A common molecular signature of patients with sickle cell disease revealed by microarray meta-analysis and a genome-wide association study

Cherif Ben Hamda, Raphael Sangeda, Liberata Mwita, Ayton Meintjes, Siana Nkya, Sumir Panji, Nicola Mulder, Lamia Guizani-Tabbane, Alia Benkahla, Julie Makani, et al.

HAL Id: pasteur-02009067
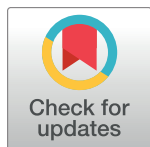## https://riip.hal.science/pasteur-02009067

Submitted on 6 Feb 2019

A common molecular signature of patients with sickle cell disease revealed by microarray meta-analysis and a genome-wide association study

Cherif Ben Hamda[1,2,3]*, Raphael Sangeda[4], Liberata Mwita[4], Ayton Meintjes[5], Siana Nkya[4], Sumir Panji[5], Nicola Mulder[5], Lamia Guizani-Tabbane[2,6], Alia Benkahla[1,2], Julie Makani[3], Kais Ghedira[1,2]*, H3ABioNet Consortium[¶]

1 Laboratory of Bioinformatics, Biomathematics and Biostatistics, Institute Pasteur of Tunis, Tunis, Tunisia, 2 University of Tunis El Manar, Tunis, Tunisia, 3 Faculty of Science of Bizerte, Jarzouna, University of Carthage, Tunisia, 4 Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania, 5 University of Cape Town, Cape Town, South Africa, 6 Laboratory of Medical Parasitology, Biotechnology and Biomolecules, Institute Pasteur of Tunis, Tunis, Tunisia

¶ The complete membership of the H3ABioNet Consortium can be found in the Acknowledgments.
* kais.ghedira@pasteur.rns.tn (KG); cherifbenhamda@gmail.com (CBH)

Abstract

A chronic inflammatory state to a large extent explains sickle cell disease (SCD) pathophysiology. Nonetheless, the principal dysregulated factors affecting this major pathway and their mechanisms of action still have to be fully identified and elucidated. Integrating gene expression and genome-wide association study (GWAS) data analysis represents a novel approach to refining the identification of key mediators and functions in complex diseases. Here, we performed gene expression meta-analysis of five independent publicly available microarray datasets related to homozygous SS patients with SCD to identify a consensus SCD transcriptomic profile. The meta-analysis conducted using the MetaDE R package based on combining p values (maxP approach) identified 335 differentially expressed genes (DEGs; 224 upregulated and 111 downregulated). Functional gene set enrichment revealed the importance of several metabolic pathways, of innate immune responses, erythrocyte development, and hemostasis pathways. Advanced analyses of GWAS data generated within the framework of this study by means of the atSNP R package and SIFT tool identified 60 regulatory single-nucleotide polymorphisms (rSNPs) occurring in the promoter of 20 DEGs and a deleterious SNP, affecting CAMKK2 protein function. This novel database of candidate genes, transcription factors, and rSNPs associated with SCD provides new markers that may help to identify new therapeutic targets.

Introduction

Sickle cell disease (SCD) is a rare hemoglobinopathy [1] characterized by high morbidity and mortality. It affects millions of people throughout the world, including 300 thousand newborns

annually [2]. The disease represents a severe genetic blood disorder caused by a single muta-tion in β globin [3] that leads to a defective hemoglobin S (HBS) protein, which upon deox-ygenation polymerizes rapidly [4] and gives red blood cells rigidity and sickle shape. Losing flexibility, the misshapen red blood cells obstruct blood flow, disrupt homeostasis, and pro-mote chronic hemolysis. The released heme iron into plasma creates an oxidative microenvi-ronment resulting in inflammation and endothelial-cell activation, leading to tissue damage in almost all organs and systems.

The SCD pathophysiology is largely explained by the chronic inflammatory state [5] and partly by the dysregulated pro- and antiapoptotic agents [6]. Nonetheless, the principal dysre-gulated factors in these major pathways and their mechanisms of action still need to be fully identified and elucidated.

Several technologies [7,8] have been developed to identify candidate markers and have highlighted potential molecular pathways related to SCD. Among the high-throughput geno-mic technologies that have contributed to the identification of candidate genes and to our understanding of complex interactions in multisystem diseases is the microarray technology. Indeed, a microarray is a gene expression profiling technology that is widely used to simulta-neously assess the expression levels of thousands of genes [9] in different disease contexts. Such high-throughput technique is considered an important tool for clinical practice and for the development of diagnostics [10,11]. Thus, its massive use in biomedical studies in the last decade has resulted in a huge quantity of data for a large number of physiological states and disease conditions [12]. In parallel, the constant evolution of bioinformatics tools provided the scientific community with a more effective and reproducible meta-analysis workflow [13–20] allowing for better handling of differences in study design and platforms and providing enhancement of the analytical performance, which results in a more robust and reliable analy-sis of gene expression signatures [21–24].

On the other hand, a genome wide association study (GWAS) is an approach that involves rapid scanning of markers (genetic variants) across the genomes of different individuals [25]. It focuses on association between genetic variations such as single-nucleotide polymorphisms (SNPs) and a specific disease. Recent studies have shown that integrative approach combining GWAS and gene expression data is more effective than analyzing each data type individually [26–29]. Indeed, SNPs present in the promoter regions of differentially expressed genes (DEGs) and more specifically in regulatory elements [30] are likely to be regulatory SNPs (rSNPs) that could mediate the binding of critical transcription factors (TFs) and consequently alter the expression of target genes [31]. Therefore, joint identification of rSNPs and a gene expression meta-signature will potentially advance the understanding of the disease mecha-nisms [30,31].

In the present study, we performed a gene expression meta-analysis to identify a consensus transcriptomic profile related to SCD, and by means of GWAS data, we investigated the involvement of rSNPs in the expression of the identified key candidate genes.

Materials and methods

Search criteria and data collection

Analysis of the publicly available microarray datasets was carried out as per PRISMA guide-lines [32] (S1 Table and **Fig 1**). A search involving a combination of the following keywords ("sickle cell disease" and "Homo sapiens") was performed in two public microarray expression data repositories: NCBI Gene Expression Omnibus (GEO) Datasets [http://www.ncbi.nlm.nih.gov/geo/] and ArrayExpress database [http://www.ebi.ac.uk/arrayexpress/]. Studies that were publicly available by September 2016 were extracted. "Globin depleted" and "Homozygote SS"
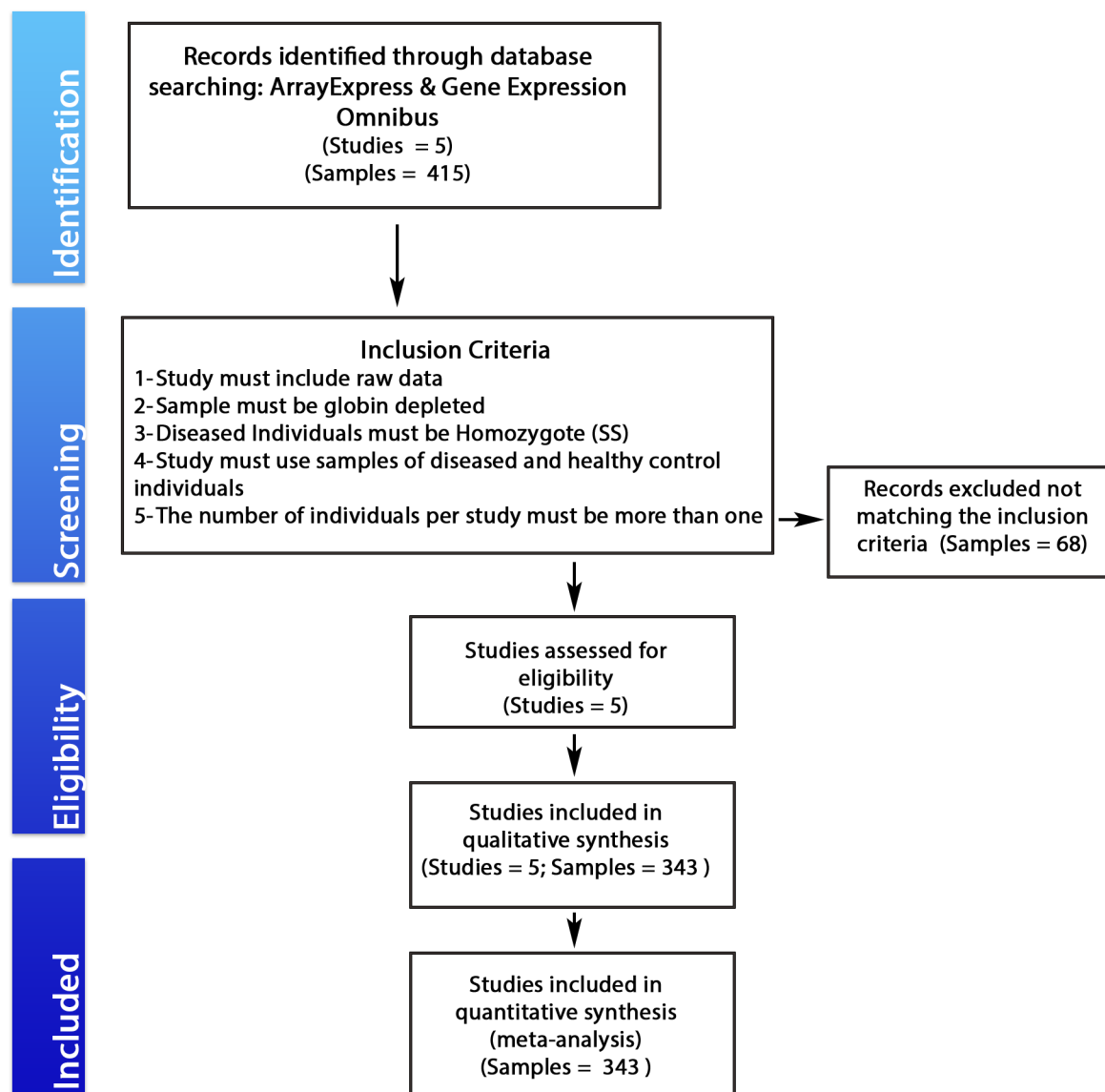
Fig 1. A PRISMA flow diagram of a systematic database search. The selection process of eligible microarray datasets for meta-analysis of the shared transcriptomic signatures between Sickle cell disease patients, according to Prisma 2009 flow diagram.

https://doi.org/10.1371/journal.pone.0199461.g001

Table 1. Characteristics of included individual microarray dataset.

GEO Accession Number	Tissue	Number of samples (Control/Cases)	Microarray Platform	Reference study
GSE11524	Peripheral blood	30(12/18)	Affymetrix Human Genome U133 plus 2.0	[33]
GSE16728	Peripheral blood	20(10/10)*	Affymetrix Human Genome U133 plus 2.1	[34]
GSE53441	Peripheral blood	34(10/24)	Affymetrix Human Genome U133 plus 2.2	[35]
GSE31757	Whole blood	8(3/5)	Affymetrix Human Exon 1.0 ST	[7]
GSE35007	Whole blood	251(61/190)	Illumina HumanHT-12 v4	[36]

Inclusion criteria: Only samples homozygotes SS and submitted to Globin reduction were selected.

*: Consists of two studies with two datasets in total

https://doi.org/10.1371/journal.pone.0199461.t001

were set as sample inclusion criteria. Table 1 provides details on all the datasets included in this study.

Data preparation and quality control

Data preprocessing and meta-analysis were carried out using MetaQC [37] and MetaDE [18], respectively. These packages of the R suite were chosen for quality control (QC) and identification of differentially expressed genes (DEGs) in microarray meta-analysis. The input gene expression table was prepared from the downloaded raw data following MetaDE guidelines [18]. In brief, probe IDs of an individual study were annotated to retrieve their official gene symbols using Bioconductor. The interquartile range method was applied to select the probe ID with the largest interquartile range of expression values when multiple probes matched an identical gene symbol. Individual studies that contained samples from different populations were divided into substudies and treated as independent datasets during the meta-analysis. Afterwards, the processed data were analyzed with MetaQC to compute six QC measures: (1) internal homogeneity of coexpression structure among studies (IQC); (2) external consistency of coexpression structure correlating with a pathway database (EQC); (3) accuracy of DEG detection (AQCg) or pathway identification (AQCp); and (4) consistency of differential expression ranking of genes (CQCg) or pathways (CQCp). The datasets showing good performance on at least five QC criteria were retained. Table 2 presents the quantitative QC measures of each dataset.

Differential expression analysis

DEG analysis was performed with Limma software package [38] for each dataset independently using an adjusted p value ≤0.05, based on Benjamini–Hochberg false discovery rate (FDR) and the moderated t test. In this meta-analysis, DEGs across diseases and healthy controls were selected by means of a maximum p value (maxP) [39]. This method combines p values and targets DEGs that have small p values in all studies (FDR < 0.05). Significantly up- and downregulated DEGs were defined as those that showed a fold change (FC) greater than 1.4 in either direction.

Functional annotation and gene-regulatory network inference

The candidate genes identified via our meta-signature were used as input for different bioinformatics enrichment tools. Gene Ontology (GO) terms were identified in the TRANSPATH database [40] on the geneXplain platform [41]; a different pathway gene set libraries (KEGG Pathway, Reactome, and Wiki pathway), were requested by means of the EnrichR web tool

Table 2. MetaQC quantitative quality control measures for gene expression data.

Dataset	IQC	EQC	CQCg	CQCp	AQCg	AQCp	Rank
GSE35007	2.53	3.7	301.34	307.65	145.57	220.18	2.17
GSE11524	5.33	3.6	148.75	307.65	101.43	210.4	2.5
GSE16728-A	3.58	3.7	145.79	307.65	78.46	157.71	3
GSE31757	3.97	1.17*	102.06	307.65	44.78	138.97	3.92
GSE53441	7.02	3.82	58.5	73.32	27.43	37.71	4
GSE16728-B	0.41*	3.6	33.06	212.76	18.27	88.93	5.42

Inclusion Criteria: Dataset with good performance in at least five quality control criteria

*: Low performance

[42]. The selection criterion for significantly enriched pathways and GO terms was a p value of 0.05 or less. Furthermore, for better result interpretation, we constructed a biological network describing the most significant over-represented GO biological process terms obtained at the previous step. In brief, we conducted two-sided (enrichment/depletion) hypergeometric distribution tests, with a p value significance level of ≤ 0.05, followed by the Bonferroni adjustment using the ClueGO plugin [43] of the Cytoscape [44] software.

Protein–protein (PP) interactions including physical and functional association across our set of genes were identified in stringdb 10.0 [45]. To gain more insights into post-transcriptional regulation of expression of DEGs, we used the complete list of our DEGs, their log FC, and the p value as input for the RNEA R [46] package to reconstruct a subnetwork of gene regulation.

GWAS data integration

GWAS was carried out on 1,213 individuals (HbSS and HbSb0) and gave a collection of 15,153,765 SNPs and indels related to SCD [47]. From those we selected 1,000 SNPs (with the best p-value ranking) affecting the genes that we identified as DE by our meta-analysis. To identify the location of each SNP within the related candidate gene, we used the list of the 1,000 SNPs as input for the VariantAnnotation Bioconductor package with 2,000 as value representing the number of base pairs upstream of the 5'- gene end and 200 representing those downstream of the 3'- gene end. SNPs occurring in coding regions were used as input for SIFT (Sorting Intolerant from Tolerant) web server—able to identify nonsynonymous SNPs and predict the effect of coding variants on protein function. The remaining SNPs (occurring in non-coding regions) were used as input for atSNP (A:ffinity T:esting for regulatory SNP: s) R package [42] to predict the regulatory variants (rSNPs). This software later computes an affinity score for the genomic sequences around each SNP (±30 bp) against a library of TF motifs for both alleles (reference and SNP allele) using their corresponding position-weighted matrices (PWMs). SNPs with a significant change in the affinity scores between the reference and SNP alleles were then hypothesized as rSNPs. We selected ENCODE PWMs as a TF motif library and set a p value corrected by the Benjamini–Hochberg criterion in the rank test between alleles (pval_rank_bh) to less than 0.1 to select significant rSNPs.

The main steps of the bioinformatics workflow integrating microarray meta-analysis and GWAS data are described in **Fig 2**.

Results

Data mining and quality assessment

Five case-control studies (GSE11524, GSE16728, GSE53441, GSE31757 and GSE35007) satisfying the fixed inclusion criteria were selected for this meta-analysis. GSE16728 includes two cohorts; each cohort was regarded as a different dataset and was analyzed separately. Sample sources were either peripheral blood or whole blood subjected to globin mRNA depletion [48]. The retained studies included 343 samples representing a total of 247 homozygous SCD patients and 96 healthy controls. Table 1 provides detailed information about the datasets, sample types, study references, and the microarray platforms used here. All expression-processed raw data that yielded a high score at the QC step are presented in Table 2.

Identification of the gene expression meta-signature

A shared transcriptional signature among SCD patients was identified by a meta-analysis of the retained datasets by combining p values according to the stringent maxP method [39]. The
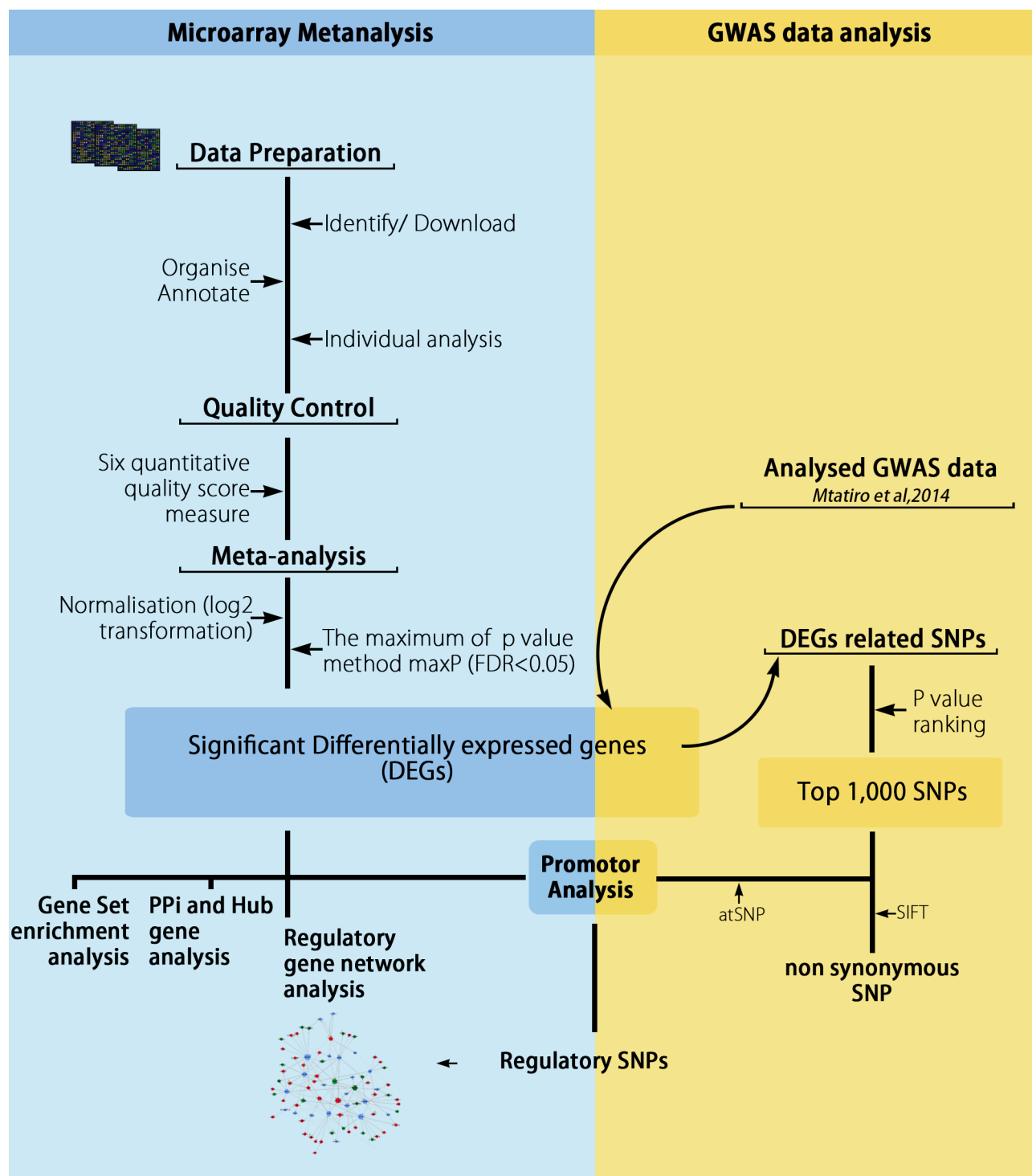
Fig 2. Workflow of microarray meta-analysis integrating GWAS data. A flow diagram depicting the process involved in the meta-analysis of the selected microarray datasets with integration of GWAS data.

raw data were loaded into the R environment and analyzed using the R package MetaDE. Non-DEGs as well as those showing small variation and expression intensities across the

Table 3. Top 20 shared DEGs identified in the meta-analysis ranked by average Log2FC.

Entrez Gene ID	HGNC Gene symbol	Gene Description	Average Log2FC	FDR
Top 10 upregulated Genes				
9829	DNAJC6	DnaJ heat shock protein family (Hsp40) member C6	4,938	0,017
221687	RNF182	Ring finger protein 182	4,003	0,002
759	CA1	carbonic anhydrase I	3,366	1,38E-18
3045	HBD	Hemoglobin subunit delta	3,294	1,38E-18
2994	GYPB	Glycophorin B (MNS blood group)	3,246	1,38E-18
9911	TMCC2	Transmembrane and coiled-coil domain family 2	3,166	1,38E-18
2993	GYPA	Glycophorin A (MNS blood group)	3,070	1,38E-18
8140	SLC7A5	Solute carrier family 7 (amino acid transporter light chain, L system), member 5	2,999	0,0004
66008	TRAK2	Trafficking protein, kinesin binding 2	2,985	0,023
493856	CISD2	CDGSH iron sulfur domain 2	2,980	1,38E-18
Top 10 downregulated Genes				
79971	WLS	Wntless Wnt ligand secretion mediator	-1,766	0,036
6934	TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)	-1,562	0,016
29909	GPR171	G protein-coupled receptor 171	-1,523	0,027
79668	PARP8	Poly(ADP-ribose) polymerase family member 8	-1,344	0,002
64167	ERAP2	Endoplasmic reticulum aminopeptidase 2	-1,343	0,048
91526	ANKRD44	Ankyrin repeat domain 44	-1,300	0,001
5788	PTPRC	Protein tyrosine phosphatase, receptor type, C	-1,275	0,005
1362	CPD	Carboxypeptidase D	-1,209	0,030
50852	TRAT1	T cell receptor associated transmembrane adaptor 1	-1,169	0,009
23224	SYNE2	Spectrin repeat containing, nuclear envelope 2	-1,159	0,047

https://doi.org/10.1371/journal.pone.0199461.t003

majority of datasets and contributing to the FDR were filtered out. The statistical framework (maxP) identified highly significant biomarkers that are differentially expressed in all studies: 335 DEGs including 224 overexpressed and 111 underexpressed genes satisfying the significance threshold of FDR < 0.05 (see S2 Table).

The top 10 upregulated and top 10 downregulated DEGs are presented in Table 3, along with the average FC in expression. Among these, DnaJ heat shock protein family (Hsp40) member C6 (*DNAJC6*), ring finger protein 182 (*RNF182*), and carbonic anhydrase I (*CA1*) were the most significantly overexpressed genes, whereas Wntless Wnt ligand secretion mediator (*WLS*), Transcription factor 7-like 2 (T-cell specific, HMG-box) (*TCF7L2*), and G protein-coupled receptor 171 (*GPR171*) were the most underexpressed genes (Table 3).

Network construction

PP interaction analysis was conducted to identify the key hub genes among our 335 DEGs. Using the STRING database [45], we generated a PP interaction network among 175 nodes with a high confidence score of 0.7. For better visualization of the network, we extracted the most connected components that represent the 139 major nodes of the network. These included 139 proteins and 278 edges representing the interactions within **Fig 3A.** Based on network topology measures, a list of top 10 hub genes was compiled (see S3 Table). Four key hub genes that showed a higher score in both degree and betweenness-centrality metrics are coding for S-phase kinase-associated protein 1 (SKP1; degree = 15, betweenness centrality = 0.35), HECT and RLD domain-containing E3 ubiquitin protein ligase 5 (HERC5; degree = 13, betweenness centrality = 0.14), NSF attachment protein alpha (NAPA; degree = 12, betweenness centrality = 0.32), and erythrocyte membrane protein band 4.2 (EPB42;
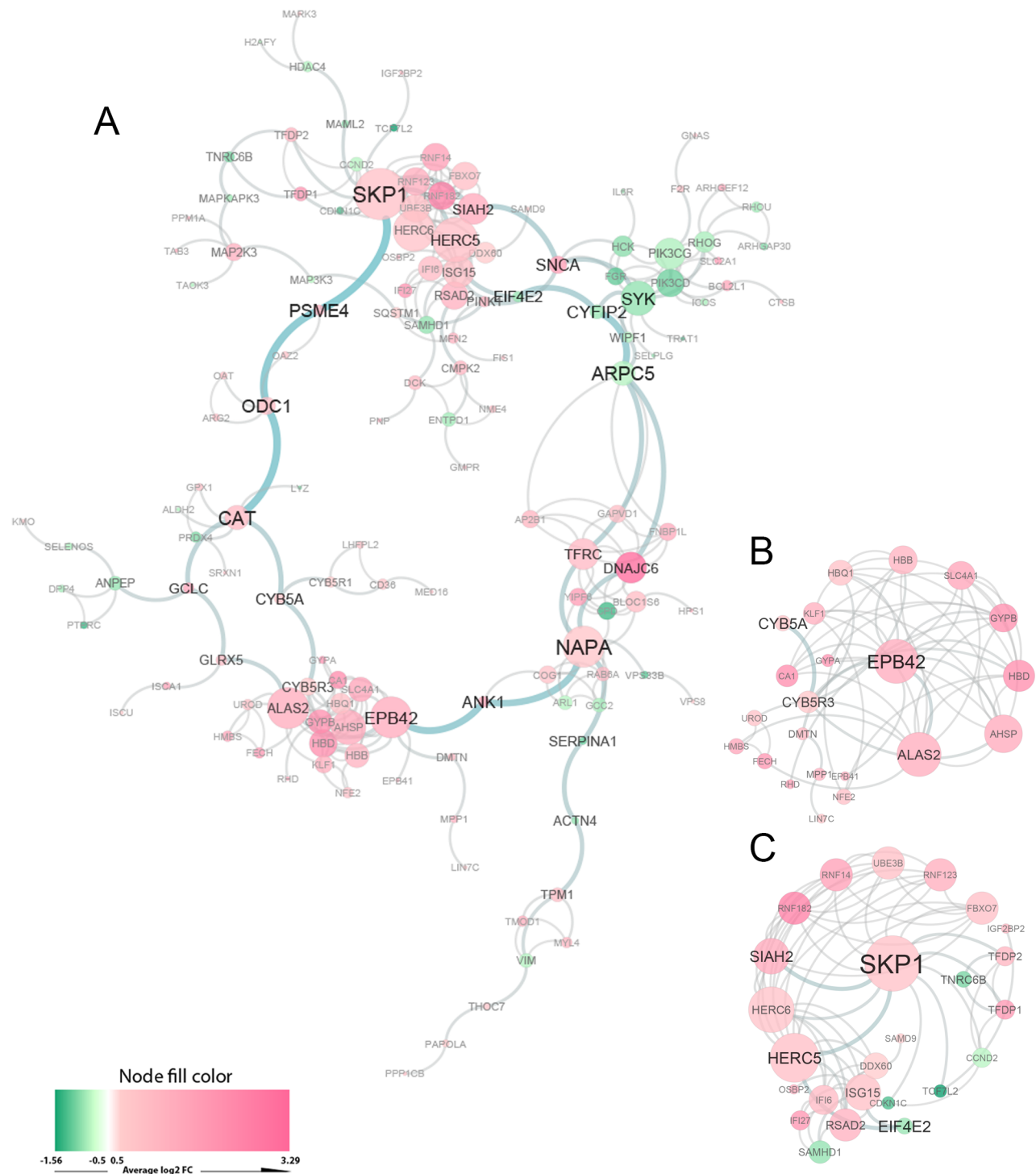
Fig 3. Network-based meta-analysis of hub genes. A: Protein interaction network analysis indicates a central role for SKP1, NAPA, EPB42, and ARPC5 in SCD anemia. All 335 genes served as input for the STRING database with the high confidence interaction score 0.7, and a network was built by means of Cytoscape. The network topology was analyzed by the Cytoscape NetworkAnalyzer tool, and then network topology measures such as the degree (represented by the node size scale), betweenness (represented by police size scale), closeness centrality, and clustering coefficient were calculated. **B and C:** The top-ranked subnetwork identified by the OH-PIN algorithm (threshold: 2, overlapping score 0.5) using CytoCluster (a Cytoscape plugin).

https://doi.org/10.1371/journal.pone.0199461.g003

degree = 12, betweenness centrality = 0.22). Uploading the 139 proteins via the CytoCluster Cytoscape plugin extracted the top-ranked clusters as a subnetwork. These included SKP1 (25 nodes and 73 edges) and EPB42 (22 nodes and 52 edges). Both subnetworks are independent protein complexes with their interacting nodes likely to work collectively to perform biological functions listed in Fig 3B and 3C.

To gain insight into the regulatory system upstream of the identified DEGs, we employed the Regulatory Network Enrichment Analysis (RNEA) bioinformatics tool that extracts lists of prioritized TFs. A regulatory subnetwork showing the interaction of activated regulators with their target genes and with one another is shown in Fig 4.

Identification of regulatory and nonsynonymous SNPs

Identifying potential rSNPs is crucial for understanding disease mechanisms. From a subset of the most significant 1,000 DEGs associated SNPs (retrieved from GWAS data analysis based on P-value) and using the atSNP R package [15], we evaluated the regulatory potential of 789 SNPs (occurring in non-coding regions) by comparing them with a motif library of 2,065 PWMs from the ENCODE project [16]. The analysis revealed a total of 60 significant rSNPs that affect gene regulation of 20 candidate disease genes out of 335 DEGs. These SNPs affect binding affinity of TFs for their target promoter (full results are in S4 Table). Besides, we identified seven TFs (Fig 4) carrying rSNPs such as NFE2, which plays a prominent role in the hematopoietic stem cell differentiation pathway.

We next evaluated the functional impact of the remaining coding variants (211 SNPs out of 1,000 SNPs) in the affected proteins. According to SIFT, one nonsynonymous SNP (rs1132780) was predicted as damaging with a significant score of 0.042 (results are in S5 Table). Rs1132780 occurs in the coding region of the calcium/calmodulin-dependent protein kinase kinase 2 (CAMKK2) gene and affects all its catalytic domain isoforms with an amino acid change from arginine to cysteine at position 363 of the protein.

Gene set enrichment analysis for identification of over-represented biological pathways and GO terms

Enriched GO biological process terms and biological pathways associated with our complete list of DEGs and showing a p value <0.05 were identified by means of both geneXplain and EnrichR tools (full results are in S6 Table). Among the top significant terms identified were those associated with an innate immune response, oxidative stress, hemostasis, and hemopoiesis (Fig 5). Relevant GO terms identified only on the basis of the list of DEGs carrying rSNPs were associated mainly with B-cell proliferation, T-cell differentiation, B-cell activation involved in an immune response, and T-cell lineage commitment.

Discussion

SCD is a monogenic disorder characterized by chronic hemolytic anemia and episodic vaso-occlusion. These major pathological signs are primarily triggered by the polymerization of defective hemoglobin S [49] and end with a sustained inflammatory response and tissue damage. Multiple cell types are involved in the chronic inflammatory state and are associated with SCD severity, including leukocytes, activated monocytes, neutrophils, and platelets [50–52]. In addition, SCD patients show higher levels of several molecules that promote inflammation, including heme and proinflammatory cytokines, such as TNF-α, IL-1β, and IL-8 [53–55], which enhance the vaso-occlusive process [53,56] and aggravate the disease.

The inflammatory response thus plays a prominent role in SCD pathophysiology [51] and needs to be resolved to help control the disease and prevent the subsequent damage to organs.
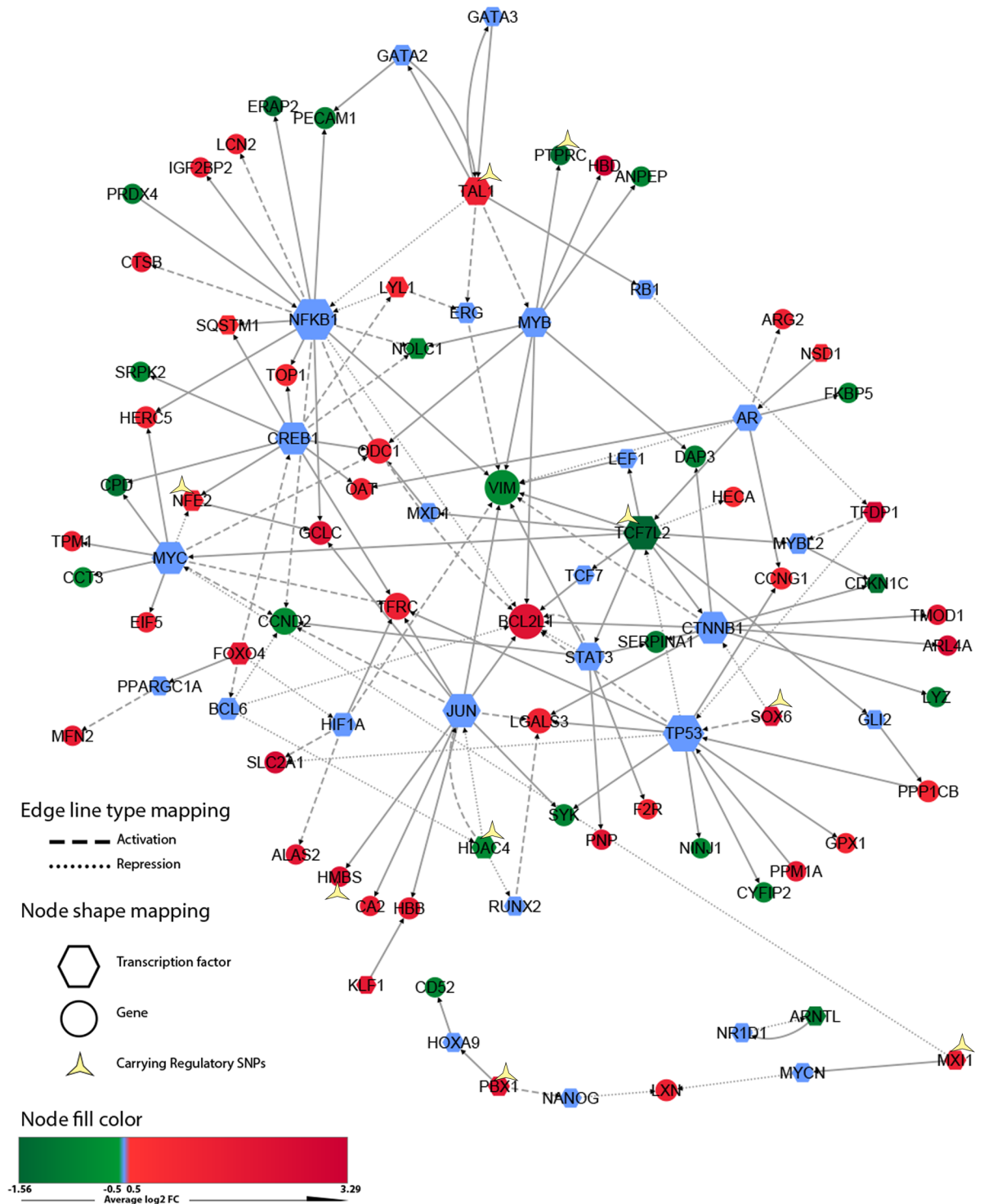
Fig 4. Transcriptional regulatory subnetwork based on microarray meta-analysis. Regulatory network analysis was performed using RNEA R/package to determine the regulation complexes upstream of DEGs identified in the meta-analysis. Genes carrying in their promoter, a significant regulatory SNPs are marked by a yellow star. Each node represents a DEG or enriched transcription factor, depending on their shapes. The node size indicate greater significance of the enrichment. The edges reflect the relationships between the nodes.

https://doi.org/10.1371/journal.pone.0199461.g004

Fig 5. Over-representation of pathways and GO categories in biological networks identified by the meta-analysis. Network representation of an enriched pathway integrating biological processes on the DEG list according to the ClueGO Cytoscape plugin. Hypergeometric (right-handed) enrichment distribution tests were conducted with a p value significance level of ≤0.05, followed by the Bonferroni adjustment for the terms, and thus leading term groups were selected based on the highest significance. Each node represents a biological process. The edges reflect the relationships between the terms based on the similarity of their associated genes. The node size and deeper color indicate greater significance of the enrichment.

Nevertheless, the mechanisms that determine the appearance of the inflammatory response and its persistence (which exacerbates the symptoms of SCD patients) and the factors involved still need to be clarified.

Here, we set up a full complete 'omics bioinformatics pipeline to further delineate the relation between SCD patients' transcriptomics profile and the pathophysiology of the disease focusing especially on the inflammatory response.

Several individual microarray-based gene expression studies related to SCD have been published in recent years and were made available for public reuse. On the other hand, the small number of individuals analyzed in each study is an obstacle for the identification of a consensus set of DEGs. A meta-analysis of the available data should reduce study bias, increase statistical power, and improve the overall biological-process understanding [57,58], which may lead to therapeutic-target discovery [59,60].

In the present study, we not only increased the sample size of individuals subjected to the meta-analysis but also overcame the heterogeneity of the disease expression across both SCD patients and individual datasets by means of a bioinformatics approach that takes into account the between-study and within-study variation.

Our meta-analysis, using all public eligible SCD raw data [7,33–36] produced a consensus shared DEG profile. Indeed, we identified 335 DEGs, which include 224 upregulated and 111 downregulated genes having FDR < 0.05. Among these, we found the already known set of

genes previously reported to be associated with SCD, e.g., *KLF1* (Krüppel-like factor 1) [61,62] and *HBD* (hemoglobin subunit delta), as well as more than 130 candidate genes found to be differentially expressed by a preceding meta-analysis despite a different statistical framework that was applied to only two datasets [63].

The gene set enrichment and pathway analysis of our DEG set revealed the importance of **innate immunity** (regulation and activation of innate immune response "GO:0002218, GO:0045088," innate immune system "R-HSA-168249," interferon alpha/beta signaling "R-HSA-909733," the type I interferon signaling pathway "GO:0060337," and the IL-6 signaling pathway, *Homo sapiens* "WP364"), **hemostasis** (blood coagulation "GO:0007596," platelet activation "hsa04611," and positive regulation of the MAPK cascade "GO:0043410"), a **response to stress** (oxidative stress "WP408," glutathione metabolism "WP100") and **hemopoiesis** (erythrocyte development "GO:0048821," myeloid cell development "GO:0061515") pathways. **Heme biosynthesis** and **apoptosis** were also found among the significant biological terms (a complete list is provided in S6 Table).

The results of the PP interaction analysis supported those obtained by the gene enrichment analysis. Indeed, among the top predicted hub genes sorted by degree and betweenness centrality, we found genes that play a prominent role in innate immunity pathways (*HERC5*, *HERC6*), hemopoiesis (*AHSP*), and heme biosynthesis (*ALAS2*). The latter (5'-aminolevulinate synthase) encodes a major regulatory enzyme in erythrocytes [64]. This protein is the rate-limiting enzyme in heme synthesis and contributes to regulation of heme (iron-containing molecule) synthesis preventing free-iron accumulation and subsequent organ damage [65].

Our findings showing the increase in *ALAS2* transcription (logFC = 1.9) are in agreement with the previously reported higher heme concentration measured in SCD patients' groups (steady state and crisis) when compared with healthy individuals [66]. This high heme level has been shown to induce inflammation and to increase vascular permeability, adhesion molecule expression, and leukocyte recruitment [67,68] and has been suggested to contribute to severe clinical manifestations of SCD [69]. Accordingly, we did not notice either the transcription of heme oxygenase gene (*HMOX1*; coding for the enzyme regulating catalytic cleavage of heme groups) or the expression of the master TF *NRF2* (its protein product regulates the transcription of most of antioxidant genes). This observation is in line with another study showing that a large proportion of SCD patients have a relatively modest HMOX1 plasma concentration due in part to a hyporesponsive *HMOX1* promoter [70]. Nevertheless, our results revealed transcriptional induction of other antioxidant genes such as *GPX1*, *GCLC*, *CAT*, *PINK1*, *SESN3*, *UBQLN1*, *CD36*, and *SNCA*.

Carbon monoxide (CO), one of the products of HMOX1, has been widely studied for its therapeutic potential [71–75]. Indeed, administration of CO and biliverdin inhibits vascular inflammation and vaso-occlusion in mouse models of SCD [72]. Furthermore, HMOX1 induction has had a beneficial effect in several pathological conditions [76] including chronic nephropathy [77]: an SCD-related pathology. HMOX1 inducers may thus help to decrease heme concentration and consequently reduce the inflammatory responses that worsen the patients' symptoms; i.e., HMOX1 inducers may be beneficial to SCD patients.

For a better understanding of the transcriptional mechanisms that regulate our DEGs, we constructed a gene-regulatory subnetwork associated with SCD (**Fig 4**). According to the network topology analysis, the nuclear factor-kappa B (NF-κB) remained the top hub-enriched TF. NF-κB regulates several physiological responses, including transcription of a large set of proinflammatory-cytokine genes [78] usually upregulated in SCD [53–55]. Our results indicate the increased transcription of various proinflammatory-cytokine genes including *IL-6*, *IL-12*, and *IL-18* and cytokine receptor genes such as *IL-17RD* and *IL-17RC* even if the increased expression of these transcripts was below the cutoff.
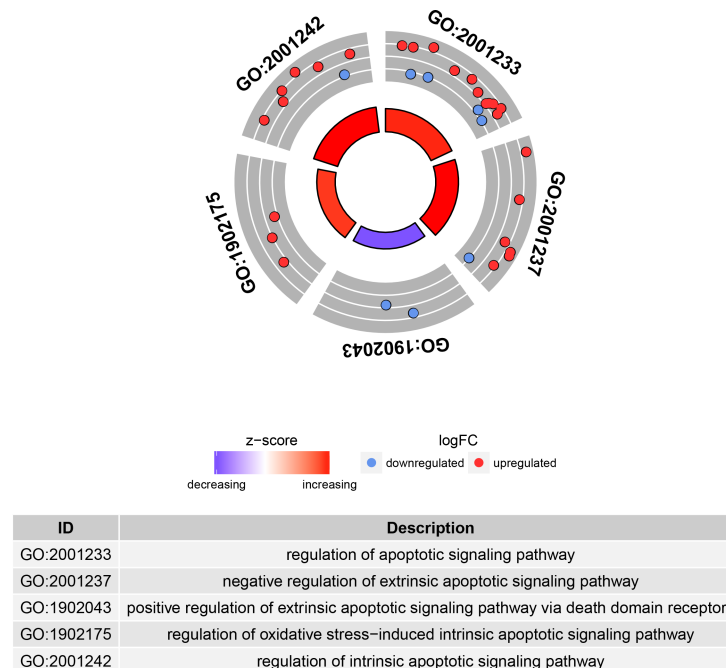
Fig 6. Downregulated and upregulated DEGs involved in apoptosis pathways. The inner ring is a bar plot where the height of the bar indicates the significance of the term ($-\log_{10}$ adjusted p value), and color corresponds to the z-score. The outer ring displays scatterplots of the expression levels (log2FC) for the genes in each term.

https://doi.org/10.1371/journal.pone.0199461.g006

At the same time, using GWAS data, from the identified polymorphisms we extracted the rSNPs occurring in the promoter region of our DEGs and gained an important insight into the regulatory complexes governing these gene expression patterns. Our analysis revealed the occurrence of 60 rSNPs in promoter DNA sequences of 20 genes (S4 Table), including *NFE2*, *HDAC4*, *PI3Kδ*, and *PI3Kγ*.

Among the top significant enriched GO terms and biological pathways associated with the subset of these 20 candidate disease-associated genes were those primarily related to T-cell activation/differentiation; alpha-beta, gamma-delta, and regulatory T-cell differentiation; T-cell lineage commitment; and B-cell proliferation or activation. This result is highly consistent with the alteration of a lymphocyte count [79], phenotype, and function associated with SCD [80]. This alteration in the acquired immunity may explain the increased risk of severe bacterial infections among these patients [81,82]. Indeed, different studies have shown that SCD patients have a decreased T-lymphocyte count and increased B-lymphocyte count either at baseline or in acute crisis [83]. Of note, T-cell lymphopenia–associated inflammatory responses have been previously linked to the inactivation of *PI3Kδ* and *PI3Kγ* [84], genes of two PI3K isoforms that we found to be significantly downregulated and to carry rSNPs (log$_2$FC: -0.94 and -0.56, respectively) in this analysis. This finding suggests that the downregulation of these kinase genes, predominantly expressed in the hematopoietic system [85], may be implicated in the profound lymphopenia observed in SCD patients.

Our gene-regulatory network also showed significant overexpression of the TF nuclear factor-erythroid 2 (*NFE2*). The expression of this TF is activated by CREB1 and repressed by MYC TFs. Promoter analysis revealed that *NFE2* is affected by five rSNPs including rs35702801 and rs34155291 (S7 Table) that are predicted to disrupt several MYC motifs but not to affect the affinity of CREB1 binding. These rSNPs by altering TF MYC binding may hence explain the overexpression of *NFE2* observed in SCD patients. Some studies have shown

that elevated expression of *NFE2* modulates proinflammatory cytokine IL-8 expression in myeloid cells [86] and causes a high neutrophil count in a murine model [87]. We suggest here that overexpression of *NFE2* may contribute to the neutrophilia associated with the severity of SCD. This neutrophilia may be—as suggested by others—accentuated by the alteration of neutrophil apoptosis, which leads to neutrophil accumulation in blood [88]. Accordingly, our results reveal downregulation of the DEGs involved in the positive ("GO:1902043") and negative ("GO:2001237") regulation of the extrinsic apoptotic signaling pathway ([Fig 6]).

Additionally, the GWAS revealed a nonsynonymous SNP occurring in *CAMKK2*'s protein-coding region (S5 Table), transforming arginine to cysteine at position 363 present in all CAMKK2 protein isoforms including the active isoforms 1, 2, and 3. Thus, besides the down-regulation, the amino acid change affecting the catalytic domain was predicted by the SIFT analysis to be a deleterious SNP with a significant damage score (0.041). CAMKK2 belongs to a family of Ca^{2+}/CaM-dependent, highly versatile serine/threonine kinases that catalyze the phosphorylation of (and have high affinity for) mainly three substrates: CAMK1, CAMK4, and AKT. CAMKK2 was reported to be exclusively expressed in the myeloid lineage [89] and participates in the regulation of several relevant physiological and pathophysiological processes, including hematopoiesis, cancer, inflammation, and immune responses. It was recently reported that under stressful conditions in immune cells, the loss of CAMKK2 promotes differentiation (both in bone marrow and peripheral-blood cells) of common myeloid progenitors toward granulocyte and monocyte precursor cells rather than megakaryocytes or erythrocytes [89]. This nonsynonymous SNP occurring in *CAMKK2*'s protein-coding region may thus also contribute to the increase in the neutrophil number (the most abundant type of granulocytes). Yet, the granulopoiesis state can be reversed by the re-expression of *CAMKK2*. Further research is needed to determine whether the CAMKK2 protein present in the blood of SCD patients shows an altered activity.

In conclusion, besides previously reported genes, our meta-signature contains a new subset of genes identified for the first time as associated with SCD. Among these, *ALAS2* belonging to heme metabolism is well known to be associated with SCD, as are PI3K delta and gamma isoforms involved in T-cell differentiation. The present analysis yielded a large database of rSNPs and candidate genes that may be helpful for future studies dealing with the pathogenesis and complexity of SCD.

Supporting information

S1 Table. Preferred reporting items for systematic reviews and meta-analyses (PRISMA). Guidelines Checklist.
(PDF)

S2 Table. Complete list of differentially-expressed genes obtained from the meta-analysis.
(XLSX)

S3 Table. Top 10 hub genes.
(XLSX)

S4 Table. The rSNPs located in the promoter region and altering TF-binding sites' affinity identified by the atSNP R package.
(XLSX)

S5 Table. Evaluation of the functional impact of the coding variants in the affected candidate gene proteins using Sift software.
(XLSX)

S6 Table. Gene set enrichment analyses of biological processes and metabolic pathways using multiple tools.
(XLSX)

S7 Table. rSNPs located in the promoter region of NFE2 and altering MYC-binding sites' affinity.
(XLSX)

Hugh Patterton University of the Free State, Bloemfontein, South Africa 9300;

Özlem Tastan Bishop, David Brown from Research Unit in Bioinformatics, Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, South Africa 6140;

Nicki Tiffin, Peter Van Heusden, Hocine Bendou, Alan Christoffels, Jean-Baka Domelevo Entfellner, Junaid Gamieldien, Mahjoubeh Jalali Sefid Dashti, Junaid Gamieldien from South African National Bioinformatics Institute/Medical Research Council of South Africa Bioinformatics Unit, University of the Western Cape, Cape Town, South Africa 7530;

Gustavo A. Salazar, Gerrit Botha, Emile Chimusa from Computational Biology Group, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa 7925;

Winston Hide, Shannan Ho Sui from Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA/Sheffield Institute for Translational Neuroscience, Department of Neuroscience, University of Sheffield, Sheffield S10 2HQ, United Kingdom;

Yasmina Jauferally Fakim, Zahra Mungloo-Dilmohammud and Shakuntala Baichoo from University of Mauritius, Reduit, Mauritius 80837;

C. Victor Jongeneel, Suresh Maslamoney, Dunfunk Mugutso, Victoria Nembaware from National Center for Supercomputing Applications and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA;

Samar Kassim from Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Ain Shams University, Cairo, Egypt 11566;

Ahmed Mansour Alzohairy from Genetics Department, Faculty of Agriculture, Zagazig University, Sharkia, Egypt 44519;

Jonathan Kayondo from Uganda Virus Research Institute, Entebbe, Uganda 31301;

Sylvester Lyantagaye from University of Dar es Salaam, Dar es Salaam, Tanzania, 73;

Daniel Masiga, Raphael Isokpehi, Benard Kulohoma and Anne Fischer from International Centre of Insect Physiology and Ecology, Nairobi, Kenya 00100;

Odile Ouwe Missi Oukem-Boyer from Centre de Recherche Medicale et Sanitaire, Niamey, Niger, BP 10887;

Ellis Owusu-Dabo Kumasi Centre for Collaborative Research in Tropical Medicine/Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, PMB;

Nzovu Ulenga from Management and Development for Health, Dar es Salaam, Tanzania, 61;

Flora Elias Mlyango from University of Dar es Salaam, Dar es Salaam, Tanzania, 73;

Bruno Mmbando from Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania 00255;

Mohamed Alibi, Amel Ghouila and Oussema Souiai from Institute Pasteur of Tunis, Tunis, Tunisia 1002;

Gloria Rendon, Chris Fields, Dan Lapine and Liudmila Sergeevna Mainzer from National Center for Supercomputing Applications and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA;

Samson Pandam Salifu from Kumasi Centre for Collaborative Research in Tropical Medicine/Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, PMB.

Author Contributions

Conceptualization: Julie Makani, Kais Ghedira.

Data curation: Cherif Ben Hamda, Liberata Mwita, Siana Nkya.

Formal analysis: Cherif Ben Hamda, Raphael Sangeda, Ayton Meintjes, Kais Ghedira.

Funding acquisition: Nicola Mulder, Julie Makani.

Investigation: Cherif Ben Hamda, Kais Ghedira.

Methodology: Cherif Ben Hamda, Raphael Sangeda, Ayton Meintjes, Siana Nkya, Kais Ghedira.

Project administration: Nicola Mulder.

Resources: Siana Nkya.

Supervision: Kais Ghedira.

Validation: Cherif Ben Hamda, Siana Nkya.

Visualization: Cherif Ben Hamda.

Writing – original draft: Cherif Ben Hamda, Raphael Sangeda, Liberata Mwita, Lamia Guizani-Tabbane, Alia Benkahla, Kais Ghedira.

Writing – review & editing: Cherif Ben Hamda, Raphael Sangeda, Liberata Mwita, Sumir Panji, Nicola Mulder, Lamia Guizani-Tabbane, Alia Benkahla, Julie Makani, Kais Ghedira.

References

1. Orphanet [Internet]. [cited 2 Oct 2017]. Available: http://www.orpha.net/consor/cgi-bin/index.php

2. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Dewi M, et al. Global epidemiology of Sickle haemoglobin in neonates: A contemporary geostatistical model-based map and population estimates. Lancet. 2013; 381: 142–151. https://doi.org/10.1016/S0140-6736(12)61229-X PMID: 23103089

3. Baralle F. Complete nucleotide sequence of the 5' noncoding region of human alpha-and beta-globin mRNA. Cell. 1977; 12: 1085–1095. doi:0092-8674(77)90171-4 [pii] PMID: 597858

4. Browning JA, Staines HM, Robinson HC, Powell T, Ellory JC, Gibson JS. The effect of deoxygenation on whole-cell conductance of red blood cells from healthy individuals and patients with sickle cell disease. Blood. 2007; 109: 2622–2629. https://doi.org/10.1182/blood-2006-03-001404 PMID: 17138828

5. Chies J a, Nardi NB. Sickle cell disease: a chronic inflammatory condition. Med Hypotheses. 2001; 57: 46–50. https://doi.org/10.1054/mehy.2000.1310 PMID: 11421623

6. Solovey A, Gui L, Ramakrishnan S, Steinberg MH, Hebbel RP. Sickle cell anemia as a possible state of enhanced anti-apoptotic tone: survival effect of vascular endothelial growth factor on circulating and unanchored endothelial cells. Blood. 1999; 93: 3824–3830. Available: http://bloodjournal.hematologylibrary.org/content/93/11/3824.short%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/10339489 PMID: 10339489

7. Raghavachari N, Barb J, Yang Y, Liu P, Woodhouse K, Levy D, et al. A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. BMC Med Genomics. 2012; 5: 28. https://doi.org/10.1186/1755-8794-5-28 PMID: 22747986

8. Desai AA, Lei Z, Bahroos N, Maienschein-Cline M, Saraf SL, Zhang X, et al. Association of circulating transcriptomic profiles with mortality in sickle cell disease. Blood. 2017; 129: 3009–3016. https://doi.org/10.1182/blood-2016-11-752279 PMID: 28373264

9. Schulze a, Downward J. Navigating gene expression using microarrays—a technology review. Nat Cell Biol. 2001; 3: E190–E195. https://doi.org/10.1038/35087138 PMID: 11483980

10. Trevino V, Falciani F, Barrera-Saldana HA. DNA microarrays: a powerful genomic tool for biomedical and clinical research. Mol Med. 2007; 13: 527–541. https://doi.org/10.2119/2006-00107.Trevino PMID: 17660860

11. Li X, Quigg RJ, Zhou J, Gu W, Nagesh Rao P, Reed EF. Clinical utility of microarrays: current status, existing challenges and future outlook. Curr Genomics. 2008; 9: 466–474. https://doi.org/10.2174/138920208786241199 PMID: 19506735

12. Rung J, Brazma A. Reuse of public genome-wide gene expression data. Nat Rev Genet. 2013; 14: 89–99. https://doi.org/10.1038/nrg3394 PMID: 23269463

13. Choi JK, Yu U, Kim S, Yoo OJ. Combining multiple microarray studies and modeling interstudy variation. Bioinformatics. 2003. https://doi.org/10.1093/bioinformatics/btg1010

14. Smid M, Dorssers LCJ, Jenster G. Venn Mapping: Clustering of heterologous microarray data based on the number of co-occurring differentially expressed genes. Bioinformatics. 2003; 19: 2065–2071. https://doi.org/10.1093/bioinformatics/btg282 PMID: 14594711

15. Parmigiani G, Garrett-Mayer ES, Anbazhagan R, Gabrielson E. A cross-study comparison of gene expression studies for the molecular classification of lung cancer. Clin Cancer Res. 2004; 10: 2922–2927. https://doi.org/10.1158/1078-0432.CCR-03-0490 PMID: 15131026

16. DeConde RP, Hawley S, Falcon S, Clegg N, Knudsen B, Etzioni R. Combining Results of Microarray Experiments: A Rank Aggregation Approach. Stat Appl Genet Mol Biol. 2006; 5. https://doi.org/10.2202/1544-6115.1204 PMID: 17049026

17. Zintzaras E, Ioannidis JPA. Meta-analysis for ranked discovery datasets: Theoretical framework and empirical demonstration for microarrays. Comput Biol Chem. 2008; 32: 38–46. https://doi.org/10.1016/j.compbiolchem.2007.09.003 PMID: 17988949

18. Wang X, Kang DD, Shen K, Song C, Lu S, Chang LC, et al. An r package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. Bioinformatics. 2012; 28: 2534–2536. https://doi.org/10.1093/bioinformatics/bts485 PMID: 22863766

19. Xia J, Fjell CD, Mayer ML, Pena OM, Wishart DS, Hancock REW. INMEX—a web-based tool for integrative meta-analysis of expression data. Nucleic Acids Res. 2013; 41. https://doi.org/10.1093/nar/gkt338 PMID: 23766290

20. Xia J, Gill EE, Hancock REW. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc. 2015; 10: 823–844. https://doi.org/10.1038/nprot.2015.052 PMID: 25950236

21. Mayne BT, Bianco-Miotto T, Buckberry S, Breen J, Clifton V, Shoubridge C, et al. Large scale gene expression meta-analysis reveals tissue-specific, sex-biased gene expression in humans. Front Genet. 2016; 7. https://doi.org/10.3389/fgene.2016.00183 PMID: 27790248

22. Jalili M, Salehzadeh-Yazdi A, Mohammadi S, Yaghmaie M, Ghavamzadeh A, Alimoghaddam K. Meta-analysis of gene expression profiles in acute promyelocytic leukemia reveals involved pathways. Int J Hematol Stem Cell Res. 2017; 11: 1–12.

23. Jha PK, Vijay A, Sahu A, Ashraf MZ. Comprehensive Gene expression meta-analysis and integrated bioinformatic approaches reveal shared signatures between thrombosis and myeloproliferative disorders. Sci Rep. 2016; 6: 37099. https://doi.org/10.1038/srep37099 PMID: 27892526

24. Likhitrattanapisal S, Tipanee J, Janvilisri T. Meta-analysis of gene expression profiles identifies differential biomarkers for hepatocellular carcinoma and cholangiocarcinoma. Tumor Biol. 2016; 37: 12755–12766. https://doi.org/10.1007/s13277-016-5186-8 PMID: 27448818

25. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol. 2012; 8. https://doi.org/10.1371/journal.pcbi.1002822 PMID: 23300413

26. Conde L, Bracci PM, Richardson R, Montgomery SB, Skibola CF. Integrating GWAS and expression data for functional characterization of disease-associated SNPs: An application to follicular lymphoma. Am J Hum Genet. 2013; 92: 126–130. https://doi.org/10.1016/j.ajhg.2012.11.009 PMID: 23246294

27. Gorlov IP, Gallick GE, Gorlova OY, Amos C, Logothetis CJ. GWAS meets microarray: Are the results of genome-wide association studies and gene-expression profiling consistent? Prostate cancer as an example. PLoS One. 2009; 4. https://doi.org/10.1371/journal.pone.0006511 PMID: 19652704

28. Becker K, Siegert S, Toliat MR, Du J, Casper R, Dolmans GH, et al. Meta-analysis of genome-wide association studies and network analysis-based integration with gene expression data identify new suggestive loci and unravel a Wnt-centric network associated with Dupuytren's disease. PLoS One. 2016; 11. https://doi.org/10.1371/journal.pone.0158101 PMID: 27467239

29. Jin H-J, Jung S, DebRoy AR, Davuluri R V. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. Oncotarget. 2016; 7: 54616–54626. https://doi.org/10.18632/oncotarget.10520 PMID: 27409348

30. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (80-). 2012; 337: 1190–1195. https://doi.org/10.1126/science.1222794 PMID: 22955828

31. Pai AA, Pritchard JK, Gilad Y. The Genetic and Mechanistic Basis for Variation in Gene Regulation. PLoS Genet. 2015; 11. https://doi.org/10.1371/journal.pgen.1004857 PMID: 25569255

32. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009; 339: b2535–b2535. https://doi.org/10.1136/bmj.b2535 PMID: 19622551

33. Raghavachari N, Xu X, Harris A, Villagra J, Logun C, Barb J, et al. Amplified expression profiling of platelet transcriptome reveals changes in arginine metabolic pathways in patients with sickle cell disease. Circulation. 2007; 115: 1551–1562. https://doi.org/10.1161/CIRCULATIONAHA.106.658641 PMID: 17353439

34. Raghavachari N, Xu X, Munson PJ, Gladwin MT. Characterization of whole blood gene expression profiles as a sequel to globin mRNA reduction in patients with sickle cell disease. PLoS One. 2009; 4. https://doi.org/10.1371/journal.pone.0006484 PMID: 19649296

35. Van Beers EJ, Yang Y, Raghavachari N, Tian X, Allen DT, Nichols JS, et al. Iron, inflammation, and early death in adults with sickle cell disease. Circ Res. 2015; 116: 298–306. https://doi.org/10.1161/CIRCRESAHA.116.304577 PMID: 25378535

36. Quinlan J, Idaghdour Y, Goulet J-P, Gbeha E, de Malliard T, Bruat V, et al. Genomic architecture of sickle cell disease in West African children. Front Genet. 2014; 5: 26. https://doi.org/10.3389/fgene.2014.00026 PMID: 24592274

37. Kang DD, Sibille E, Kaminski N, Tseng GC. MetaQC: Objective quality control and inclusion/exclusion criteria for genomic meta-analysis. Nucleic Acids Res. 2012; 40. https://doi.org/10.1093/nar/gkr1071 PMID: 22116060

38. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat Appl Genet Mol Biol. 2004; 3: 1–25. https://doi.org/10.2202/1544-6115.1027 PMID: 16646809

39. Lu S, Li J, Song C, Shen K, Tseng GC. Biomarker detection in the integration of multiple multi-class genomic studies. Bioinformatics. 2009; 26: 333–340. https://doi.org/10.1093/bioinformatics/btp669 PMID: 19965884

40. Krull M, Voss N, Choi C, Pistor S, Potapov A, Wingender E. TRANSPATH??: An integrated database on signal transduction and a tool for array analysis. Nucleic Acids Research. 2003. pp. 97–100. https://doi.org/10.1093/nar/gkg089

41. geneXplain. Available: http://platform.genexplain.com/bioumlweb/

42. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013; 14: 128. https://doi.org/10.1186/1471-2105-14-128 PMID: 23586463

43. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009; 25: 1091–1093. https://doi.org/10.1093/bioinformatics/btp101 PMID: 19237447

44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498–2504. https://doi.org/10.1101/gr.1239303 PMID: 14597658

45. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43: D447–D452. https://doi.org/10.1093/nar/gku1003 PMID: 25352553

46. Chouvardas P, Kollias G, Nikolaou C. Inferring active regulatory networks from gene expression data using a combination of prior knowledge and enrichment analysis. BMC Bioinformatics. 2016; 17 Suppl 5: 181. https://doi.org/10.1186/s12859-016-1040-7 PMID: 27295045

47. Mtatiro SN, Singh T, Rooks H, Mgaya J, Mariki H, Soka D, et al. Genome wide association study of fetal hemoglobin in sickle cell Anemia in Tanzania. PLoS One. 2014; 9. https://doi.org/10.1371/journal.pone.0111464 PMID: 25372704

48. Liu J, Walter E, Stenger D, Thach D. Effects of Globin mRNA Reduction Methods on Gene Expression Profiles from Whole Blood. J Mol Diagnostics. 2006; 8: 551–558. https://doi.org/10.2353/jmoldx.2006.060021 PMID: 17065423

49. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. The Lancet. 2010. pp. 2018–2031. https://doi.org/10.1016/S0140-6736(10)61029-X

50. Hebbel RP. Ischemia-reperfusion injury in sickle cell anemia: Relationship to acute chest syndrome, endothelial dysfunction, arterial vasculopathy, and inflammatory pain. Hematology/Oncology Clinics of North America. 2014. pp. 181–198. https://doi.org/10.1016/j.hoc.2013.11.005 PMID: 24589261

51. Hoppe CC. Inflammatory mediators of endothelial injury in sickle cell disease. Hematology/Oncology Clinics of North America. 2014. pp. 265–286. https://doi.org/10.1016/j.hoc.2013.11.006 PMID: 24589266

52. Zhang D, Xu C, Manwani D, Frenette PS. Neutrophils, platelets, and inflammatory pathways at the nexus of sickle cell disease pathophysiology. Blood. 2016. pp. 801–809. https://doi.org/10.1182/blood-2015-09-618538 PMID: 26758915

53. Lanaro C, Franco-Penteado CF, Albuqueque DM, Saad STO, Conran N, Costa FF. Altered levels of cytokines and inflammatory mediators in plasma and leukocytes of sickle cell anemia patients and effects of hydroxyurea therapy. J Leukoc Biol. 2009; 85: 235–242. https://doi.org/10.1189/jlb.0708445 PMID: 19004988

54. Sarray S, Saleh LR, Lisa Saldanha F, Al-Habboubi HH, Mahdi N, Almawi WY. Serum IL-6, IL-10, and TNF?? levels in pediatric sickle cell disease patients during vasoocclusive crisis and steady state condition. Cytokine. 2015; 72: 43–47. https://doi.org/10.1016/j.cyto.2014.11.030 PMID: 25569375

55. Qari MH, Dier U, Mousa SA. Biomarkers of inflammation, growth factor, and coagulation activation in patients with sickle cell disease. Clin Appl Thromb Hemost. 2012; 18: 195–200. https://doi.org/10.1177/1076029611420992 PMID: 21949038

56. Conran N, Franco-Penteado CF, Costa FF. Newer aspects of the pathophysiology of sickle cell disease vaso-occlusion. Hemoglobin. 2009. pp. 1–16. https://doi.org/10.1080/03630260802625709 PMID: 19205968

57. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Medicine. 2008. pp. 1320–1332. https://doi.org/10.1371/journal.pmed.0050184 PMID: 18767902

58. Tseng GC, Ghosh D, Feingold E. Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Research. 2012. pp. 3785–3799. https://doi.org/10.1093/nar/gkr1265 PMID: 22262733

59. Kodama K, Toda K, Morinaga S, Yamada S, Butte AJ. Anti-CD44 antibody treatment lowers hyperglycemia and improves insulin resistance, adipose inflammation, and hepatic steatosis in diet- induced obese mice. Diabetes. 2015; 64: 867–875. https://doi.org/10.2337/db14-0149 PMID: 25294945

60. Kodama K, Horikoshi M, Toda K, Yamada S, Hara K, Irie J, et al. Expression-based genome-wide association study links the receptor CD44 in adipose tissue with type 2 diabetes. Proc Natl Acad Sci U S A. 2012; 109: 7049–54. https://doi.org/10.1073/pnas.1114513109 PMID: 22499789

61. Peralta R, Low A, Booten S, Zhou D, Kim A, Freier S, et al. Targeting KLF1 for the Treatment of Sickle Cell Disease Using Antisense Oligonucleotides [Internet]. Blood. 2014. Available: http://www.bloodjournal.org/content/124/21/4038.abstract

62. Gallienne AE, Dréau HMP, Schuh A, Old JM, Henderson S. Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. Haematologica. 2012; 97: 340–434. https://doi.org/10.3324/haematol.2011.055442 PMID: 22102705

63. Hounkpe BW, Fiusa MML, Colella MP, da Costa LNG, Benatti R de O, Saad STO, et al. Role of innate immunity-triggered pathways in the pathogenesis of Sickle Cell Disease: a meta-analysis of gene expression studies. Sci Rep. 2015; 5: 17822. https://doi.org/10.1038/srep17822 PMID: 26648000

64. Chiabrando D, Mercurio S, Tolosano E. Heme and erythropoieis: More than a structural role. Haematologica. 2014. pp. 973–983. https://doi.org/10.3324/haematol.2013.091991 PMID: 24881043

65. Steinbicker AU, Muckenthaler MU. Out of balance-systemic iron homeostasis in iron-related disorders. Nutrients. 2013. pp. 3034–3061. https://doi.org/10.3390/nu5083034 PMID: 23917168

66. C M.O.S., R L.C., R J.H.O., DAS T., DN V.M.L., C M.B., et al. Heme concentration as a biomarker of sickle cell disease severity: Its role in steady-state and in crisis patients [Internet]. Blood. 2015. p. 975. Available: http://www.bloodjournal.org/content/126/23/975%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed13&NEWS=N&AN=72171521

67. Wagener FADTG, Eggert A, Boerman OC, Oyen WJG, Verhofstad A, Abraham NG, et al. Heme is a potent inducer of inflammation in mice and is counteracted by heme oxygenase. Blood. 2001; 98: 1802–1811. https://doi.org/10.1182/blood.V98.6.1802 PMID: 11535514

68. Belcher JD, Beckman JD, Balla G, Balla J, Vercellotti G. Heme degradation and vascular injury. Antioxid Redox Signal. 2010; 12: 233–48. https://doi.org/10.1089/ars.2009.2822 PMID: 19697995

69. Sawicki KT, Shang M, Wu R, Chang HC, Khechaduri A, Sato T, et al. Increased Heme Levels in the Heart Lead to Exacerbated Ischemic Injury. J Am Heart Assoc. 2015; 4: e002272. https://doi.org/10.1161/JAHA.115.002272 PMID: 26231844

70. Owusu BY, Hu Y, Ghosh S, Tan F, Sey F, Ekem I, et al. Determinants Of Heme-Oxygenase-1 Upregulation In Patients With Sickle Cell Disease. Adisa O, editor. Blood. 2013; 122: 2235 LP–2235.

71. Belcher JD, Young M, Chen C, Nguyen J, Burhop K, Tran P, et al. MP4CO, a pegylated hemoglobin saturated with carbon monoxide, is a modulator of HO-1, inflammation, and vaso-occlusion in transgenic sickle mice. Blood. 2013; 122: 2757–2764. https://doi.org/10.1182/blood-2013-02-486282 PMID: 23908468

72. Belcher JD, Mahaseth H, Welch TE, Otterbein LE, Hebbel RP, Vercellotti GM. Heme oxygenase-1 is a modulator of inflammation and vaso-occlusion in transgenic sickle mice. J Clin Invest. 2006; 116: 808–816. https://doi.org/10.1172/JCI26857 PMID: 16485041

73. Kim KM, Pae H-O, Zheng M, Park R, Kim Y-M, Chung H-T. Carbon monoxide induces heme oxyge-nase-1 via activation of protein kinase R-like endoplasmic reticulum kinase and inhibits endothelial cell apoptosis triggered by endoplasmic reticulum stress. Circ Res. 2007; 101: 919–27. https://doi.org/10.1161/CIRCRESAHA.107.154781 PMID: 17823375

74. Araujo JA, Zhang M, Yin F. Heme oxygenase-1, oxidation, inflammation, and atherosclerosis. Frontiers in Pharmacology. 2012. https://doi.org/10.3389/fphar.2012.00119 PMID: 22833723

75. Araujo JA. HO-1 and CO: Fighters vs sickle cell disease? Blood. 2013. pp. 2535–2536. https://doi.org/10.1182/blood-2013-08-521922 PMID: 24113796

76. Liu X, Peyton KJ, Ensenat D, Wang H, Schafer AI, Alam J, et al. Endoplasmic reticulum stress stimu-lates heme oxygenase-1 gene expression in vascular smooth muscle. Role in cell survival. J Biol Chem. 2005; 280: 872–7. https://doi.org/10.1074/jbc.M410413200 PMID: 15546873

77. Nath K a, Grande JP, Haggard JJ, Croatt a J, Katusic ZS, Solovey a, et al. Oxidative stress and induc-tion of heme oxygenase-1 in the kidney in sickle cell disease. Am J Pathol. 2001; 158: 893–903. https://doi.org/10.1016/S0002-9440(10)64037-0 PMID: 11238038

78. Pahl HL. Activators and target genes of Rel/NF-kappaB transcription factors. Oncogene. 1999; 18: 6853–66. https://doi.org/10.1038/sj.onc.1203239 PMID: 10602461

79. Vingert B, Tamagne M, Desmarets M, Pakdaman S, Elayeb R, Habibi A, et al. Partial dysfunction of Treg activation in sickle cell disease. Am J Hematol. 2014; 89: 261–266. https://doi.org/10.1002/ajh.23629 PMID: 24779034

80. Balandya E, Reynolds T, Obaro S, Makani J. Alteration of lymphocyte phenotype and function in sickle cell anemia: Implications for vaccine responses. American Journal of Hematology. 2016. pp. 938–946. https://doi.org/10.1002/ajh.24438 PMID: 27237467

81. Makani J, Mgaya J, Balandya E, Msami K, Soka D, Cox SE, et al. Bacteraemia in sickle cell anaemia is associated with low haemoglobin: A report of 890 admissions to a tertiary hospital in Tanzania. Br J Haematol. 2015; 171: 273–276. https://doi.org/10.1111/bjh.13553 PMID: 26084722

82. Ramakrishnan M, Mo??si JC, Klugman KP, Iglesias JMF, Grant LR, Mpoudi-Etame M, et al. Increased risk of invasive bacterial infections in African people with sickle-cell disease: A systematic review and meta-analysis. The Lancet Infectious Diseases. 2010. pp. 329–337. https://doi.org/10.1016/S1473-3099(10)70055-4 PMID: 20417415

83. Glassman AB, Deas DV, Berlinsky FS BC. Lymphocyte blast transformation and peripheral lymphocyte percentages in patients with sickle cell disease. Ann O F Clin L A N D Lab SC IE N C E. 1980;Vol. 10. Available: https://www.ncbi.nlm.nih.gov/pubmed/?term=PMID%3A+6965833

84. Ji H, Rintelen F, Waltzinger C, Bertschy Meier D, Bilancio A, Pearce W, et al. Inactivation of PI3K-gamma and PI3Kdelta distorts T-cell development and causes multiple organ inflammation. Blood. 2007; 110: 2940–2947. https://doi.org/10.1182/blood-2007-04-086751 PMID: 17626838

85. Fruman DA, Cantley LC. Phosphoinositide 3-kinase in immunological systems. Semin Immunol. 2002; 14: 7–18. https://doi.org/10.1006/smim.2001.0337 PMID: 11884226

86. Wehrle J, Seeger TS, Schwemmers S, Pfeifer D, Bulashevska A, Pahl HL. Transcription factor nuclear factor erythroid-2 mediates expression of the cytokine interleukin 8, a known predictor of inferior out-come in patients with myeloproliferative Neoplasms. Haematologica. 2013; 98: 1073–1080. https://doi.org/10.3324/haematol.2012.071183 PMID: 23445878

87. Kaufmann KB, Gründer A, Hadlich T, Wehrle J, Gothwal M, Bogeska R, et al. A novel murine model of myeloproliferative disorders generated by overexpression of the transcription factor NF-E2. J Exp Med. 2012; 209: 35–50. https://doi.org/10.1084/jem.20110540 PMID: 22231305

88. Ozdogu H, Boga C, Sozer O, Sezgin N, Kizilkilic E, Maytalman E, et al. The apoptosis of blood polymor-phonuclear leukocytes in sickle cell disease. Cytom Part B—Clin Cytom. 2007; 72: 276–280. https://doi.org/10.1002/cyto.b.20160

89. Racioppi L, Noeldner PK, Lin F, Arvai S, Means AR. Calcium/calmodulin-dependent protein kinase kinase 2 regulates macrophage-mediated inflammatory responses. J Biol Chem. 2012; 287: 11579–91. https://doi.org/10.1074/jbc.M111.336032 PMID: 22334678