Published in final edited form as:

Hum Genet. 2015 August ; 134(8): 895–904. doi:10.1007/s00439-015-1572-3.

Genetic polymorphism of *APOB* is associated with diabetes mellitus in sickle cell disease

Xu Zhang^{1,*}, Wei Zhang², Santosh L. Saraf¹, Mehdi Nouraie³, Jin Han⁴, Michel Gowhari¹, Johara Hassan¹, Galina Miasnikova⁵, Adelina Sergueeva⁶, Sergei Nekhai³, Rick Kittles⁷, Roberto F. Machado⁸, Joe G N Garcia⁷, Mark T. Gladwin⁹, Martin H. Steinberg¹⁰, Paola Sebastiani¹¹, Donald A. McClain¹², and Victor R. Gordeuk^{1,*}

¹Comprehensive Sickle Cell Center, Section of Hematology/Oncology, Department of Medicine, University of Illinois at Chicago, Chicago, IL

²Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL

³Center for Sickle Cell Disease, Howard University, Washington, DC

⁴Department of Pharmacy Practice, College of Pharmacy, University of Illinois at Chicago, Chicago, IL

⁵Chuvash Republic Clinical Hospital 1, Cheboksary, Russia

⁶Cheboksary Children's Hospital, Cheboksary, Russia

⁷University of Arizona, College of Medicine, Tucson, AZ

⁸Department of Medicine, Pulmonary and Critical Care Medicine, University of Illinois at Chicago, Chicago, IL

⁹Division of Pulmonary, Allergy, and Critical Care Medicine, Vascular Medicine Institute, University of Pittsburgh, Pittsburgh, PA

¹⁰Department of Medicine, Boston University School of Medicine, Boston, MA

¹¹Department of Biostatistics, Boston University School of Public Health, Boston, MA

¹²Department of Internal Medicine, Wake Forest University School of Medicine and VA Medical Center, Winston Salem, NC

Abstract

Environmental variations have strong influences in the etiology of type 2 diabetes mellitus. In this study, we investigated the genetic basis of diabetes in patients with sickle cell disease (SCD), a Mendelian disorder accompanied by distinct physiological conditions of hypoxia and hyperactive

Correspondence should be addressed to X.Z. (xuzhangchicago@gmail.com) and V.R.G. (vgordeuk@uic.edu).

Author Contributions: X.Z., D.A.M., V.R.G. conceived and designed study; X.Z., W.Z., V.R.G. analyzed data; X.Z., W.Z., S.L.S., M.N., J.H., M.G., J.H., G.M., A.S., S.N., R.K., R.F.M., J.G.N.G., M.T.G., M.H.S., P.S., D.A.M., and V.R.G. contributed material and tools; X.Z., W.Z., S.L.S., M.N., J.H., M.G., J.H., G.M., A.S., S.N., R.K., R.F.M., J.G.N.G., M.T.G., M.H.S., P.S., D.A.M., and V.R.G. wrote paper.

Competing Financial Interests: The authors declare no competing financial interests.

erythropoiesis. Compared to the general African-American population, the prevalence of diabetes as assessed in two SCD cohorts of 856 adults was low, but it markedly increased with older age and overweight. Meta-analyses of over 5 million single nucleotide polymorphisms (SNPs) in the two SCD cohorts identified a SNP, rs59014890, the C allele of which associated with diabetes risk at $P=3.2\times 10^{-8}$ and, surprisingly, associated with decreased *APOB* expression in peripheral blood mononuclear cells (PBMCs). The risk allele of the *APOB* polymorphism was associated with overweight in 181 SCD adolescents, with diabetes risk in 592 overweight, non-SCD African Americans 45 years of age, and with elevated plasma lipid concentrations in general populations. In addition, lower expression level of *APOB* in PBMCs was associated with higher values for percent hemoglobin A1C and serum total cholesterol and triglyceride concentrations in patients with Chuvash polycythemia, a congenital disease with elevated hypoxic responses and increased erythropoiesis at normoxia. Our study reveals a novel, environment-specific genetic polymorphism that may affect key metabolic pathways contributing to diabetes in SCD.

Introduction

Type 2 diabetes mellitus (T2D) occurs when impaired insulin effectiveness is accompanied by decreased insulin production by β cells. With 366 million people diagnosed in 2011 and a trend of increasing prevalence worldwide (Lyssenko and Laakso 2013), diabetes is one of the major threats to human health. Both environmental and genetic factors contribute to the risk of T2D, as demonstrated by the 50-92% disease concordance among monozygotic twins compared to a 37% concordance in dizygotic twins (Florez et al. 2003). Since 2007, more than 60 genetic loci have been associated with T2D in large-scale genome wide association studies (GWAS) that have supported a polygenic model for T2D with many causal variants each of modest effect (Morris et al. 2012). These common genetic loci explain only about 10% of familial aggregation of the disease suggesting a role for non-genetic factors and gene-environment interactions (Permutt et al. 2005).

Sickle cell disease (SCD) is due to homozygosity for a Glu6Val mutation in *HBB* (sickle cell anemia; hemoglobin SS) or to compound heterozygous forms like hemoglobin SC and hemoglobin S- β thalassemia (Pauling et al. 1949). The hemoglobin S mutation allows deoxy-hemoglobin to polymerize, distorting sickle erythrocytes and causing hemolytic anemia and obstruction of the microvasculature that lead to acute and chronic organ damage (Rees et al. 2010). The consequent chronic hypoxia, enhanced erythropoiesis, inflammation and oxidative stress (Akohoue et al. 2007) impose distinct physiological environments that may alter metabolism in SCD. For example, metabolic measurements using indirect calorimetry and doubly labeled water technique indicated elevated resting energy expenditure but decreased activity-related energy expenditure in SCD children compared to matched healthy subjects (Barden et al. 2000). High baseline metabolism and low body mass index (BMI) (Barden et al. 2002) may provide protection from T2D in SCD (Morrison et al. 1979). In contrast, endogenous or exogenous iron overload due to hemolysis and blood transfusions can result in β -cell damage and decreased insulin production promoting diabetes (Simcox and McClain 2013). Past studies suggested a low prevalence of diabetes in patients with SCD (Morrison et al. 1979). Improvements in treatment and care have increased the life span of patients (Elmariah et al. 2014; Platt et al. 1994). This, along with

the wide availability of high calorie diets and increasing adiposity in SCD, lead us to investigate the genetic basis of diabetes in SCD.

Results

Prevalence of diabetes in SCD

The number of diabetes cases was small in our study: 16 in the University of Illinois at Chicago (UIC) cohort (Saraf et al. 2014) and 14 in the Walk-PHaSST cohort (Machado et al. 2011) (Supplemental Table 1). We first compared the 856 adult patients from these two cohorts with 2579 non-Hispanic black individuals from the 2009-2012 National Health and Nutrition Examination Survey (Figure 1). As expected, percent overweight (BMI ≥ 25 kg/m²) and percent diabetes, defined as diagnosis of diabetes in the medical record and treatment with a glucose-lowering agent, were lower in SCD compared to control individuals. Percent overweight increased with age in SCD patients (34% in patients 18-44 years of age vs 54% in those 45-85 years of age) as in control individuals (69% vs 78%). In the younger age group, one (0.25%) diabetes case was found in non-overweight SCD patients and three (1.5%) were found in overweight SCD patients, representing a six-fold increase with overweight albeit not statistically significant due to small case numbers. In the younger control individuals, percent diabetes increased with overweight by three-fold (1.4% vs 4.5%). In the older age group, two (1.7%) diabetes cases were found in non-overweight SCD patients and twenty-four (17%) were found in overweight SCD patients, representing a ten-fold increase with overweight. In the older control individuals, percent diabetes did not quite double with overweight (16% vs 29%). These observations suggest that the protection from diabetes, although strong in younger and normal weight SCD patients, breaks down quickly with older age and overweight.

Compared to hemoglobin SS and Sbeta⁰-thalassemia, hemoglobin SC and Sbeta⁺-thalassemia are milder forms of SCD. Percent overweight in mild SCD was higher than that in severe SCD: 57% vs 28% ($P=1.3\times 10^{-8}$) in younger patients and 80% vs 43% ($P=2.0\times 10^{-8}$) in older patients. Among younger patients, two (1.7%) diabetes cases were found in mild SCD, compared to two (0.42%) found in severe SCD ($P=0.17$). Among older patients, twenty (25%) diabetes cases were found in mild SCD, compared to six (3.3%) in severe SCD ($P=2.7\times 10^{-7}$). Therefore, compared to severe SCD, mild SCD showed reduced protection from diabetes.

Genome-wide association with diabetes in SCD

To uncover genetic regulation of diabetes under the distinct cellular environments characterized by SCD, we carried out a genome wide meta-analysis in UIC and Walk-PHaSST cohorts (Supplemental Table 2). Considering our limited sample size, we aimed to identify common risk genetic variants with large effect. We imputed the genotypes of these cohorts to 1000 genomes data using African and European reference panels, since patients included in the meta-analysis were all African-Americans with admixed European ancestry. We scanned over five million SNPs with imputation quality r^2 (imputation dosage r^2) >0.5 and with minor allele frequency (MAF) >0.1 within each cohort, for association with diabetes. As demonstrated above, age and BMI were two strong environmental factors that

determined the risk of diabetes in SCD. In addition, primary model selection procedures carried out in the two cohorts indicated *HBB* genotype (contrasting mild vs severe SCD) and hydroxyurea treatment as predictors of disease outcome. We therefore included age, gender, *HBB* genotype, hydroxyurea treatment and BMI as covariates in the logistic models for association.

One SNP, rs59014890, was associated with diabetes risk with genome-wide significance ($P=3.2\times 10^{-8}$, Figure 2A) in the combined analysis. The C allele of the SNP increased diabetes risk in both cohorts (OR=17, $P=0.00047$ in UIC and OR=18, $P=0.000016$ in Walk-PHaSST, Table 1). For the TT, TC and CC genotypes, the percent diabetes was 5.1%, 9.5% and 13% in the UIC cohort and 0%, 3.1% and 15% in the Walk-PHaSST cohort, respectively (Supplemental Table 3). As shown in Supplemental Table 4, adjustment of covariates, especially age and BMI, improved the associations (without adjustment, OR=1.8, $P=0.22$ in UIC and OR=9, $P=0.000061$ in Walk-PHaSST). The risk allele of rs59014890 consistently increased diabetes risk among mild SCD (OR=2.9 in UIC, OR=12 in Walk-PHaSST, combined- $P=0.0024$) and severe SCD (OR=7.2 in UIC, OR=7.8 in Walk-PHaSST, combined- $P=0.0056$) patients (Supplemental Table 5).

To examine the potential function of rs59014890, we tested its association with expression levels of seven genes, located within $\pm 1\text{Mb}$ region of the SNP, in peripheral blood mononuclear cells (PBMCs) derived from two independent samples: 134 individuals from the UIC cohort, and 35 adult SCD patients from a Howard University cohort (Zhang et al. 2014b). Within each dataset, gene expression levels were regressed on the SNP dosages, adjusting for age, gender, *HBB* genotype, hydroxyurea treatment, and variation in ancestral admixture. The C allele of rs59014890 decreased expression level of *APOB* in both the UIC ($\beta=-0.021$, $P=0.028$) and Howard ($\beta=-0.030$, $P=0.15$) cohorts (Figure 2B), with a combined P value of 0.0082 (Table 1). None of the other genes showed significant expression association with rs59014890 (Supplemental Table 6).

Validation of the candidate association

As few diabetes cases have been recorded in other SCD cohorts, we were unable to validate the association between the candidate SNP and diabetes in SCD. We reasoned that a risk genetic variation causing a progressive disease such as diabetes may exhibit certain effects on metabolism in early life. Among SCD patients, overweight individuals showed greater risk of diabetes at young age compared to non-overweight individuals, and their disease risk increased markedly when they turned 45 years and older (Figure 1C and 1D). Therefore, overweight was a relevant phenotype to diabetes in young SCD patients. We tested the association of rs59014890 and its tagged SNPs ($r^2 > 0.5$) with overweight in 181 patients aged 12-20 years from an independent SCD cohort (PUSH) (Bae et al. 2012). The risk allele of rs59014890 was significantly associated with overweight in these patients (OR=2.0, $P=0.048$, Supplemental Table 7), consistent with its effect on diabetes in adult cohorts. The association was stronger in mild SCD (13 cases out of 39, OR=7.5, $P=0.0065$) relative to severe SCD (20 cases out of 142, OR=1.06, $P=0.9$).

We further assessed the effect of rs59014890 in a case-control study of diabetes in African Americans without SCD (Palmer et al. 2012). The association between rs59014890 and

diabetes was not significant (OR=1.05, P=0.6) among all 1605 analyzed individuals, but it was significant among individuals with age ≥ 45 years and BMI >25 kg/m² (OR=1.39, P=0.037, n=592), exhibiting a gene by environment (defined by age and BMI stratification) interaction effect (P=0.031, n=1605) (Supplemental Table 8). The age and BMI dependency of the *APOB* genetic effect observed in the normal population was consistent with the age and BMI dependency of diabetes observed in SCD cohorts, although in the SCD cohorts we were unable to directly test the gene by environment effect for the *APOB* polymorphism because the diabetes cases were concentrated in older and overweight patients.

Cell transcriptional specificity of the candidate genes

The cell type in which a given gene exhibits intense transcriptional regulation, reflected as the density of active transcriptional factor binding sites, should indicate that the gene functions in the cell type. As shown above, the risk allele of rs59014890 decreased expression level of *APOB* in PBMCs. To determine the regulation specificity of the locus, we compared the enrichment of DNase hypersensitivity peaks derived from a variety of cell lines within the linkage disequilibrium (LD) block ($r^2 > 0.1$) of rs59014890. We found significant enrichment of DNase hypersensitivity peaks derived from hepatocytes compared to other cell types in the Encyclopedia of DNA Elements (ENCODE) data (Supplemental Table 9, Figure 2C), suggesting that this locus is important for transcription of *APOB* in hepatocytes, cells that play critical roles in lipid and glucose metabolism.

Low-density lipoproteins-cholesterol and diabetes in SCD

APOB encodes the primary apolipoprotein component of chylomicrons and low-density lipoproteins (LDLs). To investigate the potential effect of rs59014890, we examined association results from meta-analysis of plasma lipid traits for individuals of European descent (Teslovich et al. 2010). As SNP rs59014890 was not covered by that study we identified its tagged SNPs. These SNPs were associated with plasma concentration of LDL-cholesterol ($P < 1 \times 10^{-13}$), total cholesterol ($P < 1 \times 10^{-10}$), and triglycerides ($P < 0.03$) (Supplemental Table 10). The direction of these associations indicated that the alleles associated with elevated plasma LDL-cholesterol levels in normal individuals are those with greater risk of diabetes in SCD. The lead SNPs in the reported study, however, were rs1367117 for LDL-cholesterol ($P = 4 \times 10^{-114}$) and rs1042034 for triglyceride ($P = 1 \times 10^{-45}$). SNP rs59014890 was not in LD with either rs1367117 ($r^2 = 0.002$ in European samples and $r^2 = 0.002$ in African samples) or rs1042034 ($r^2 = 0.003$ in European samples and $r^2 = 0.0004$ in African samples) based on the 1000 genomes data, suggesting rs59014890 as an independent variant of the lead SNPs.

Non-fasting lipid measurements were available for 45 individuals in the UIC cohort. Among these patients, older age, higher BMI, and diabetes in general associated with higher levels of LDL-cholesterol, total cholesterol, and triglyceride. Severe SCD genotype associated with lower levels of LDL-cholesterol and total cholesterol, but higher levels of triglyceride (Supplemental Table 11). For 36 patients with lipid measurements and genotype data available, the association of rs59014890 with diabetes was only slightly affected by adjusting lipid levels (Supplemental Table 12).

We also evaluated candidate SNPs derived from previous GWAS on T2D (Morris et al. 2012) and lipid traits (Teslovich et al. 2010) for association with diabetes in SCD. Certain excess of significance was observed for T2D SNPs (Supplemental Table 13), the top one being SNP rs11651755 in *HNF1B* gene (Bonferroni adjusted $P=0.064$). Few significant associations were found for lipid SNPs (Supplemental Table 13).

Correlation of APOB PBMC gene expression with hemoglobin A1C, serum total cholesterol and triglyceride in Chuvash polycythemia

Chuvash polycythemia is caused by *VHL*^{R200W} homozygosity that leads to chronically up-regulated hypoxic responses and increased erythropoiesis at normoxia, similar to the physiological conditions of hemolysis-induced anemia in SCD except for the absence of anemia. Also similar to SCD, Chuvash polycythemia subjects with normal BMI have improved glucose tolerance (McClain et al. 2013). Our *a priori* hypothesis was that comparing genetic and transcriptional responses in SCD and Chuvash polycythemia would help identify novel pathways of disease risk. Measurements of hemoglobin A1C and serum concentrations of total cholesterol and triglyceride were available in our Chuvash polycythemia cohort (McClain et al. 2013). We tested the correlation of these variables with expression levels of *APOB* in PBMCs derived from the Chuvash polycythemia cohort. Chuvash polycythemia is often treated by phlebotomy resulting in great variation in serum ferritin concentration, which was shown to affect insulin sensitivity (Gabrielsen et al. 2012). Therefore, we regressed out the effects of age, gender and ferritin concentration from both gene expression and the metabolic measurements. Correlations between the residual expression and metabolic measurements were then tested. As shown in Figure 3, lower expression levels of *APOB* correlated with elevated levels of hemoglobin A1C ($P=0.031$, $n=17$), serum total cholesterol ($P=0.093$, $n=30$) and triglyceride ($P=0.0023$, $n=30$). Therefore, decreased PBMC expression of *APOB* was associated with a metabolic profile in Chuvash polycythemia that is consistent with the genetic effects of *APOB* on diabetes risk in SCD.

Discussion

Our study indicates that the prevalence of diabetes mellitus is low in SCD patients compared to the general population, but that the prevalence increases sharply with older age and elevated BMI. We found that the C allele of rs59014890, which decreased *APOB* expression in PBMCs, was associated with increased diabetes risk in SCD patients. The *APOB* polymorphism also associates with the risk of overweight in adolescent SCD patients, with diabetes risk in older, overweight African-Americans without SCD, and with increases in serum lipids in general populations. In Chuvash polycythemia patients who experience chronic hypoxia, increased erythropoiesis and improved glucose tolerance similar to SCD patients, lower *APOB* expression levels in PBMCs correlated with higher levels of hemoglobin A1C, serum total cholesterol and triglyceride. Limitations to our study are that the relatively small sample size only allowed detection of large genetic effects and that we were not able to find a SCD cohort with diabetes diagnosis to replicate our finding.

As SCD exhibited strong protection from diabetes, one explanation of the association of rs59014890 with diabetes is that its risk allele ameliorates SCD thus reducing the protection from diabetes. Using the degree of hemolysis as an indicator of SCD severity (Materials and Methods), however, we found no association between rs59014890 and SCD severity in the UIC and Walk-PHaSST cohorts ($\beta=-0.025$, combined $P=0.7$), suggesting that an ameliorating effect on SCD could not fully explain the SNP's strong effect on diabetes risk.

Endogenous lipids derived from hepatocytes are assembled in association with apoB to form very low-density lipoprotein (VLDL), which converts to LDL in the bloodstream. A recent cross-sectional study of patients with specific *APOB* mutations leading to familial hypercholesterolemia suggests a causal relationship between lipid metabolism and diabetes. These mutations inhibit lipid transmembrane transport and are protective from T2D, presumably through prevention of buildup of cholesterol in beta cells of the pancreas and prevention of related cytotoxic effects (Besseling et al. 2015). From this perspective, the rs59014890 locus may be primarily a lipid locus that is mildly associated with diabetes in older, overweight non-SCD African Americans due to dyslipidemia. The strong association of rs59014890 with diabetes is likely specific to SCD, as iron overload, oxidative stress, inflammation and hypoxia may exacerbate the toxicity of any increased beta cell cholesterol related to increased plasma lipid levels. Furthermore, diabetes outcome in SCD may depend on the balance between SCD-triggered genetic effects and an overall protective effect of SCD from diabetes.

APOB mRNA levels are tightly regulated across various metabolic conditions in cell cultures (Wang et al. 2003). In addition, it has been shown in a rat hepatoma cell line that apoB-48, one of two apoB protein isoforms, is preferentially translated over apoB-100 when endogenous apoB mRNA is increased (Collins et al. 2000), indicating a distinct influence of *APOB* transcription level on relative abundance of apoB protein isoforms. The risk allele of rs59014890 decreased *APOB* expression in PBMCs in SCD, while lower *APOB* expression in PBMCs correlated with elevated hemoglobin A1C and serum lipids in Chuvash polycythemia. If our observations in PBMCs apply to hepatocytes, the major cell type in which the polymorphism executes regulatory function, the risk allele of rs59014890 might decrease *APOB* transcription level in hepatocytes leading to increased abundance of apoB-100 relative to apoB-48. Alternatively, it is possible that the risk allele of rs59014890 might have a different effect on *APOB* gene expression in hepatocytes as opposed to PBMCs, because effects of *cis*-acting variations are highly dependent on specific cell types (Brown et al. 2013). Further studies are needed to investigate the molecular mechanism underlying the observed associations.

In contrast to previous GWAS in normal populations that uncovered many common loci with small effects for T2D, our meta-analysis in SCD identified a large effect locus. Independent cohorts are needed to validate this association with more robust estimation of its effect size and to uncover the causal variant through deep sequencing of the target region. Deleterious genetic variants can affect the fitness of carriers (Gibson 2011). In SCD patients, however, large effect loci contributing to risk of diabetes may exist owing to the natural history of the disease. In tropical areas with relative paucity of food and short life expectancy, low BMI and hyperactive metabolism associated with hypoxia and heightened

erythropoiesis in young individuals may suppress the expression of genetic risk factors. In contrast, affluent modern societies could permit genetic risk effects to be revealed and exacerbated with age due to obesity, chronic inflammation and blood transfusion-related iron overload. SCD might provide a unique model to uncover genetic variants influencing key metabolic pathways and their environmental specificity.

Materials and Methods

SCD cohorts

The UIC cohort is comprised of 174 Hb SS, 38 Hb SC, 19 Hb Sbeta⁺-thalassemia, 6 Hb Sbeta⁰-thalassemia, and 1 Hb SO^{arab} (a severe form of SCD) patients. The Walk-PHaSST is comprised of 469 Hb SS, 114 Hb SC, 24 Hb Sbeta⁺-thalassemia and 11 Hb Sbeta⁰-thalassemia patients. The PUSH adolescent cohort is comprised of 139 Hb SS, 34 Hb SC, 5 Hb Sbeta⁺-thalassemia and 3 Hb Sbeta⁰-thalassemia patients with age ranging between 12-20 years old. The Howard cohort is comprised of 26 Hb SS, 7 Hb SC, and 2 Hb Sbeta⁺-thalassemia patients.

Comparative NHANES data and prevalence of overweight and diabetes

Prevalence of overweight and diabetes were assessed in 856 individuals between 18-85 years old, pooled from the UIC and the Walk-PHaSST cohorts. National Health and Nutrition Examination Survey 2009-2010 and 2011-2012 data were combined, resulting in 2579 general population African Americans between 18-85 years old and with unambiguous T2D diagnosis. Fisher's exact tests were applied for comparisons.

Genotype data processing and imputation

Genomic DNA isolated from peripheral blood mononuclear cells was labeled and hybridized to the Affymetrix Axiom genome-wide Pan-African array (UIC cohort), the Illumina Human 610-Quad SNP array (Walk-PHaSST cohort and PUSH cohort), and the Affymetrix Genome-Wide Human SNP Array 6.0 (Howard cohort). All samples had genotype call rate >95%. SNPs deviating from Hardy Weinberg equilibrium ($P < 0.0001$) or with a minor allele frequency (MAF) <0.01 were removed. Population outliers were identified based on principal components analysis of the genotype data (Yang et al. 2011). Proportion of identity-by-descent was calculated pair-wise using PLINK (Purcell et al. 2007) to identify related individuals and potentially contaminated DNA samples. We imputed SNPs genotyped in UIC, Walk-PHaSST, PUSH adolescent, and Howard cohorts to 1000 genomes project phase 1 data, with European and African reference panels, using Beagle (Browning and Browning 2007) version 4. SNPs with imputation quality $r^2 \geq 0.5$ or MAF ≥ 0.1 within each cohort were excluded from association analysis.

Meta-analysis

Patients <18 years old were excluded from meta-analysis of the UIC and Walk-PHaSST cohorts. A logistic linear regression model was applied adjusting for age, gender, *HBB* genotype, hydroxyurea treatment, and BMI. The proportion of European ancestry using the first principal component of genotype data was also adjusted, since it was the primary source of population structure. Firth's bias reduction method (Firth 1993) was used for parameter

estimation to account for the small number of diabetic individuals. P values were estimated using χ^2 test with 1 degree of freedom. The combined P values were estimated using a Z-score approach (Willer et al. 2010). Tagged SNPs of the candidate SNP were defined based on LD estimated by pair-wise r^2 in the Walk-PHaSST cohort.

Genetic association with SCD severity

Principal components analysis was carried out on serum concentrations of lactate dehydrogenase, total bilirubin, and aspartate aminotransferase for 183 UIC and 428 Walk-PHaSST patients. Degree of hemolysis was defined as the first principal component. Hemolysis was regressed on SNP dosage of rs59014890, adjusting for age, gender, *HBB* genotype, and hydroxyurea treatment. Inverse variance meta-analysis was carried out to combine the association results.

Genetic association with overweight in PUSH SCD adolescent cohort

rs59014890 and its tagged SNPs ($r^2 > 0.5$) were tested for association with overweight in the PUSH adolescent cohort. We used the 2000 Centers for Disease Control and Prevention Growth Charts for the United States as reference to identify overweight thresholds (85% BMI quantiles) according to age and gender. Logistic regression was applied adjusting for age, gender, *HBB* genotype, hydroxyurea treatment, and European ancestry based on principal components analysis of the genotype data.

Genetic association with diabetes in a case-control study of African Americans without SCD

The individual level genotype data and phenotype data were obtained from dbGaP with study accession phs000140.v1.p1. SNP information was obtained from Affymetrix Genome-Wide Human SNP Array 6.0 annotation release 34. Quality control and preprocessing of genotype data was as described above. SNPs 1Mb away from rs59014890 were imputed to 1000 genomes phase 1 data using European and African reference panels. Individuals <18 years old and phenotype outliers, defined as those with BMI four standard deviations away from mean, were excluded from analysis. To test for SNP main effect, a logistic linear regression model was applied adjusting for age, gender, BMI, and population stratification (top three principal components of genotype data). To test for SNP by environment (age 45 years old and BMI >25 kg/m²) interaction, the logistic linear regression model included effects of SNP dosage, environment, SNP by environment interaction, gender, and population stratification. P values were estimated using χ^2 test with 1 degree of freedom.

Gene expression datasets

Message RNA was purified from PBMCs, labeled and hybridized to Affymetrix Exon 1.0 ST Array for 35 patients of Howard cohort (Zhang et al. 2014a), and to Affymetrix Gene 2.0 ST Array for 134 SCD patients selected from UIC cohort. Probe sequences were aligned to human genome assembly GRCh37 to select for probes with unique perfect alignment. Probes that interrogate multiple gene transcripts and that contain SNPs with 1% minor allele frequency in dbSNP dataset were removed. Probe intensities were log₂ transformed, background corrected and quantile normalized. Probe intensity was subtracted by the

corresponding probe mean across samples. Gene expression level was summarized as mean intensity across probes within gene, using Gencode version 19. Batch effects of RNA labeling and array hybridization were adjusted using an empirical Bayes method (Johnson et al. 2007).

Correlation between gene expression and rs59014890

Expression levels of seven genes located within $\pm 1\text{Mb}$ of rs59014890 were extracted from the expression data from the UIC cohort and the Howard cohort. Within each dataset, gene expression levels were regressed on imputed dosage of rs59014890, adjusting for age, gender, *HBB* genotype, hydroxyurea treatment and ancestral admixture. Inverse variance meta-analysis was carried out to combine the association results.

Correlation between APOB gene expression and metabolic measurements in Chuvash polycythemia

Gene expression data from Chuvash polycythemia were processed as described (Zhang et al. 2014b) and the metabolic measurements were obtained from a previous study (McClain et al. 2013). Serum triglyceride concentrations were square-root transformed to approximate normal distribution. Correlation between the residual gene expression levels and residual metabolic measurements, after regressing out age, gender and log (serum ferritin concentration) effects, were assessed using Pearson's one-sided correlation test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported in part by grants R01 HL079912-04, 2 R25-HL03679-08, and 1P30HL107253 (V.R.G.); KL2TR000048 (S.L.S.); P50HL118006 (M.N.); R01HL111656 and K23HL098454 (R.F.M.).

References

- Akhoue SA, Shankar S, Milne GL, Morrow J, Chen KY, Ajayi WU, Buchowski MS. Energy expenditure, inflammation, and oxidative stress in steady-state adolescents with sickle cell anemia. *Pediatr Res.* 2007; 61:233–8.10.1203/pdr.0b013e31802d7754 [PubMed: 17237728]
- Bae HT, Baldwin CT, Sebastiani P, Telen MJ, Ashley-Koch A, Garrett M, Hooper WC, Bean CJ, Debaun MR, Arking DE, Bhatnagar P, Casella JF, Keefer JR, Barron-Casella E, Gordeuk V, Kato GJ, Minniti C, Taylor J, Campbell A, Luchtman-Jones L, Hoppe C, Gladwin MT, Zhang Y, Steinberg MH. Meta-analysis of 2040 sickle cell anemia patients: BCL11A and HBS1L-MYB are the major modifiers of HbF in African Americans. *Blood.* 2012; 120:1961–2.10.1182/blood-2012-06-432849 [PubMed: 22936743]
- Barden EM, Kawchak DA, Ohene-Frempong K, Stallings VA, Zemel BS. Body composition in children with sickle cell disease. *Am J Clin Nutr.* 2002; 76:218–25. [PubMed: 12081838]
- Barden EM, Zemel BS, Kawchak DA, Goran MI, Ohene-Frempong K, Stallings VA. Total and resting energy expenditure in children with sickle cell disease. *J Pediatr.* 2000; 136:73–9. [PubMed: 10636978]
- Besseling J, Kastelein JJ, Defesche JC, Hutten BA, Hovingh GK. Association between familial hypercholesterolemia and prevalence of type 2 diabetes mellitus. *JAMA.* 2015; 313:1029–36.10.1001/jama.2015.1206 [PubMed: 25756439]

- Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* 2013; 9:e1003649.10.1371/journal.pgen.1003649 [PubMed: 23935528]
- Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* 2007; 81:1084–97.10.1086/521987 [PubMed: 17924348]
- Collins HL, Sparks CE, Sparks JD. B48 is preferentially translated over B100 in cells with increased endogenous apo B mRNA. *Biochem Biophys Res Commun.* 2000; 273:1156–60.10.1006/bbrc.2000.3074 [PubMed: 10891388]
- Elmariiah H, Garrett ME, De Castro LM, Jonassaint JC, Ataga KI, Eckman JR, Ashley-Koch AE, Telen MJ. Factors associated with survival in a contemporary adult sickle cell disease cohort. *Am J Hematol.* 2014; 89:530–5.10.1002/ajh.23683 [PubMed: 24478166]
- Firth D. Bias reduction of maximum likelihood estimates. *Biometrika.* 1993; 80:27–38.
- Florez JC, Hirschhorn J, Altshuler D. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet.* 2003; 4:257–91.10.1146/annurev.genom.4.070802.110436 [PubMed: 14527304]
- Gabrielsen JS, Gao Y, Simcox JA, Huang J, Thorup D, Jones D, Cooksey RC, Gabrielsen D, Adams TD, Hunt SC, Hopkins PN, Cefalu WT, McClain DA. Adipocyte iron regulates adiponectin and insulin sensitivity. *J Clin Invest.* 2012; 122:3529–40.10.1172/JCI44421 [PubMed: 22996660]
- Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet.* 2011; 13:135–45.10.1038/nrg3118 [PubMed: 22251874]
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007; 8:118–27.10.1093/biostatistics/kxj037 [PubMed: 16632515]
- Lyssenko V, Laakso M. Genetic screening for the risk of type 2 diabetes: worthless or valuable? *Diabetes Care.* 2013; 36(Suppl 2):S120–6.10.2337/dcS13-2009 [PubMed: 23882036]
- Machado RF, Barst RJ, Yovetich NA, Hassell KL, Kato GJ, Gordeuk VR, Gibbs JS, Little JA, Schraufnagel DE, Krishnamurti L, Girgis RE, Morris CR, Rosenzweig EB, Badesch DB, Lanzkron S, Onyekwere O, Castro OL, Sachdev V, Wacławski MA, Woolson R, Goldsmith JC, Gladwin MT. walk PI, Patients. Hospitalization for pain in patients with sickle cell disease treated with sildenafil for elevated TRV and low exercise capacity. *Blood.* 2011; 118:855–64.10.1182/blood-2010-09-306167 [PubMed: 21527519]
- McClain DA, Abuelgasim KA, Nouraie M, Salomon-Andonie J, Niu X, Miasnikova G, Polyakova LA, Sergueeva A, Okhotin DJ, Cherqaoui R, Okhotin D, Cox JE, Swierczek S, Song J, Simon MC, Huang J, Simcox JA, Yoon D, Prchal JT, Gordeuk VR. Decreased serum glucose and glycosylated hemoglobin levels in patients with Chuvash polycythemia: a role for HIF in glucose metabolism. *J Mol Med (Berl).* 2013; 91:59–67.10.1007/s00109-012-0961-5 [PubMed: 23015148]
- Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012; 44:981–90.10.1038/ng.2383 [PubMed: 22885922]
- Morrison JC, Schneider JM, Kraus AP, Kitabchi AE. The prevalence of diabetes mellitus in sickle cell hemoglobinopathies. *J Clin Endocrinol Metab.* 1979; 48:192–5.10.1210/jcem-48-2-192 [PubMed: 429473]

- Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, An SS, Hester JM, Cooke JN, Bostrom MA, Rudock ME, Talbert ME, Lewis JP, Consortium D, Investigators M. Ferrara A, Lu L, Ziegler JT, Sale MM, Divers J, Shriner D, Adeyemo A, Rotimi CN, Ng MC, Langefeld CD, Freedman BI, Bowden DW, Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bostrom KB, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieveverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, et al. A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS One*. 2012; 7:e29202.10.1371/journal.pone.0029202 [PubMed: 22238593]
- Pauling L, Itano HA, et al. Sick cell anemia a molecular disease. *Science*. 1949; 110:543–8. [PubMed: 15395398]
- Permutt MA, Wasson J, Cox N. Genetic epidemiology of diabetes. *J Clin Invest*. 2005; 115:1431–9.10.1172/JCI24758 [PubMed: 15931378]
- Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, Klug PP. Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med*. 1994; 330:1639–44.10.1056/NEJM199406093302303 [PubMed: 7993409]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–75.10.1086/519795 [PubMed: 17701901]
- Rees DC, Williams TN, Gladwin MT. Sick cell disease. *Lancet*. 2010; 376:2018–31.10.1016/S0140-6736(10)61029-X [PubMed: 21131035]
- Saraf SL, Zhang X, Kanas T, Lash JP, Molokie RE, Oza B, Lai C, Rowe JH, Gowhari M, Hassan J, Desimone J, Machado RF, Gladwin MT, Little JA, Gordeuk VR. Haemoglobinuria is associated with chronic kidney disease and its progression in patients with sickle cell anaemia. *Br J Haematol*. 2014; 164:729–39.10.1111/bjh.12690 [PubMed: 24329963]
- Simcox JA, McClain DA. Iron and diabetes risk. *Cell Metab*. 2013; 17:329–41.10.1016/j.cmet.2013.02.007 [PubMed: 23473030]
- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemssen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010; 466:707–13.10.1038/nature09270 [PubMed: 20686565]
- Wang AB, Liu DP, Liang CC. Regulation of human apolipoprotein B gene expression at multiple levels. *Exp Cell Res*. 2003; 290:1–12. [PubMed: 14516783]
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–1.10.1093/bioinformatics/btq340 [PubMed: 20616382]
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011; 88:76–82.10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]

- Zhang X, Zhang W, Ma SF, Desai AA, Saraf S, Miasniakova G, Sergueeva A, Ammosova T, Xu M, Nekhai S, Abbasi T, Casanova NG, Steinberg MH, Baldwin CT, Sebastiani P, Prchal JT, Kittles R, Garcia JG, Machado RF, Gordeuk VR. Hypoxic response contributes to altered gene expression and precapillary pulmonary hypertension in patients with sickle cell disease. *Circulation*. 2014a; 129:1650–8.10.1161/CIRCULATIONAHA.113.005296 [PubMed: 24515990]
- Zhang X, Zhang W, Ma SF, Miasniakova G, Sergueeva A, Ammosova T, Xu M, Nekhai S, Nourai M, Wade MS, Prchal JT, Garcia JG, Machado RF, Gordeuk VR. Iron deficiency modifies gene expression variation induced by augmented hypoxia sensing. *Blood Cells Mol Dis*. 2014b; 52:35–45.10.1016/j.bcmd.2013.07.016 [PubMed: 23993337]

Figure 1.

Age and BMI dependency of diabetes prevalence in SCD. Percent overweight is lower in SCD than in control individuals in (A) younger (34% vs 69%, $P < 1 \times 10^{-16}$) and (B) older (54% vs 78%, $P = 1.8 \times 10^{-15}$) individuals. Percent overweight increased with age in SCD (34% vs 54%, $P = 4.3 \times 10^{-8}$) and in control individuals (69% vs 78%, $P = 2.8 \times 10^{-8}$). (C) Percent diabetes increased with overweight in younger SCD subjects (0.25% in non-overweight vs 1.5% in overweight, $P = 0.11$) and in control individuals (1.4% in non-overweight vs 4.5% in overweight, $P = 0.0057$). (D) Percent diabetes increased sharply with overweight in older SCD individuals (1.7% in non-overweight vs 17% in overweight, $P = 1.5 \times 10^{-5}$), whereas it barely doubled in control individuals (16% in non-overweight vs 29% in overweight, $P = 4.4 \times 10^{-6}$).

Figure 2. Genetic association of diabetes in SCD. (A) Manhattan plot of GWAS scan. The dashed horizontal line represents the significance threshold. (B) Association between the C allele of rs59014890 and *APOB* gene expression in UIC (black points) and Howard (red points) cohorts. Residual expression and residual allele dosage values, after adjusting for age, gender, *HBB* genotype, hydroxyurea treatment and population stratification, are plotted. (C) Heatmap of $-\log_{10}P$ from pair-wise comparison for enrichment of DNase hypersensitivity peaks within LD block of rs59014890. astrocyte_C: astrocytes-cerebellar; astrocyte_H: astrocytes-hippocampal; hepatocyte: primary hepatocytes; pancreatic duct: pancreatic duct cells immortalized with E6E7 gene of HPV; pancreatic islet: dedifferentiated human pancreatic islets; stellate: hepatic stellate cells; skeletal muscle: skeletal muscle cells.

Figure 3.

Correlation between *APOB* gene expression in PBMCs with percent hemoglobin A1C (A), serum total cholesterol (B) and triglyceride (C) concentrations in Chuvash polycythemia. Residual plots of metabolite measurements against gene expression levels are shown, with r and P from Pearson's one-sided correlation test.

Table 1

The association of rs59014890 with *APOB* gene expression and diabetes risk in the UIC and Walk-PHaSST cohorts. r^2 : imputation quality r^2 ; RAF: risk allele frequency.

SNP	chr	position	risk allele	APOB expression		diabetes										
				beta	P	UIC					Walk					P combined
						r ²	RAF	OR	P	r ²	RAF	OR	P			
rs59014890	2	21509092	C	-0.023	0.0082	1	0.24	17	4.7×10 ⁻⁴	0.89	0.23	18	1.6×10 ⁻⁵	3.2×10 ⁻⁸		