 8A. ChIP analysis using an anti-NFI antibody confirmed actual binding of NFI to these sites (Figure 8B and 8C). We extended this experiment by counting NFI motifs in the FAIRE peaks on a genome-wide scale. Interestingly, percent fractions of genes harboring NFI binding motifs in the FAIRE peaks were higher when the genes were bound by PPARγ and induced during differentiation (Figure 8D), indicating a significant degree of specificity for the NFI's action on the adipogenic transcriptional program.

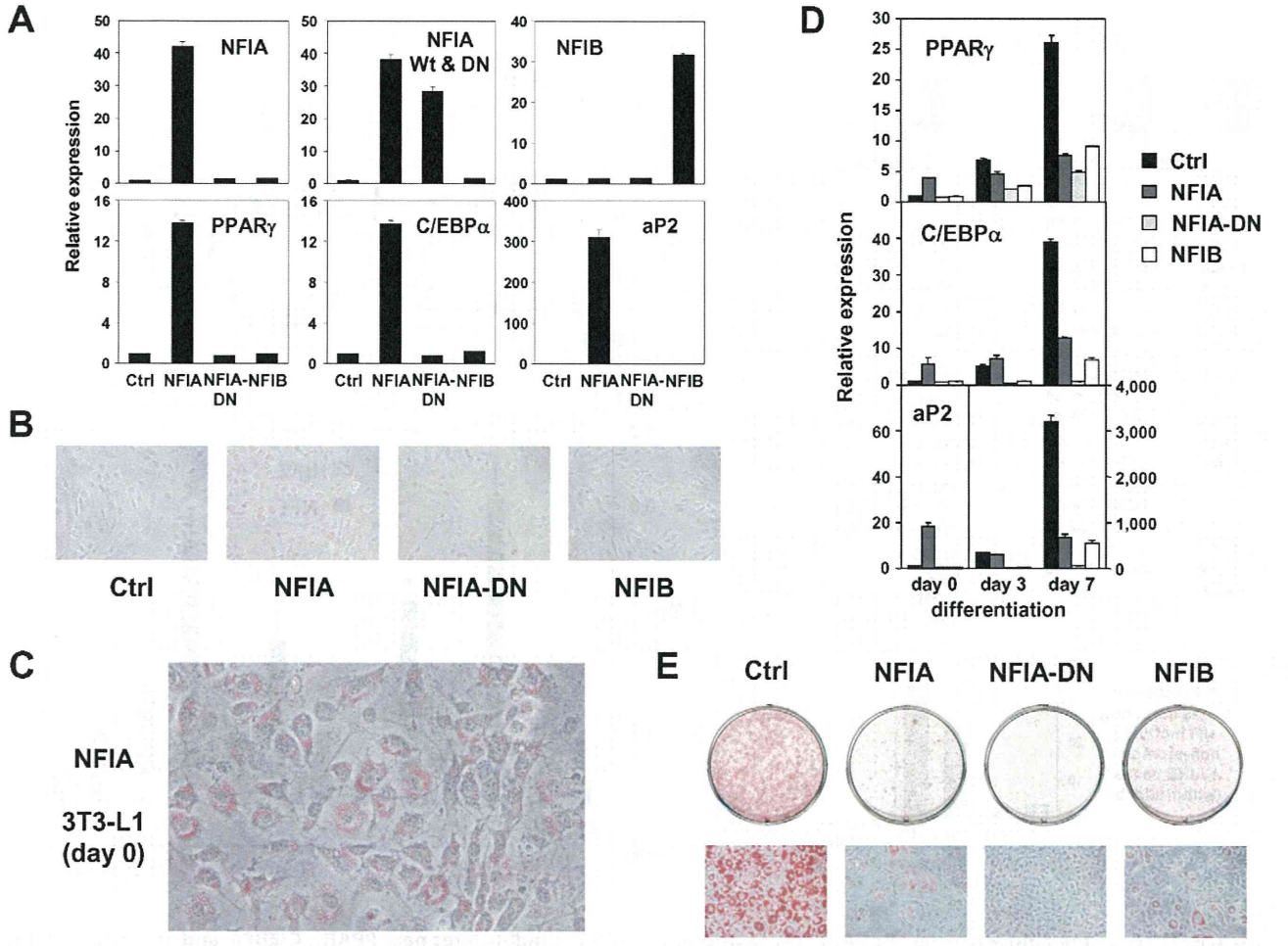

Figure 7. Overexpression of NFIA, NFIB, and dominant negative NFIA in 3T3-L1 cells. (A) Expression analysis of overexpressed NFI factors (upper panel) and adipogenic PPARγ, C/EBPα and aP2 (lower panel). Note, overexperssion of NFIA resulted in a robust induction of adipogenic factors. (B) Microscopic pictures of 3T3-L1 cells overexpressing NFI factors at confluence stained by oil red O (day 0). (C) Close examination of NFIA-overexpressing cells revealed formation of lipid droplets without adipogenic stimulus before differentiation. (D) Time course of expression levels of PPARγ, C/EBPα and aP2 during differentiation. Note, the induction of these genes by NFIA overexpression was overtaken by that of control cells, and on day 7, NFIA and NFIB overexpressing cells showed attenuated differentiation. Dominant negative NFIA showed almost complete suppression. (E) Oil red O staining of 3T3-L1 overexpressing NFI factors on day 7.
doi:10.1371/journal.pgen.1002311.g007

Collectively, we demonstrated that the combination of FAIRE-seq and computational motif analyses is useful in identifying novel regulators of adipocyte differentiation.

Comparison of FAIRE Peaks between Undifferentiated 3T3-L1 and NIH-3T3 Cells

The 3T3-L1 adipogenic cell line was established by isolating clonal sublines of mouse fibroblast line 3T3 [59]. Lastly, we compared FAIRE peaks between 'undifferentiated' 3T3-L1 and NIH-3T3 cells. As shown in Figure 2A, a substantial proportion of FAIRE peaks was unique to either 3T3-L1 or NIH-3T3 cells. We defined non-promoter FAIRE peaks as specific to 3T3-L1and NIH-3T3—as we did for the adipocyte- or preadipocyte-specific FAIRE peaks in Figure 2B. The 3T3-L1- or NIH-3T3-specific FAIRE peaks were enriched in the vicinity of genes whose expression levels were higher in 3T3-L1 or NIH-3T3, respectively (Figure S9A). Motif analysis of the 3T3-L1-specific FAIRE peaks showed that the binding motif for EBF (Figure S9B) had the highest enrichment ratio (1.81) and a statistically significant p-value of 3.9E-3. Although the p-value of the motif for PPARγ/RXR did not reach statistical significance, that motif had an enrichment ratio of 1.84. These two factors were among the handful that were proven to transform NIH-3T3 cells into adipocytes when ectopically introduced [16,60].

Discussion

We demonstrated that genome-wide mapping of open chromatin regions by FAIRE-seq is a simple, accurate method that allows a snapshot view of regulatory elements in the genome. Although open chromatin regions detected by FAIRE-seq include promoters of transcribed genes, enhancers and insulators, open chromatin regions that vary in two different conditions likely contain regulatory elements that play roles in the specific biological process. By comparing open chromatin regions in preadipocytes and adipocytes, we identified the adipocyte- and preadipocyte-specific FAIRE peaks in the genome. Functionally, we demonstrated that the adipocyte-specific FAIRE peaks were associated with genes up- regulated by adipogenesis while the preadipocyte-specific FAIRE peaks were associated with genes down-regulated by adipogenesis (Figure 2C, 2D and 2E). Adipocyte gene

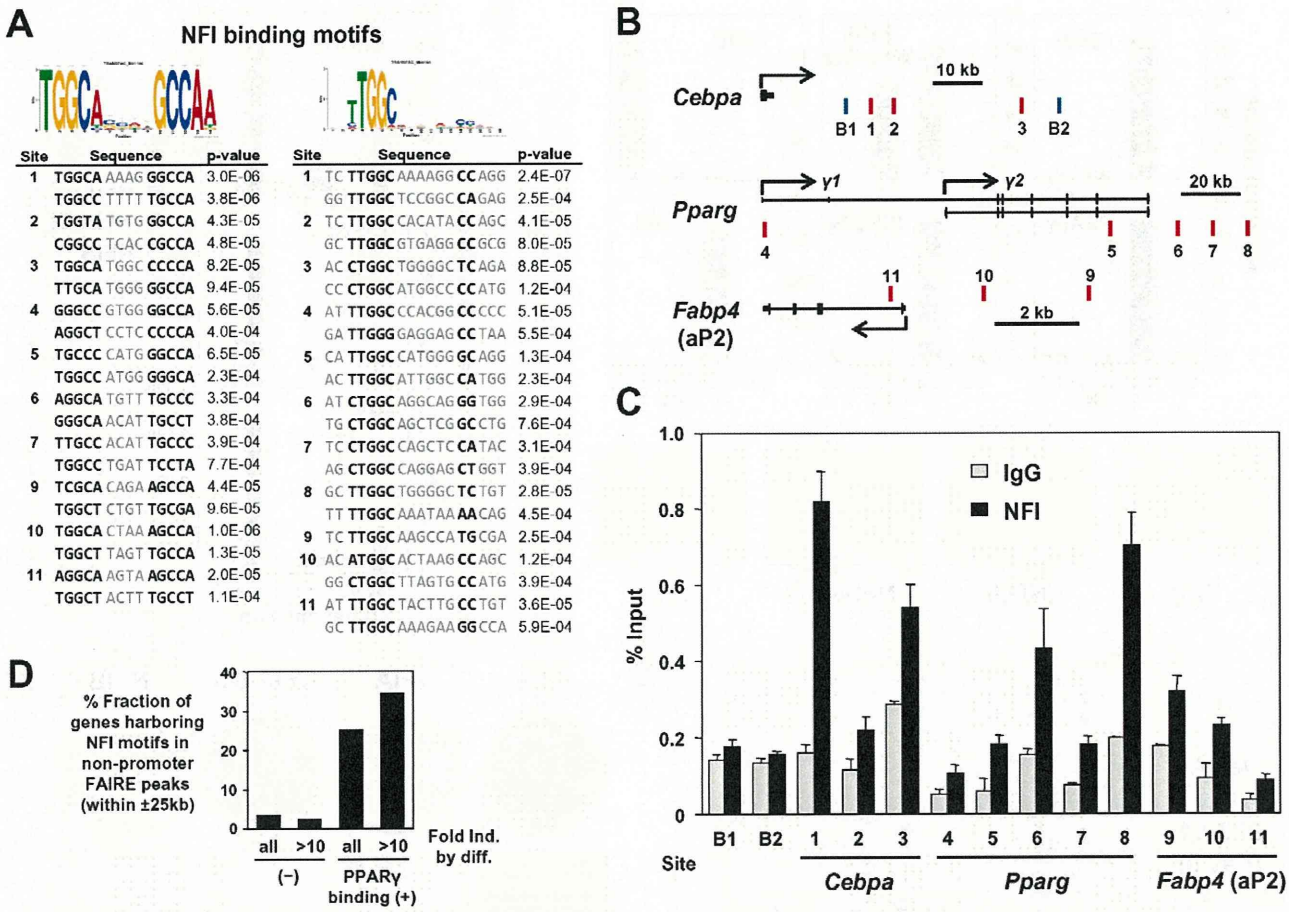

Figure 8. NFI occupy the adipocyte-specific FAIRE peaks and/or the PPARγ binding sites near PPARγ, C/EBPα, and aP2 genes. (A) The NFI binding motifs identified in the adipocyte-specific FAIRE peaks and/or the PPARγ binding sites in the vicinity of PPARγ, C/EBPα and aP2. For site numbers, see (B). (B) Genomic location of the regions examined. B1 and B2 are unrelated genomic regions used as background negative controls. (C) ChIP-qPCR analysis using an anti-NFI antibody (H-300). (D) Percent fraction of genes harboring NFI motifs in non-promoter FAIRE peaks (within ±25 kb) were higher when the genes were bound by PPARγ (within ±25 kb) and induced during differentiation.
doi:10.1371/journal.pgen.1002311.g008

expression appears mediated through multiple regulatory elements distal to transcription start sites (TSSs): greater induction of gene expression by differentiation means greater likelihood that more adipocyte-specific FAIRE peaks are associated with the gene (Figure 2D and 2E). This implies that optimal gene transcriptional regulation may require coordinated actions of multiple regulatory elements. Therefore, although valuable and informative, the proximal promoter assay may not always be sufficient (e.g., AdipoR2, see Figure S2B and Figure 3). Nevertheless, the importance of proximal promoter regions is obvious given the fact that many proximal promoter regions are successfully used to generate tissue-specific transgenic lines. Recently, Mikkelsen et al. demonstrated in adipocytes that many *cis*-regulatory elements are often not conserved between human and murine adipocytes even though the expression pattern of genes is conserved [28]. They observed that such motifs were located within linage-specific transposon insertions. Existence of multiple regulatory elements around biologically important genes *could* be a mechanism by which cells maintain key gene regulations against genomic changes during evolution. Clustering of regulatory elements could also result from an accumulative effect of such evolutional genomic changes.

Computational motif analysis is used to discover new transcription-factor binding motifs in sequences inferred from genome-wide studies such as ChIP-seq [61]. In genome-wide ChIP analysis of transcription factors, motif analysis is used to obtain their accurate binding motifs *and* discover unknown DNA binding factors that co-localize with the transcription factors of interest, for example, see [27,62,63]. The analyses, however, relied on prior knowledge about transcription factors and the regions to be analyzed are limited to their biding sites. In contrast, the combination of motif analyses and mapping of regulatory elements by FAIRE-seq does not require such prior knowledge, hence offers a distinct advantage in unbiased screening for novel transcription factors important in given biological processes. In our study, we retrieved the motifs for PPARγ and C/EBPs and for known regulators that top the list of the motifs identified in the adipocyte- or preadipocyte-specific FAIRE peaks (Figure 5, Figures S6 and S7). Furthermore, we demonstrated that NFIA and NFIB were functionally required for proper adipocyte differentiation (Figure 6). These results demonstrated that motif analyses of cell type–specific FAIRE peaks are useful in identifying regulators of a biological process in an unbiased manner.

To our knowledge, few studies have employed motif analysis and our unbiased approaches in investigating enhancer-like DNA regions. Mikkelsen et al. recently employed ChIP-seq for H3K27ac to define enhancer regions specific for adipocyte differentiation. Both studies similarly detected the motifs for PPAR, C/EBPs and