



Genomic architecture of sickle cell disease in West African children

Jacklyn Quinlan^{1,2†}, Youssef Idaghdour^{2,3†}, Jean-Philippe Goulet², Elias Gbeha², Thibault de Malliard², Vanessa Bruat², Jean-Christophe Grenier², Selma Gomez⁴, Ambaliou Sanni⁴, Mohamed C. Rahimy^{5*} and Philip Awadalla^{2*}

¹ Department of Social and Preventive Medicine, Faculty of Medicine, School of Public Health, University of Montreal, Montreal, QC, Canada

² Department of Pediatrics, Faculty of Medicine, Sainte-Justine Research Center, University of Montreal, Montreal, QC, Canada

³ Biology, Division of Science and Mathematics, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

⁴ Laboratoire de Biochimie et Biologie Moléculaire, Faculté des Sciences et Techniques, Université d'Abomey-Calavi, Cotonou, Benin

⁵ Faculté des Sciences de la Santé, Centre de Prise en charge Médicale Intégrée du Nourrisson et de la Femme Enceinte atteints de Drépanocytose, Université d'Abomey-Calavi, Cotonou, Benin

Edited by:

Jason Wolf, University of Bath, UK

Reviewed by:

Alison Motsinger-Reif, North Carolina State University, USA
Marcin Kierczak, Swedish University of Agricultural Sciences, Sweden
Ricardo A. Verdugo, University of Chile, Chile

*Correspondence:

Mohamed C. Rahimy, Faculté des Sciences de la Santé, Centre de Prise en charge Médicale Intégrée du Nourrisson et de la Femme Enceinte atteints de Drépanocytose, Université d'Abomey-Calavi, PO Box 2640, Cotonou, Benin
e-mail: mrahimy@bj.refer.org;
Philip Awadalla, Department of Pediatrics, Faculty of Medicine, Sainte-Justine Research Center, Philip Awadalla, 3175 Cote Sainte-Catherine, B-456 Pavillon Vidéotron, Montreal, QC H3T 1C5, Canada
e-mail: philip.awadalla@umontreal.ca

† These authors have contributed

equally to this work.

Sickle cell disease (SCD) is a congenital blood disease, affecting predominantly children from sub-Saharan Africa, but also populations world-wide. Although the causal mutation of SCD is known, the sources of clinical variability of SCD remain poorly understood, with only a few highly heritable traits associated with SCD having been identified. Phenotypic heterogeneity in the clinical expression of SCD is problematic for follow-up (FU), management, and treatment of patients. Here we used the joint analysis of gene expression and whole genome genotyping data to identify the genetic regulatory effects contributing to gene expression variation among groups of patients exhibiting clinical variability, as well as unaffected siblings, in Benin, West Africa. We characterized and replicated patterns of whole blood gene expression variation within and between SCD patients at entry to clinic, as well as in follow-up programs. We present a global map of genes involved in the disease through analysis of whole blood sampled from the cohort. Genome-wide association mapping of gene expression revealed 390 peak genome-wide significant expression SNPs (eSNPs) and 6 significant eSNP-by-clinical status interaction effects. The strong modulation of the transcriptome implicates pathways affecting core circulating cell functions and shows how genotypic regulatory variation likely contributes to the clinical variation observed in SCD.

Keywords: sickle cell disease, genomics, transcriptome, eSNP mapping, gene-by-environment interactions

INTRODUCTION

Sickle cell disease (SCD) is an autosomal recessive genetic disorder particularly common among individuals of Sub-Saharan African ancestry, affecting 1 in 100 West African individuals and 1 in 500 African-Americans (World Health Organization, 2006). Genetic mutations that cause SCD result in structural changes to wild-type hemoglobin (HbAA), the oxygen carrying protein inside red blood cells (RBCs). The most common form of SCD in West Africa is caused by a single point mutation in codon 6 of the β -globin gene which leads to an amino acid substitution of glutamic acid to valine (HbSS). The second most common abnormal Hb mutation in West Africa, HbC, results in an amino acid change at the same position in the beta globin gene, but with lysine replacing glutamic acid. These hemoglobin mutations compromise the delivery of oxygen and result in tissue and organ damage. Despite the monogenic origin of the disease, SCD patients exhibit a broad spectrum of clinical variation (Driss et al., 2009) ranging from patients with mild forms of the disease

that rarely require medical interventions to patients with severe complications warranting frequent hospitalization and aggressive clinical follow-up. Homozygous HbSS and compound heterozygous HbSC individuals suffer from SCD with overlapping yet distinctive clinical and biochemical features (Hannemann et al., 2011). Inter-individual clinical variation is also pervasive within each of these SCD groups, but its basis is poorly understood and likely reflects a combination of the effects of several factors including haplotypic variation in the β -globin locus region, the action of genetic modifiers elsewhere in the genome, and a wide range of environmental factors (Weatherall, 2001; Sankaran et al., 2010).

Mapping genetic variants associated with SCD clinical phenotypes have largely been limited to candidate gene approaches and genome-wide association studies (Adams et al., 2003; Menzel et al., 2007; Lettre et al., 2008; Sebastiani et al., 2010; Solovieff et al., 2010; Thein, 2011; Steinberg and Sebastiani, 2012). One of the most characterized modulators of clinical expression of SCD is fetal hemoglobin (HbF). Higher HbF levels have been

associated with reduced rates of acute pain episodes, leg ulcers, less frequent acute chest syndromes, and reduced disease severity (Akinsheye et al., 2011). Heritability for HbF is high (h^2 approximately 0.60–0.89) and genetic factors that control its expression have been mapped (Sankaran et al., 2010). However, HbF levels have no clear association with other SCD clinical manifestations such as stroke and silent cerebral infarction, priapism, urine albumin excretion, and systemic blood pressure (Akinsheye et al., 2011). Several other genetic modifiers contributing to the variation in clinical expression of SCD have also been identified (Thein, 2011; Steinberg and Sebastiani, 2012), nonetheless most of the variation remains unexplained. Since disease in general involves differential expression (Emilsson et al., 2008; Cookson et al., 2009; Berry et al., 2010; Idaghdour et al., 2012), a systems genetics approach to map genetic variation associated with gene expression traits correlated with clinical phenotypes (Idaghdour and Awadalla, 2012) is likely to reveal regulatory variation modulating SCD.

We recruited HbSS and HbSC patients from a West African SCD cohort, using a two-phase sampling design, and generated a global map of gene expression variation and its genetic regulatory variation. Patients in the cohort were part of an established comprehensive clinical care program which includes an intensive socio-medical intervention program to impact the disease course (Rahimy et al., 2003). The vast majority of the children in the cohort show a severe clinical phenotype at some stage of the follow-up program. We set out to establish the extent of the effects of SCD relating to the Hb genotype and clinical follow-up on whole blood gene expression profiles, and also to identify genetic regulatory variation associated with gene expression traits. Furthermore, we test the hypothesis that gene expression variation associated with the disease or in response to clinical follow-up can be dependent on patient's regulatory genotypes, which in turn may explain inter-individual differences in disease severity. In so doing, we captured new genes associated with SCD clinical variation and identified genetic regulatory effects that explain a substantial percentage of transcriptional variation in SCD patients.

MATERIALS AND METHODS

STUDY POPULATION

Ethics approval for the study was granted by the Sainte-Justine Research Center Ethics Committee and by the Faculté des Sciences de la Santé of the University of Abomey-Calavi in Benin, West Africa. Informed consent was obtained for all participants in the study. Patients were part of a large cohort of SCD children longitudinally followed-up at the Centre de Prise en charge Médicale Intégrée du Nourrisson et de la Femme Enceinte atteints de Drépanocytose (CPMI-NFED), the National Institute of SCD Infants and Pregnant Women in Cotonou, The Republic of Benin. In total, 250 SCD patients aged between 6 months and 9 years old (mean age 4 years) and 61 healthy control siblings (not HbSS or HbSC and have at least one normal hemoglobin allele) were sampled under informed consent. A two-phase sampling design was used from February–December 2010 (Figure S1). The initial discovery phase included patients recruited mostly before the end of April, 2010, and the replication phase included patients that were recruited mostly between April 2010 and December 2010 (Figure

S2). The distribution of SCD patients newly enrolled (E) and followed (FU), Hb genotypes, and sex were proportionate in both phases. The 61 healthy siblings were of roughly equal age (mean age 3 years) and sex proportions as the SCD patients, and were also recruited at the CPMI-NFED (Figure S2).

SCD CLINICAL STATUS AND SEVERITY SCORE

The CPMI-NFED has an established comprehensive clinical care program that includes an intensive socio-medical intervention program to impact the disease course (Rahimy et al., 2003). The vast majority of children in the cohort show a severe clinical phenotype. SCD patients experiencing an acute event are labeled acute (A). For the purpose of the present study, two (2) sampling clinical categories were assigned to patients: patients sampled at enrolment into the program and in steady-state are labeled as entry (E), and patients already being followed at the SCD Center were labeled as FU. At the Center, most patients that are followed obtain a steady-state condition with general clinical improvement that involves increased velocity of linear physical growth and marked reduction in the frequency and severity of SCD-related acute events; however, some followed patients experience no such improvement.

Age-matched healthy siblings were recruited as controls (Ctls). Three quarters of our Ctls are heterozygous HbAS and 1/4 are homozygous HbAA. Only 14 probes were differentially expressed between HbAA and HbAS individuals at FDR 1%. Furthermore, none of the variance in the Ctls was explained by this effect as evidenced by variance component analysis and by the lack of clustering based on Hb genotype in PCA analysis (Figure S2). For these reasons, we grouped HbAA and HbAS individuals and used them as a control sample.

A quantitative SCD severity score (SV) was calculated using an online SCD severity calculator (<http://www.bu.edu/sicklecell/projects/>) (Sebastiani et al., 2007) where each patient was assigned a score based on their sex, Hb genotype, mean corpuscular volume (MCV), and white blood cell (WBC) counts. Ctls were assigned a score of 0.

SAMPLE PREPARATION

The same collection procedure was followed for all samples in order to reduce technical heterogeneity. A total of 10 ml of peripheral whole blood was collected for each patient between 9:00 am and 2:00 pm and stored at -30°C . Shipment to Montreal was done at -20°C . Approximately 3 ml of this blood was collected for RNA work in TEMPUS blood RNA Tubes (Life Technologies); and approximately 5 ml of this blood was collected in EDTA tubes for DNA work; the remainder of the blood was used for complete blood counts using an automated KX-21 blood analyzer (Sysmex Corporation, Japan), identification of the hemoglobin phenotype by high-performance liquid chromatography (HPLC) and Capillary Electrophoresis, and thick smear analysis for parasitemia quantification. Total RNA was isolated using the TEMPUS RNA extraction kit (Life Technologies) following the manufacturer's recommendations. A globin mRNA reduction step was performed using GLOBINclear-Human kit (Life Technologies). Total RNA extractions were quantified and quality was checked using the RNA 6000 Nano LabChip kit and 2100 Bioanalyzer (Agilent Technologies). Only samples of

high RNA quality (Agilent's RNA Integrity Number >7.5) were retained for expression profiling. DNA samples were extracted using QIAamp DNA Kit (Qiagen). Quantity and quality was checked using Agilent's DNA 6000 Nano LabChip kit and the 2100 Bioanalyzer (Agilent Technologies).

GENOTYPING THE β -GLOBIN LOCUS

Identification of the rs334 genotype and characterization of haplotype structure in the Hb locus was performed using Sequenom MassARRAY technology for 237 patients using 600 ng of genomic DNA and following the manufacturer's recommended protocols. Individuals with less than 75% call rate were excluded (Figure S2).

DATA SETS

Four datasets were used for gene expression analyses: the discovery phase (126 SCD patients and 31 Ctls), the replication phase (124 SCD patients and 30 Ctls), combined dataset I (250 SCD patients and 61 Ctls), and combined data set II (160 SCD patients and 56 Ctls). Combined data set II includes only HbSS SCD patients excluding those sampled during acute crises. For the joint genotypic and gene expression data analyses, the combined dataset II ($n = 173$) included 120 SCD patients and 53 Ctls (Figure S1, Supplementary File 3).

GENE EXPRESSION PROFILING

Illumina's HumanHT-12 v4 BeadArrays were used to generate expression profiles of more than 48,000 probes using 500 ng of labeled cRNA for each sample following the manufacturer's recommended protocols. All expression data are available at NCBI Gene Expression Omnibus (GEO) under the series number GSE35007. The individual expression arrays are listed as GSM860207 through GSM860517. To minimize chip and batch effects, a randomized design was used. Hybridization was performed on two different dates and 4 samples from the first hybridization batch were re-hybridized with the second batch. These technical replicates clustered adjacent to one another in hierarchical analysis, indicating a negligible batch effect on the data. This was confirmed by testing for batch effect in the probe-by-probe analysis of variance. The expression intensities were averaged for each probe in the statistical analysis. The raw intensities were extracted using the Gene Expression Module in Illumina's BeadStudio software. Expression intensities were log₂ transformed and quantile normalized using JMP Genomics v5.0 (SAS) after an outlier filtering procedure was applied. In total, 28,595 probes with expression at or above background levels averaged across all the arrays were retained for further analyses. These represent probes remaining after removal of 18,404 probe measurements that were considered to lay below background detection levels indicated by the inflection point in a plot of rank-ordered normalized intensities. Also, 427 probes overlaying SNPs included in the Illumina's OmniExpress BeadChip were removed from the analysis. Pathway and gene ontology analysis was performed using Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005).

GENOME WIDE GENOTYPING

Genome-wide genotyping data was generated for over 733,200 SNPs using Illumina's HumanOmni Express BeadChip arrays

following manufacturer's protocols and extracted using the Genotyping Module in Illumina's BeadStudio software. Marker properties were calculated using PLINK (Purcell et al., 2007). Only SNPs with minor allelic frequency >5%, a call rate >99% and SNPs that are in Hardy-Weinberg Equilibrium (HWE) were included (p -value > 0.001). This resulted in a final set of 568,921 SNPs for further analysis. Global genotypic variation and ancestry was inferred using Eigenstrat (Price et al., 2006) and STRUCTURE (Pritchard et al., 2000); we detected limited population structure in our sample (Figures S2, S3). Insignificant population structure and limited genetic differentiation were also observed when 541 genotypes from a subset of genes known to influence hemoglobin levels (alpha-globin, G6PD, BCL11A, MYB, and HBS1L) were used in PCA and gene-wise F_{st} analysis to estimate the magnitude of genetic differentiation among the clinical statuses investigated and between them and the Ctls.

GENE EXPRESSION DATA ANALYSIS

All statistical analyses of the gene expression data were performed using JMP Genomics v5.0 (SAS), and SAS 9.3 (SAS). Principal Component analysis (PCA) and Variance Component analysis (VCA) of the gene expression data were performed such that the first three expression PCs (ePCs) were modeled either simultaneously or individually as a function of various effects in the data: Hemoglobin genotype, clinical status (E vs. FU vs. Ctls), sex, and pair-wise combination of fixed effects. SAS GLM was used to evaluate the magnitude and significance of differentially expressed probes. Probe-level differential expression analysis was performed using analysis of covariance. Variance was partitioned among the Hemoglobin genotype (Hb), clinical status effect, sex, and total blood cell counts (RBCs and WBCs) as covariates. The effects of date of sampling, phase (discovery vs. replication), age (in years), and gPCs were tested and found to be marginal. Pairwise contrasts (Hb genotype \times Sex, Hb genotype \times ClinStatus, and ClinStatus \times Sex) also were evaluated and found to be insignificant. Results from the following full ANCOVA model are detailed in **Figure 2**: Expression = $\mu + \text{Hb genotype} + \text{ClinStatus} + \text{Sex} + \text{WBC} + \text{RBC} + \epsilon$.

The error ϵ was assumed to be normally distributed with mean equal to zero. The 3-way clinical status effect (E vs. FU vs. Ctls) was evaluated. A statistical significance threshold of 1% FDR was applied separately to each term in the analysis of covariance.

eSNP MAPPING

Multiple linear regression analyses were performed using PLINK to test for significant associations between gene expression levels for each probe and SNP genotype. Only well-annotated, autosomal probes with validated chromosomal location and gene function based on the most recent annotation in NCBI and UCSC as of October 2011 were included for the association tests. In the process we aligned all probes to the reference genome (hg19), excluded ambiguous and all non-RefSeq probes, and removed 427 probes overlaying known SNPs from the analysis. This resulted in a total of 19,431 expressed probes that were tested for association with 560,675 SNPs. SNPs with a minor allelic frequency <5%, an exact HWE P -value < 0.001, or >1% missing data were excluded. We distinguished between local and distal associations

based on the chromosomal location of the probe-SNP pair; a local association implicates a probe and a SNP located on the same chromosome while a distal association implicates a probe and a SNP located on different chromosomes. We applied Bonferroni correction for all eSNP associations in this study by accounting for both the number of SNPs and loci tested. Since 560,675 SNPs were tested for association with 19,431 probes, a genome-wide Bonferroni threshold for distal-associations corresponds to $0.05/(19,431 \text{ probes} \times 560,675 \text{ SNP}) = 4.59 \times 10^{-12}$ and for local associations to a Bonferroni threshold of $0.05/(19,431 \text{ probes} \times 200 \text{ SNPs}) = 1.28 \times 10^{-08}$ considering an average number of 200 SNPs tested against each probe.

In the multiple linear regression eSNP analysis, we tested for association between probe expression levels and SNP genotype while accounting for clinical status effect (ClinStatus), sex and blood cell counts (white and RBC counts, WBC and RBC), assuming that the error ϵ is normally distributed with a mean of zero, where:

$$\text{Model 1 : Expression} = \mu + \text{SNP} + \text{ClinStatus} + \text{WBC} \\ + \text{RBC} + \text{Sex} + \epsilon$$

The significant associations were compared to the associations reported in 12 published eQTL studies of peripheral blood or its derivatives at nominal P -values $< 10^{-7}$. These published associations were accessed using the eQTL Browser (<http://eqtl.uchicago.edu/cgi-bin/gbrowser/eqtl/>) and compared to our results.

INTERACTION EFFECTS

We tested for SNP-by-clinical status interaction effects using 7002 probes differentially expressed for the 3-way clinical status effect (E vs. FU vs. Ctls, FDR 1%) using combined data set II. For this analysis and to reduce the effect of outlier expression values, we further filtered the set of genotypes and included only SNPs with a minor allelic frequency $< 5\%$, an exact HWE test P -value < 0.001 and $> 1\%$ missing data calculated in each of the sub-groups of patients separately (455,750 SNPs). This resulted in a final set of 455,750 SNPs tested against 7002 probes while accounting for clinical status effect (ClinStatus), sex, cell counts (white and RBCs, WBC and RBC) and including a term for $\text{SNP} \times \text{ClinStatus}$ in the model:

$$\text{Model 2 : Expression} = \mu + \text{SNP} + \text{ClinStatus} + \text{WBC} + \text{RBC} \\ + \text{Sex} + \text{SNP} \times \text{ClinStatus} + \epsilon$$

where ϵ is assumed to be normally distributed with a mean of zero.

A genome-wide Bonferroni correction was applied by accounting for both the number of SNPs and loci considered in this analysis. Since 455,750 SNPs were tested for association with 7002 probes, a genome-wide Bonferroni threshold for distal-associations corresponds to $0.05/(7002 \text{ probes} \times 455,750 \text{ SNP}) = 1.57 \times 10^{-11}$ and for local associations to a Bonferroni threshold of $0.05/(7002 \text{ probes} \times 200 \text{ SNPs}) = 3.57 \times 10^{-08}$ considering an average number of 200 SNPs tested against each probe.

To account for relatedness in our samples, we generated a matrix of pairwise relatedness estimates (IBD) for all possible pairs of individuals in our cohort. Only autosomal SNPs with a MAF > 0.1 , missingness of 0%, and that were not in linkage disequilibrium ($r^2 < 0.3$) were included in estimating relatedness (final number of SNPs = 1992 SNPs). We used this matrix to estimate the random effects of relatedness in a Q-K mixed model framework using the GLIMMIX procedure in SAS (Yu et al., 2006). This procedure is computationally intensive and was applied only to the associations deemed initially statistically significant for the interaction effect prior to accounting for relatedness.

RESULTS

STUDY DESIGN AND CASE DESCRIPTION

A total of 311 children from Cotonou, Benin, West Africa, were recruited for this study (Table 1, Figure S1). Here, we distinguish between two groups of SCD patients: those who were newly admitted into the program and were labeled as entry (E) and those sampled after being followed and were labeled as FU (see Materials and Methods for details). The initial discovery phase included 126 SCD patients recruited in early 2010, and the replication phase included 124 SCD patients recruited in late 2010. The distribution of clinical categories, Hb genotypes, and sex were proportionate in both phases (Figure S2). In addition, 61 healthy siblings with at least one normal hemoglobin allele and of similar age and proportions of sex were recruited at the CPMI-NFED (Table 1, Figure S2). SCD patients were also assigned a SV (Sebastiani et al., 2007) based on sex, Hb genotype, and cell counts. For all 311 participants, whole blood samples were collected, and complete blood cell counts (CBCs), genome-wide gene expression profiling and genome-wide genotyping were generated. Variabilities in RBC and WBC counts were treated as covariates in the analyses of variance.

DIFFERENTIAL GENE EXPRESSION ANALYSIS—DISCOVERY PHASE

Analysis of gene expression shows that SCD has substantial influence on the whole blood transcriptome. In the discovery phase ($n = 157$), unsupervised hierarchical clustering analysis of the genome-wide gene expression correlation matrix revealed that individual gene expression profiles cluster largely according to Hb genotype, SCD SV, and clinical status (E vs. FU vs. Ctls; Figures 1A,B). PCA revealed the presence of strong correlation structure in the data such that the first three expression principal components (ePC1-3) explain over a third of the total variance (Figure S4). VCA of the first three ePCs further confirms the substantial effect of Hb genotype (explaining 45.6% of the variance) followed by clinical status (explaining 7% of the variance) (Figure 1C). Variance of ePC1 was explained primarily by Hb genotype ($> 70\%$) while ePC2 and 3 were dominated by the effect of clinical status, explaining 20% of the variance of each PC; sex and interaction effects had negligible effects on the variance (Figure S4). Repeating this analysis with only SCD patients ($n = 126$) revealed that a third of the variance (31%) was captured by the first three ePCs, with Hb genotype and the FU effect explaining 19.5 and 8.6% of the variance, respectively (Figure 1C).

Table 1 | Characteristics of study participants.

Dataset	Analysis	n	Sex %F	Hb genotype			ClinStatus		
				HbSS	HbSC	Ctls	E	FU	A
Discovery									
	Gene expression	157	0.49	99	27	31	70	38	18
	Genotyping	142	0.49	89	23	30	62	33	17
Replication									
	Gene expression	154	0.48	91	33	30	64	42	18
	Genotyping	129	0.49	68	32	29	53	31	16
Combined I									
	Gene expression	311	0.49	190	60	61	134	80	36
	Genotyping	263	0.48	151	54	58	110	64	31
Combined II									
	Gene expression	216	0.52	160	n/a	56	102	58	n/a
	Genotyping	173	0.55	120	n/a	53	79	41	n/a

Patient characteristics in the discovery phase, replication phase, and combined datasets I and II. Combined data set II includes only HbSS SCD patients excluding those sampled during acute crises. Approximately, equal numbers of girls and boys were sampled and approximately 2/3 of the SCD patients have the HbSS genotype. Controls have either the HbAS or HbAA genotype. Mean age of the patients and controls was 4 and 3 years, respectively. Sampling was done between February 2010 and December 2010. The numbers of individuals gene expression profiled (Illumina HumanHT-12) or genotyped (Illumina OmniExpress) in each phase are shown. Each main variable Clinical status: (Entry, E; Follow-up, FU; Acutes, A; and Controls, Ctls; Sex (% F, % Female); and Hb Genotype: (HbSS; HbSC; Ctls, C) was sampled in equal proportions in the discovery phase, replication phase, and in the combined datasets. n/a, non-applicable.

Next, we evaluated the magnitude and significance of differentially expressed genes between SCD clinical status and Ctls. Given that a fraction of the variation in ePCs is likely due to differences in the proportion of cell types between SCD patients, we performed a probe-by-probe analysis of covariance (ANCOVA) of the discovery sample that accounts for total blood cell counts (RBC and WBC counts), in addition to sex, and genetic ethnicity using individuals' scores at significant genotypic PC axes (see Materials and Methods for details). This analysis revealed significant differences between SCD patients (E and FU) and Ctls with a quarter of the transcriptome being differentially expressed for the 3-way clinical status effect at 1% False Discovery Rate (FDR) (Figure S5). Thousands of genes were also significantly differentially expressed between Hb genotypes (HbSS, HbSC, Ctls) while minor differences were observed between sexes (Figure 2A) and no effect of the genome-wide genotypic ethnicity effect (gPCs) was detected (Figure S5). Since meaningful population structure in the sample was not observed (Figures S2, S3) and since no probes were significant for the gPC effect (FDR 1%) (Figure S5), genetic ancestry is unlikely to contribute significantly to the observed gene expression differences in our sample.

REPLICATION OF DIFFERENTIAL EXPRESSION AMONG SCD PATIENT GROUPS AND CONTROLS

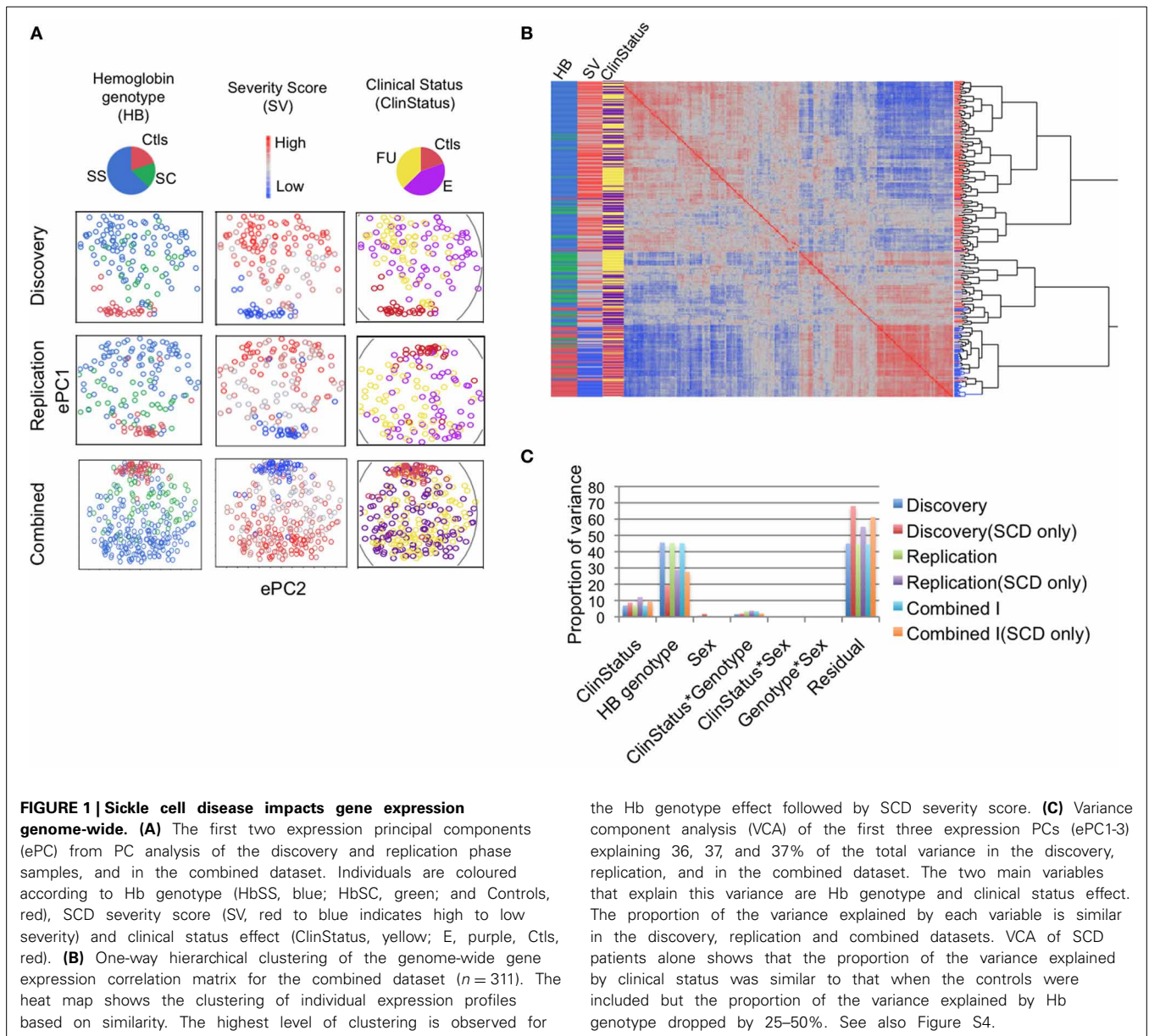
To test the consistency of the patterns of gene expression differentiation observed in the discovery phase, we performed the analyses described above on the replication group ($n = 154$) and the combined dataset (combined dataset I, $n = 311$, see Table 1) and observed similar results (Figures 1, 2A, S4, S5, Supplementary File 5). Unsupervised analysis identified similar clustering by Hb genotype, SCD SV and the clinical status effect (Figures 1A,B), with Hb genotype and the clinical status effect explaining 45.2 and 6.9% of the variance of the first three ePCs in the replication phase, respectively (Figure 1C, Figure S4).

When only SCD patients were included, Hb genotype and the FU effect explained 28.8 and 12.1% of the variance in the first three ePCs, respectively (Figure 1C). The magnitude and significance of differentially expressed probes for the clinical status and Hb genotype effects were highly consistent in both replication and discovery phases (Figures 2A, S5).

Next we focused on 160 SCD HbSS patients and 56 Ctls (combined data set II, $n = 216$, see Table 1) to characterize the transcriptional signatures associated with SCD clinical status and follow-up. HbSC individuals were excluded from this analysis given their small sample size relative to the HbSS group. SCD patients undergoing an acute event were also excluded to focus on the steady state of the disease. An ANCOVA of this dataset accounting for sex and total cell counts revealed that over seven thousand probes were significantly differentially expressed (1% FDR) for the clinical status effect (Figure 2A) and 739 probes for the FU effect (Figure 2B). The effect of clinical status is visually shown in a heat map generated using a 2-way hierarchical clustering of per-group mean expression levels of differentially expressed probes (Figure 2C). The supervised and unsupervised gene expression analysis of both the discovery and replication samples documented the relative contribution of Hb genotype and clinical status to the transcriptional variation observed in a West African SCD cohort and characterized the effects taking place after clinical follow-up. These analyses show that SCD has a substantial influence on whole blood transcriptome with Hb genotype and clinical status explaining the majority of the variation.

IDENTIFICATION OF BIOLOGICALLY RELEVANT PATHWAYS THROUGH ENRICHMENT ANALYSIS

In order to identify the biological pathways subject to the effects of differential expression associated with SCD disease and clinical follow-up, GSEA (Subramanian et al., 2005) was performed using the results of differential expression analysis described above

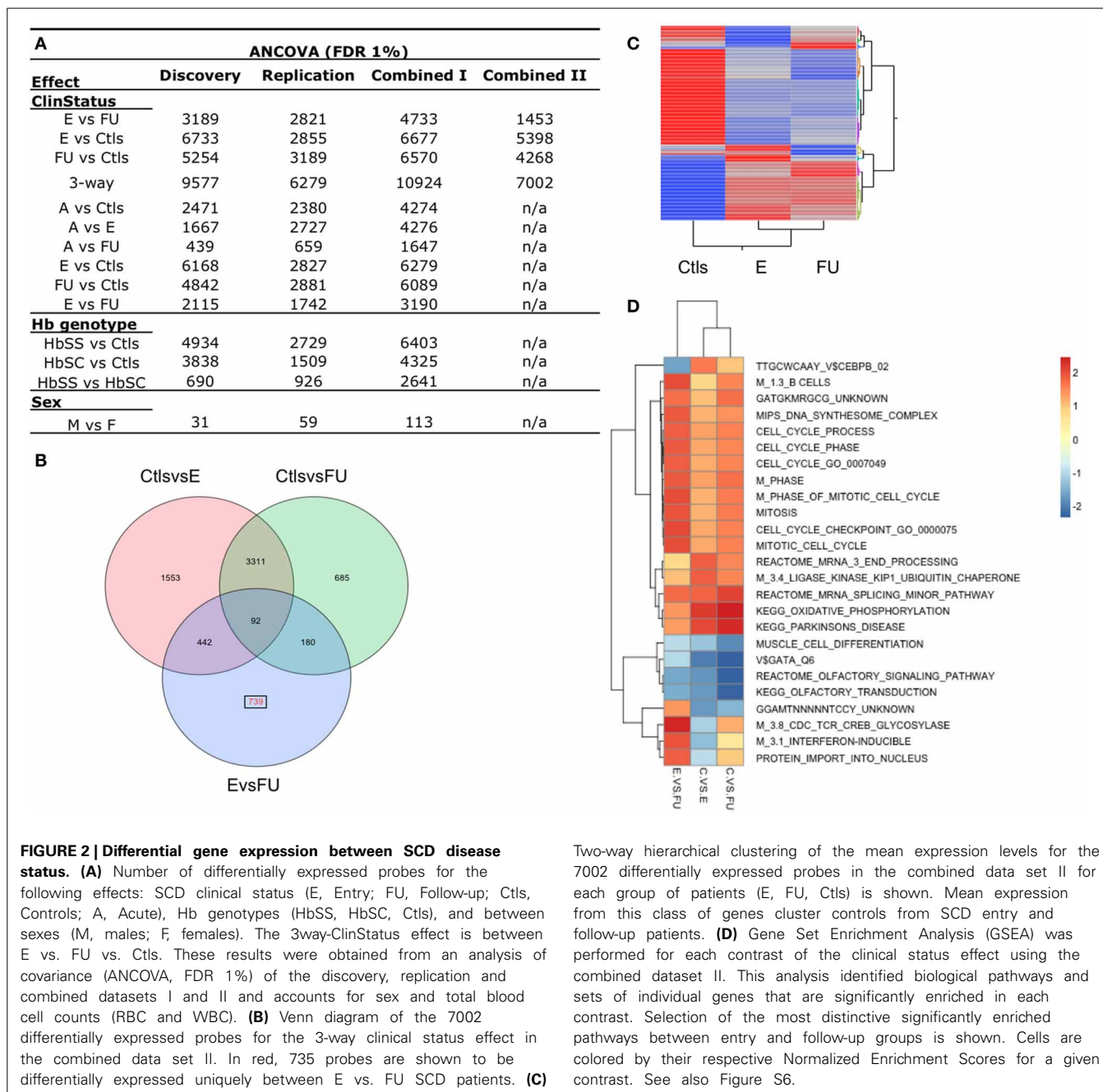


for the discovery, replication and combined datasets (I, II). We focused on gene sets with Normalized Enrichment Scores (NES) greater than 0.25 in either E or FU relative to the Ctls as shown in **Figure 2D**. This analysis showed that the strong modulation of the transcriptome implicates pathways affecting core circulating cell functions. A strong activation of pathways associated with B-lymphocytes development, stress (glucocorticoid, interferon and oxidative phosphorylation associated pathways) and cell proliferation in E compared to the FU group was observed. We also note a significant up-regulation of genes specific to platelet function and erythrocyte membrane in SCD individuals relative to the control and to a lesser extent in FU relative to the E group.

THE GENETIC ARCHITECTURE OF TRANSCRIPT ABUNDANCE IN SCD

The genetic architecture of transcript abundance in SCD was investigated through genome-wide association analysis of gene

expression traits in SCD patients and Ctls. Given the high degrees of correlation in the results of gene expression analyses for the discovery and replication phases and to increase mapping power, we performed the analysis on a subset of the combined dataset II ($n = 173$: 120 HbSS SCD patients and 53 Ctls) for which both gene expression and genotypic data were available. The expression data for the combined dataset was re-normalized in order to minimize potential batch effects, resulting in a final set of 19,431 probes tested against 560,675 SNP genotypes using multiple regression analyses and applying Bonferroni correction for multiple testing. For local associations, the genome-wide significance threshold corresponds to testing on average 200 SNPs against each probe. For distal associations, each probe was tested against each of the 560,675 SNPs. We ran a model that accounted for participant clinical status (E, FU, and Ctls), total blood cell counts (RBS, WBC), and sex (Model 1, see Materials and Methods).



Three hundred and ninety genome-wide significant peak SNP-probe associations were identified corresponding to 371 local and 19 distal effects (Figure 3). These associations explain on average a third of the variance in transcript abundance.

eSNP-BY-CLINICAL STATUS INTERACTIONS

Differential expression analysis revealed 7002 probes significantly differentially expressed (1% FDR) for the clinical status effect. In order to identify which of these genes are under strong genetic regulatory effects that are dependent on clinical status we tested for the SNP-by-ClinStatus interaction effect by including it as term in Model 1 (See Materials and Methods for details).

Bonferroni correction for multiple testing in this analysis was applied. The markers included in this analyses were limited to SNPs with MAF >5% in each clinical group (See Materials and Methods for details). This analysis revealed 11 significant interaction effects, six of which remained genome-wide significant after accounting for relatedness in the entire sample using a Q-K mixed model (Yu et al., 2006) (see Materials and Methods for details): ZSCAN12L1 (p-value = 4.26 × 10⁻¹⁰), C9ORF173 (p-value = 8.94 × 10⁻⁹), CAPZA1 (p-value = 1.33 × 10⁻⁸), SVIL (p-value = 2.41 × 10⁻⁸), MEF2A (p-value = 1.69 × 10⁻⁸), and C1ORF88 (p-value = 5.42 × 10⁻⁹). These interactions are visualized in Figures 4, S7. Figures 4A–C shows three local eSNP

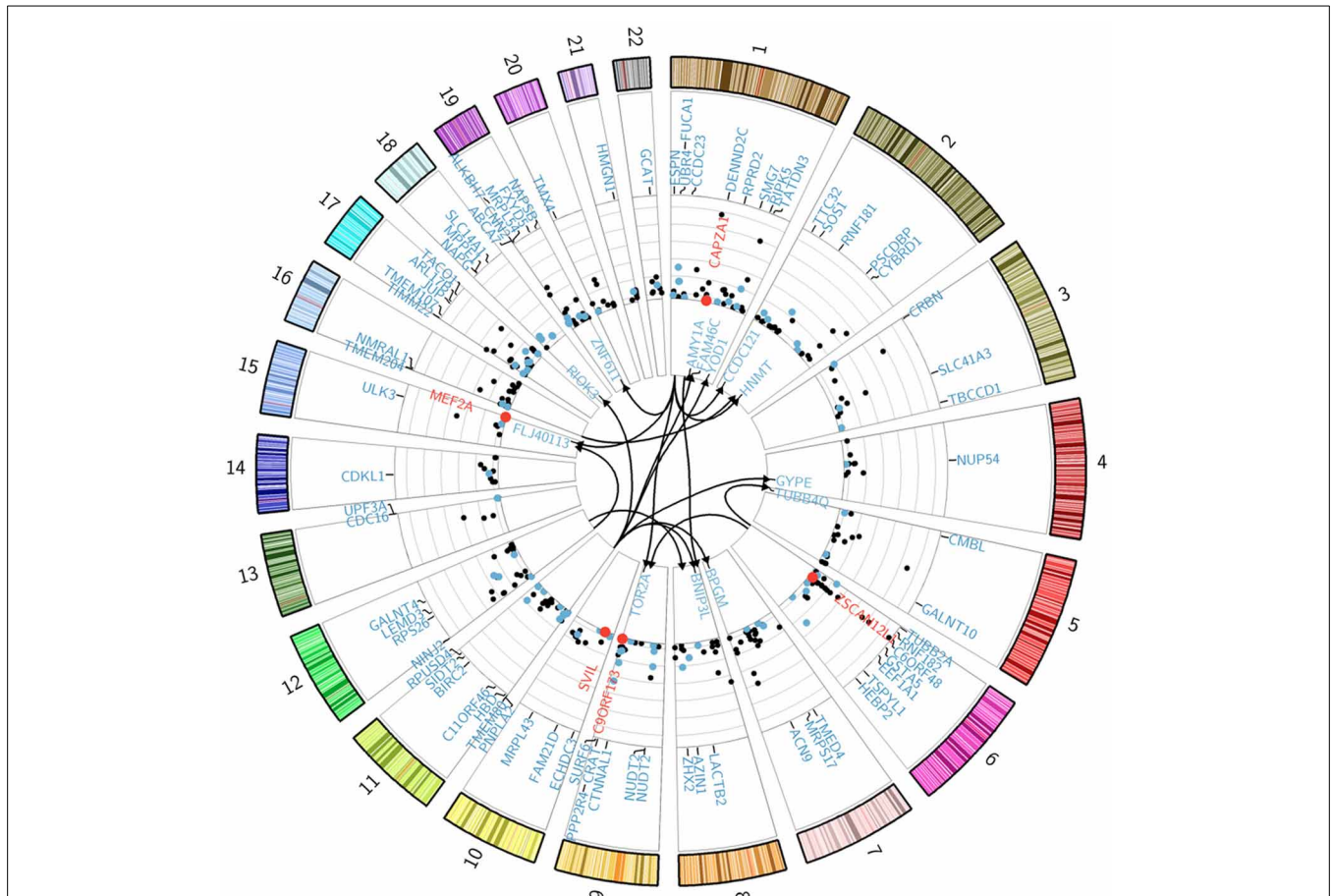


FIGURE 3 | Genetic regulation of gene expression in SCD patients. The Circularized Manhattan plot shows genome-wide significant SNP-probe associations for the analysis that used the combined II dataset. Bonferroni correction for multiple testing was applied to all of our analyses with a genome-wide significance threshold of $p < 0.05/(19,431 \text{ probes} \times 200 \text{ SNPs}) = 1.28 \times 10^{-08}$ (NLP = 7.89) for local associations in model 1 and $p < 0.05/(19,431 \text{ probes} \times 560,675 \text{ SNP}) = 4.59 \times 10^{-12}$ (NLP = 11.34) for distal-associations in model 1; while model 2 thresholds were $p < 0.05/(7002 \text{ probes} \times 200 \text{ SNPs}) = 3.57 \times 10^{-08}$ (NLP = 7.45) for local associations and

$p < 0.05/(7002 \text{ probes} \times 455,750 \text{ SNP}) = 1.57 \times 10^{-11}$ (NLP = 10.80) for distal-associations. Distal associations are shown in the center of the plot. All genes involved in an interaction effect are differentially expressed and shown in red. eSNP genes from model 1 that are differentially expressed for the clinical status effect are shown in blue. The y-axis of the Manhattan plot indicates significance values ($-\log_{10} p$ -values) for the local-associations. Genes under eSNP control that are not differentially expressed for the clinical status effect (in the ANCOVA analysis at FDR 1%) are shown in black. See also Table S1.

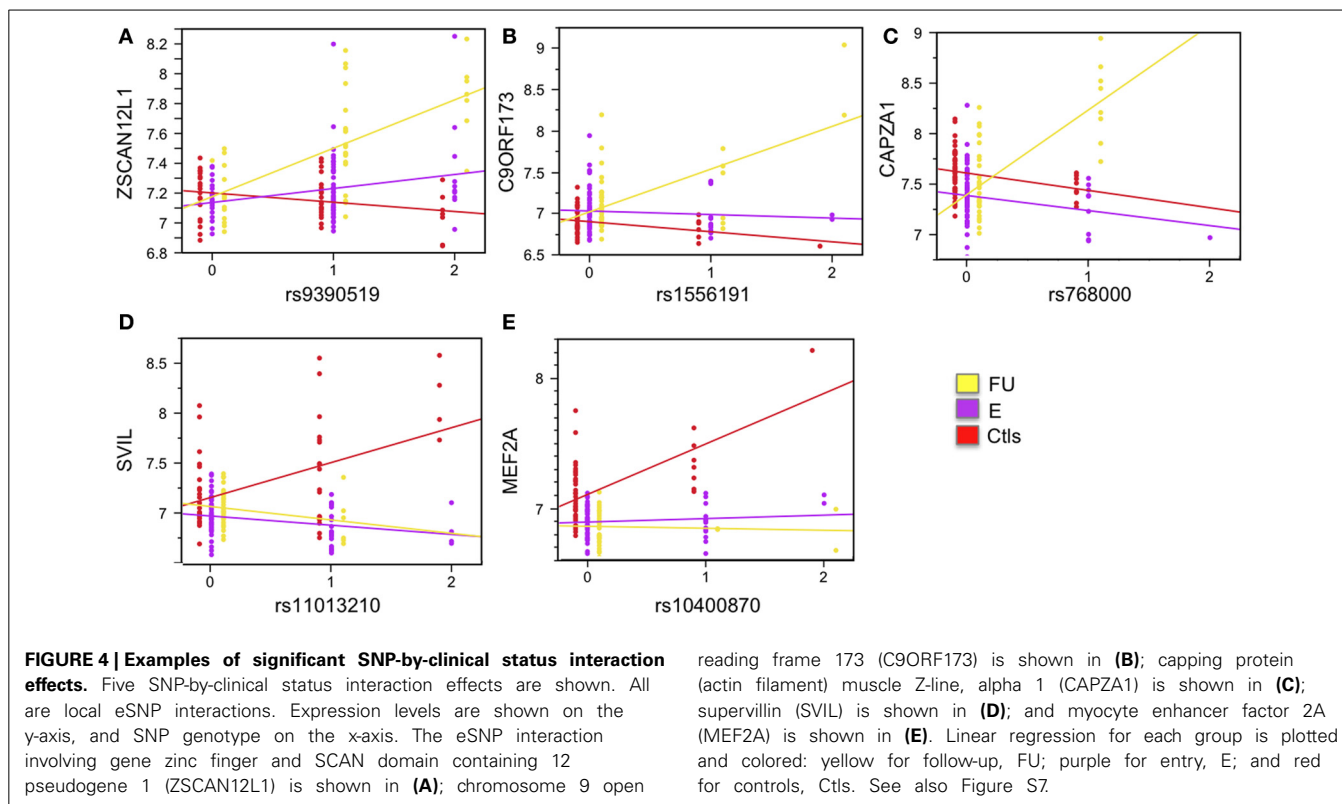
interaction effects where higher expression levels of the corresponding gene in the FU group relative to both the E group and the Ctls is driven by the minor allele of the eSNP in question. **Figures 4D,E** shows two associations where the higher expression levels in the Ctls relative to SCD patients is observed only in the presence of the minor allele for the corresponding eSNP.

DISCUSSION

Here, we characterized the transcriptomes of SCD patients. We first identified the extent of gene expression variation in SCD patients that is explained by clinical phenotypes and measured the magnitude and significance of gene expression differences for SCD clinical status in an initial discovery phase. The unsupervised analysis of gene expression profiles shows that SCD has substantial influence on the human transcriptome, explaining over a third of the total variance, followed by Hb genotype and SCD

clinical status. Significant differences in gene expression profiles between SCD clinical status and Ctls were also observed, with over a quarter of the transcriptome being differentially expressed. We replicated these findings in a replication cohort.

Using GSEA, we identified and replicated biological pathways involved in the clinical course of SCD. This analysis shows that the strong modulation of the transcriptome implicates pathways affecting core circulating cell functions. Enrichment analysis also showed that several biological pathways previously reported to be associated with SCD (Jison et al., 2004) are subject to differential expression between the three clinical groups. Furthermore, we observed strong activation of pathways associated with B-lymphocyte development, stress and cell proliferation in the E compared to the FU group. Enrichment of genes that were uniquely differentiated between the E and FU patients identified a significant up-regulation in B-lymphocytes expressing



phosphorylated CD5, B-cell Receptor Signaling and upstream regulation of B-cells by PAX5. PAX5 expression has been shown to increase the quantity and the commitment of B cells (Horcher et al., 2001). These observations reflect perturbed cellular profiles in the E groups and more stable profiles after clinical FU. Furthermore, markers of mitosis, cell cycle and DNA synthesis were identified in the analysis on combined data set II and likely suggest a more stable state of blood cells in the FU group in general. The strong interferon related signature also suggests a more perturbed and potentially more pathogenic state of blood cells prior to clinical follow-up. The overexpression of activated B lymphocyte markers in the E group tends to point in that same direction. Previous studies have shown that changes in B cell function occurs during vaso-occlusive crisis (VOC) in patients with SCD (Venkataraman and Westerman, 1985). Thus, follow-up of SCD patients may act on these pathways. We also note a significant up-regulation of genes specific to platelet function and erythrocyte membrane in SCD individuals relative to the control group and to a lesser extent in FU relative to the E group. Activation of platelets in SCD patients was previously associated with clinical complications such as vasculopathy (Raghavachari et al., 2007) and hemolysis-associated pulmonary hypertension (Villagra et al., 2007). We observed a strong inflammatory response signature in Acute patients consistent with the processes induced during SCD crises events such as VOC (Musa et al., 2010).

We characterized the genetic architecture of transcript abundance in SCD patients and Ctls and identified 390 genome-wide significant peak SNP-probe associations. Four genes that are associated with an eSNP were previously associated with SCD

phenotypes in reported association studies (Table S1). Almost half of the eSNP genes (150 eSNP genes) that we identified overlapped with previously reported significant eQTL associations (Table S1). Out of these, 58 were exact SNP-gene eSNP pairs. The overlap between our distal eSNP associations and those published in a recent paper that examined the effects of trans eQTLs as putative drivers of disease (Westra et al., 2013) identified three distal eSNPs (rs11171739, rs10493008, and rs6489721) in our SCD cohort that were also associated with genes in complex traits. Although the SNP-gene associations in our study were not exact matches with those reported in the (Westra et al., 2013) paper, it is possible that these 3 trans eSNPs are drivers for SCD related phenotypes.

Differential expression analysis revealed thousands of genes differentially expressed between clinical categories. We identified 11 eSNP interaction effects that are dependent on clinical status for this class of genes, six of which remained genome-wide significant after accounting for relatedness in the entire sample using a Q-K mixed model: *ZSCAN12L1*, *C9ORF173*, *CAPZA1*, *SVIL*, *MEF2A*, and *C1ORF88*. These genes represent novel SCD associations, form an interacting network generated using Ingenuity Pathway Analysis (www.ingenuity.com) (Figure S8) and have some overlapping clinical manifestations, particularly with respect to cardiovascular disease.

For example, *CAPZA1*, capping protein (actin filament) muscle Z-line alpha 1, is a gene located on chromosome 1 that encodes the alpha subunit of the barbed-end actin binding protein (Kuhlman and Fowler, 1997). *CAPZA1* has recently been associated with blood pressure variation in a meta-analysis of

GWAS (Kato et al., 2011). In our study, we see an interaction between cis-acting eSNP rs768000 and clinical status such that SCD patients with the minor allele have higher expression of *CAPZA1* when they are followed-up. The *SVIL* gene encodes isoforms of supervillin, a protein that has been associated with KIRD2DL that regulates the inhibitory signal of natural killer cells that recognize MHC class I molecules (Liu et al., 2011). Recently, a study suggested an inhibitory role for *SVIL* in platelet adhesion and arterial thrombosis using human GWAS and mice knockout approaches (Edelstein et al., 2012). Edelstein et al. (2012) identified that platelets express *SVIL*; platelet thrombus formation is associated with human *SVIL* variants and low *SVIL* expression. We show a local association between *SVIL* gene expression and a SNP on chromosome 10, rs11013210. In our study, Ctls have a significant increase in *SVIL* gene expression when they have the minor allele for the SNP rs11013210. Finally, the *MEF2A* gene encodes a protein that is a DNA-binding transcription factor that activates many muscle-specific, growth factor-induced, and stress-induced genes (Zhao et al., 2012). Defects in this gene have been associated with autosomal dominant coronary artery disease 1 with myocardial infarction (Liu et al., 2012). *MEF2A* affects the proliferation, migration and phenotype of vascular smooth muscle cells (Zhao et al., 2012; Papait et al., 2013). In our samples, we see a local association between *MEF2A* gene expression and SNP rs10400870 genotype on chromosome 15 that is dependent on clinical status. Ctls have a significant increase in *MEF2A* gene expression when they have the minor allele for rs10400870.

In an attempt to estimate the contribution of the genetically controlled fraction of transcript abundance on the association between expression traits and clinical FU categories, we compared the effect of clinical status on global gene expression before and after conditioning for the eSNP effect on transcript abundance. To run this test we used the combined dataset II (focusing on the SNP effect and ClinStatus within the HbSS group) and applied it to all expressed genes. We used a full ANCOVA model with and without the SNP effect and applied the same stringent filtering criteria as described for the analyses above. We extracted *p*-values for the ClinStatus effect for all the genes and contrasted the two models. This comparison revealed overall a relatively high degree of correlation (Figure S9) suggesting that quantitatively most of the transcriptional signal differentiating the clinical categories is robust. However and as shown in Figure S9, we observe a global shift in the significance values toward lower values once genotypic effects are accounted for. To further show this trend, we limited our comparison to genes differentially expressed at Bonferroni level for the ClinStatus effect in at least one of the two tested models ($n = 2188$ genes). After fitting the genotype effect on each of these genes, 41% ($n = 905$) remain significant for the ClinStatus effect and 58% ($n = 1275$) are no longer significant. It is also worth noting that directionality of the effect between the two models was consistent for all tested genes. Moreover and as expected, our top interacting genes (based on *p*-values for the SNP \times ClinStatus interaction effect) were among the list of genes that were subject to the highest drop in statistical significance for association with ClinStatus after fitting the eSNP effect. Taken together, these results suggest that transcript abundance of over

half of expressed genes associated with clinical follow-up is under significant regulatory genetic effect.

Using eQTL approaches, transcriptional genotype-by-environment interactions have previously been reported in humans (Smirnov et al., 2009; Romanoski et al., 2010; Barreiro et al., 2012; Idaghdour et al., 2012) but mostly using *in vitro* systems (with the exception of Idaghdour et al., 2012). Here we demonstrate the existence of these effects *in vivo* in SCD. The genes implicated in these interactions show differential eSNP effects depending on SCD follow-up status. These interactions show how the genetic control of gene expression through allelic variation is likely to impact processes implicated in the response to SCD as well as to clinical follow-up programs.

In summary, using a two-stage sampling design, we identified and replicated a strong transcriptional signature of the effect of follow-up in SCD patients that implicates core biological pathways involved in the pathobiology of the disease. We have provided a genome-wide picture of regulatory variation *in vivo* in SCD patients and highlighted genotype-by-clinical status interaction effects that likely contribute to the clinical heterogeneity observed in SCD patients including those enrolled in SCD clinical follow-up programs. These results further our understanding of the transcriptional events occurring in SCD patients and their genetic regulatory control. The genetic and transcriptional markers reported here can potentially guide follow-up programs. These markers detected in whole blood, a readily and ethically accessible source of biological material in children, will be particularly useful in populations where the disease is most prevalent.

ACCESSION NUMBERS

All expression data are available at NCBI GEO under the series number GSE35007. The individual expression arrays are listed as GSM860207 through GSM860517.

AUTHOR CONTRIBUTIONS

Philip Awadalla and Mohamed C. Rahimy conceived the study. Mohamed C. Rahimy followed the SCD patients and oversaw characterization of SCD patient clinical categories. All hematological analysis was performed at the NSCDC under Mohamed C. Rahimy's direction. Philip Awadalla, Elias Gbeha, Selma Gomez, Jacklyn Quinlan, Mohamed C. Rahimy, Ambaliou Sanni, and Youssef Idaghdour collected the samples. Elias Gbeha, Jacklyn Quinlan and Youssef Idaghdour processed the samples and performed the genomic experiments. Vanessa Bruat, Thibault de Malliard and Jean-Christophe Grenier provided bioinformatics support for statistical analysis of the data by Philip Awadalla and Jacklyn Quinlan. Jean-Philippe Goulet performed enrichment analysis. Philip Awadalla, Youssef Idaghdour and Jacklyn Quinlan wrote the paper.

ACKNOWLEDGMENTS

We are thankful to all of the study participants and their families, as well as the staff of the Centre de Prise en charge Médicale Intégrée du Nourrisson et de la Femme Enceinte atteints de Drépanocytose (CPMI-NFED) who facilitated in collecting the samples. We thank Julie Hussin, Alan Hodgkinson and

Mélanie Capredon for helpful discussions and comments on the manuscript. The research is funded by a Human Frontiers in Science Program Grant RGP0054/2006-C to Philip Awadalla. Jacklyn Quinlan is supported by doctoral Fellowships from the Fonds de la Recherche en Santé du Québec and by the Réseau de médecine génétique appliquée and Youssef Idaghdour is supported by a Banting Post-doctoral Fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00026/abstract>

REFERENCES

- Adams, G. T., Snieder, H., Mckie, V. C., Clair, B., Brambilla, D., Adams, R. J., et al. (2003). Genetic risk factors for cerebrovascular disease in children with sickle cell disease: design of a case-control association study and genomewide screen. *BMC Med. Genet.* 4:6. doi: 10.1186/1471-2350-4-6
- Akinsheye, I., Alsultan, A., Solovieff, N., Ngo, D., Baldwin, C. T., Sebastiani, P., et al. (2011). Fetal hemoglobin in sickle cell anemia. *Blood* 118, 19–27. doi: 10.1182/blood-2011-03-325258
- Barreiro, L. B., Tailleux, L., Pai, A. A., Gicquel, B., Marioni, J. C., and Gilad, Y. (2012). Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1204–1209. doi: 10.1073/pnas.1115761109
- Berry, M. P., Graham, C. M., McNab, F. W., Xu, Z., Bloch, S. A., Oni, T., et al. (2010). An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 466, 973–977. doi: 10.1038/nature09247
- Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* 10, 184–194. doi: 10.1038/nrg2537
- Driss, A., Asare, K. O., Hibbert, J. M., Gee, B. E., Adamkiewicz, T. V., and Stiles, J. K. (2009). Sickle cell disease in the post genomic era: a monogenic disease with a polygenic phenotype. *Genomics Insights* 2009, 23–48.
- Edelstein, L. C., Luna, E. J., Gibson, I. B., Bray, M., Jin, Y., et al. (2012). Human genome-wide association and mouse knockout approaches identify platelet supervillin as an inhibitor of thrombus formation under shear stress. *Circulation* 125, 2762–2771. doi: 10.1161/CIRCULATIONAHA.112.091462
- Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., et al. (2008). Genetics of gene expression and its effect on disease. *Nature* 452, 423–428. doi: 10.1038/nature06758
- Hannemann, A., Weiss, E., Rees, D. C., Dalibalta, S., Ellory, J. C., and Gibson, J. S. (2011). The properties of red blood cells from patients heterozygous for HbS and HbC (HbSC Genotype). *Anemia* 2011, 248527. doi: 10.1155/2011/248527
- Horcher, M., Souabni, A., and Busslinger, M. (2001). Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. *Immunity* 14, 779–790. doi: 10.1016/S1074-7613(01)00153-4
- Idaghdour, Y., and Awadalla, P. (2012). Exploiting gene expression variation to capture gene-environment interactions for disease. *Front. Genet.* 3:228. doi: 10.3389/fgene.2012.00228
- Idaghdour, Y., Quinlan, J., Goulet, J. P., Berghout, J., Gbeha, E., Bruat, V., et al. (2012). Evidence for additive and interaction effects of host genotype and infection in malaria. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16786–16793. doi: 10.1073/pnas.1204945109
- Jison, M. L., Munson, P. J., Barb, J. J., Suffredini, A. F., Talwar, S., Logun, C., et al. (2004). Blood mononuclear cell gene expression profiles characterize the oxidant, hemolytic, and inflammatory stress of sickle cell disease. *Blood* 104, 270–280. doi: 10.1182/blood-2003-08-2760
- Kato, N., Takeuchi, F., Tabara, Y., Kelly, T. N., Go, M. J., Sim, X., et al. (2011). Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* 43, 531–538. doi: 10.1038/ng.834
- Kuhlman, P. A., and Fowler, V. M. (1997). Purification and characterization of an alpha 1 beta 2 isoform of CapZ from human erythrocytes: cytosolic location and inability to bind to Mg²⁺ ghosts suggest that erythrocyte actin filaments are capped by adducin. *Biochemistry* 36, 13461–13472. doi: 10.1021/bi970601b
- Lette, G., Sankaran, V. G., Bezerra, M. A., Araujo, A. S., Uda, M., Sanna, S., et al. (2008). DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11869–11874. doi: 10.1073/pnas.0804799105
- Liu, H. P., Yu, M. C., Jiang, M. H., Chen, J. X., Yan, D. P., Liu, F., et al. (2011). Association of supervillin with KIR2DL1 regulates the inhibitory signaling of natural killer cells. *Cell. Signal* 23, 487–496. doi: 10.1016/j.cellsig.2010.11.001
- Liu, Y., Niu, W., Wu, Z., Su, X., Chen, Q., Lu, L., et al. (2012). Variants in exon 11 of MEF2A gene and coronary artery disease: evidence from a case-control study, systematic review, and meta-analysis. *PLoS ONE* 7:e31406. doi: 10.1371/journal.pone.0031406
- Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., et al. (2007). A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat. Genet.* 39, 1197–1199. doi: 10.1038/ng2108
- Musa, B. O., Onyemelukwe, G. C., Hambolu, J. O., Mammam, A. I., and Isa, A. H. (2010). Pattern of serum cytokine expression and T-cell subsets in sickle cell disease patients in vaso-occlusive crisis. *Clin. Vaccine Immunol.* 17, 602–608. doi: 10.1128/CVI.00145-09
- Papait, R., Cattaneo, P., Kunderfranco, P., Greco, C., Carullo, P., Guffanti, A., et al. (2013). Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20164–20169. doi: 10.1073/pnas.1315155110
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Raghavachari, N., Xu, X., Harris, A., Villagra, J., Logun, C., Barb, J., et al. (2007). Amplified expression profiling of platelet transcriptome reveals changes in arginine metabolic pathways in patients with sickle cell disease. *Circulation* 115, 1551–1562. doi: 10.1161/CIRCULATIONAHA.106.658641
- Rahimy, M. C., Gangbo, A., Ahougnan, G., Adjou, R., Deguenon, C., Goussanou, S., et al. (2003). Effect of a comprehensive clinical care program on disease course in severely ill children with sickle cell anemia in a sub-Saharan African setting. *Blood* 102, 834–838. doi: 10.1182/blood-2002-05-1453
- Romanoski, C. E., Lee, S., Kim, M. J., Ingram-Drake, L., Plaisier, C. L., Yordanova, R., et al. (2010). Systems genetics analysis of gene-by-environment interactions in human cells. *Am. J. Hum. Genet.* 86, 399–410. doi: 10.1016/j.ajhg.2010.02.002
- Sankaran, V. G., Lettre, G., Orkin, S. H., and Hirschhorn, J. N. (2010). Modifier genes in Mendelian disorders: the example of hemoglobin disorders. *Ann. N.Y. Acad. Sci.* 1214, 47–56. doi: 10.1111/j.1749-6632.2010.05821.x
- Sebastiani, P., Nolan, V. G., Baldwin, C. T., Abad-Grau, M. M., Wang, L., Adewoye, A. H., et al. (2007). A network model to predict the risk of death in sickle cell disease. *Blood* 110, 2727–2735. doi: 10.1182/blood-2007-04-084921
- Sebastiani, P., Solovieff, N., Hartley, S. W., Milton, J. N., Riva, A., Dworkis, D. A., et al. (2010). Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. *Am. J. Hematol.* 85, 29–35. doi: 10.1002/ajh.21572
- Smirnov, D. A., Morley, M., Shin, E., Spielman, R. S., and Cheung, V. G. (2009). Genetic analysis of radiation-induced changes in human gene expression. *Nature* 459, 587–591. doi: 10.1038/nature07940
- Solovieff, N., Milton, J. N., Hartley, S. W., Sherva, R., Sebastiani, P., Dworkis, D. A., et al. (2010). Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. *Blood* 115, 1815–1822. doi: 10.1182/blood-2009-08-239517
- Steinberg, M. H., and Sebastiani, P. (2012). Genetic modifiers of sickle cell disease. *Am. J. Hematol.* 87, 795–803. doi: 10.1002/ajh.23232
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Thein, S. L. (2011). Genetic modifiers of sickle cell disease. *Hemoglobin* 35, 589–606. doi: 10.3109/03630269.2011.615876
- Venkataraman, M., and Westerman, M. P. (1985). B-cell changes occur in patients with sickle cell anemia. *Am. J. Clin. Pathol.* 84, 153–158.

- Villagra, J., Shiva, S., Hunter, L. A., Machado, R. F., Gladwin, M. T., and Kato, G. J. (2007). Platelet activation in patients with sickle disease, hemolysis-associated pulmonary hypertension, and nitric oxide scavenging by cell-free hemoglobin. *Blood* 110, 2166–2172. doi: 10.1182/blood-2006-12-061697
- Weatherall, D. J. (2001). Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.* 2, 245–255. doi: 10.1038/35066048
- Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243. doi: 10.1038/ng.2756
- World Health Organization, T. W. H. A. (2006). “Sickle cell anaemia”, in *Report by the Secretariat in Provisional Agenda Item 11.4. A59/9*. Geneva: World Health Organization.
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhao, W., Zhao, S. P., and Peng, D. Q. (2012). The effects of myocyte enhancer factor 2A gene on the proliferation, migration and phenotype of vascular smooth muscle cells. *Cell Biochem. Funct.* 30, 108–113. doi: 10.1002/cbf.1823
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 22 September 2013; accepted: 24 January 2014; published online: 14 February 2014.
- Citation: Quinlan J, Idaghdour Y, Goulet J-P, Gbeha E, de Malliard T, Bruat V, Grenier J-C, Gomez S, Sanni A, Rahimy MC and Awadalla P (2014) Genomic architecture of sickle cell disease in West African children. *Front. Genet.* 5:26. doi: 10.3389/fgene.2014.00026
- This article was submitted to *Genetic Disorders*, a section of the journal *Frontiers in Genetics*.
- Copyright © 2014 Quinlan, Idaghdour, Goulet, Gbeha, de Malliard, Bruat, Grenier, Gomez, Sanni, Rahimy and Awadalla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.