GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Screening for feature genes associated with hereditary hemochromatosis and functional analysis with DNA microarrays

J. Wang, X. Zhou, J. Zhao, Z. Li and X. Li

Department of Hematology, Institute of Surgery Research, Daping Hospital, Third Military Medical University, Chongqing, China

Corresponding author: J. Wang
E-mail: wangjinwwjjsh@hotmail.com

Genet. Mol. Res. 12 (4): 6240-6248 (2013)
Received December 18, 2012
Accepted June 5, 2013
Published December 4, 2013
DOI http://dx.doi.org/10.4238/2013.December.4.11

ABSTRACT. The aim of this study was to identify feature genes that are associated with hereditary hemochromatosis (HHC; iron overload) in cardiac and skeletal muscle of mice. First, the expression profile GSE9726 was downloaded from Gene Expression Omnibus database which included 12 samples. Then the differentially expressed genes (DEGs) were identified by R language. Furthermore, the KUPS software was used to identify relationships in interactions among common DEGs in the cardiac and skeletal muscles. We then used the EASE software to obtain enriched pathways in a gene interaction network. Finally, we used the plugins of the Cytoscape software, i.e., Mcode and Bingo, to conduct module analysis. By comparing diseased and normal tissue samples, 5 and 6 genes in the cardiac and skeletal muscles, respectively, were identified as DEGs. We observed that the *S100a8* and *S100a9* genes were common DEGs in both tissues examined. In addition, we constructed an interaction network with common DEGs and their interacting components, and identified *S100a8* and *S100a9* as being associated with immune responses. In view of the relationship between the early stage of myelodysplastic syndrome and the immune system, we hypothesize that

the expression of the *S100a8* and *S100a9* genes is a feature that can be used for diagnosis during the early stage of the myelodysplastic syndrome and that the 2 genes could be used as targets in treating this disease.

Key words: Myelodysplastic syndrome; Iron overload; Module analysis; Hereditary hemochromatosis; Functional enrichment analysis

INTRODUCTION

Myelodysplastic syndrome (MDS) is a heterogeneous clonal disease characterized by bone marrow (BM) myeloid cells dysplasia, ineffective hematopoiesis, and blood cytopenias. MDS also poses an increased risk of transformation to acute myeloid leukemia (AML) (Jadersten and Hellstrom-Lindberg 2009) -approximately 30% of MDS cases progress to AML. In the United States alone, the estimated number of MDS cases reported per year is about 40,000-76,000 with a gross medical cost of about US $30,000 per 44 persons per year (Beck et al. 2011). The genetic landscape of MDS remains largely unknown. Currently proposed models of the disease suggest gene mutations, deregulated gene expression, and epigenetic changes as key steps in MDS pathogenesis (Bejar et al., 2011).

Hereditary hemochromatosis (HHC) is a type of myelodysplastic anemia. It is a common genetic trait that increases the risk of developing systemic iron overload. The disorder is caused by mutations in genes encoding various proteins involved in the regulation of iron access to the blood (Majore et al., 2013). In HHC, iron progressively accumulates to toxic levels in most tissues, a process that is particularly damaging to the liver, heart, pancreas, and pituitary gland. If left untreated, the continuing iron deposition leads to arthritis, endocrine disorders (including delayed growth, impotence, and diabetes), cirrhosis of the liver, and heart failure. Death due to HHC is often the result of toxic effects due to the progressive iron overload resulting from sustained iron absorption and/or blood transfusions administered in order to treat the anemia (Pietrangelo, 2010). Significant changes in the symptoms of HHC have been observed in recent years, HHC is rarely fully expressed in clinical settings and the disorder displays a large phenotypic heterogeneity (Sredoja Tišma et al., 2012).

At present, microarrays are used as efficient tools to investigate and report gene expression patterns in biomedical studies and have been effectively used in phenotypic classification based on gene-expression profiles (GEPs). Due to the large number of probes/features (hundreds to thousands) deployed in a microarray, a critical step in the microarray-based discriminating analysis is feature selection (Arisi et al., 2011). Feature selection has been shown to effectively help in understanding the correlation between gene expression profiles and specific phenotypes (Xiong et al., 2001). The heart muscle is one of the specific tissues affected by iron overload, which results in cardiomyopathy in some HHC patients (Rodriguez et al., 2007). In this study, to better understand the mechanisms behind the pathological effects of iron overload on muscle cells, we performed a genome-wide expression analysis of genes in skeletal and heart muscle of mice with or without iron loading. We analyzed microarray data of samples from HHC mouse myocardium and skeletal muscle subjected to excessive iron load, to screen for marker genes of myelodysplastic anemia (iron overload). The results of this analysis may reveal novel links between iron overload and pathological manifestations of HHC, which may provide important information for treating HHC in the clinic.

MATERIAL AND METHODS

Oligonucleotide bead chip data

The chip dataset GSE9726 was downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/gds), which was obtained with the GPL4234 Sentrix Mouse-6 Expression BeadChip platform. The data set included 12 samples, including 4 normal heart and skeletal muscle samples and 4 heart and 4 skeletal muscle samples with iron overload.

Microarray analysis

Microarray analyses were performed by the Limma package in R language to identify differentially expressed genes (DEGs) in normal and iron-overload samples (Troyanskaya et al., 2001; Smyth et al., 2005; Fujita et al., 2006).

A $P < 0.05$ and a $|logFC| > 1$ were set as threshold of screening for DEGs and identifying common DEGs both in heart and skeletal muscle samples.

The KUPS software links to protein interactions in 3 different protein-interaction databases (IntAct, MINT, and HPRD). Therefore, KUPS was used to search for relationships among interactions of the common DEGs in the cardiac and skeletal muscles (see Figure 1) (Chen et al., 2011). The EASE software was then used to perform analysis of the predicted functions of the DEGs and to screen for enriched pathways of genes in the interaction network (Hosack et al., 2003) Finally, we used the plugins of the Cytoscape software (Shannon et al., 2003), Mcode and Bingo, to perform module analysis (Benza et al., 2010).

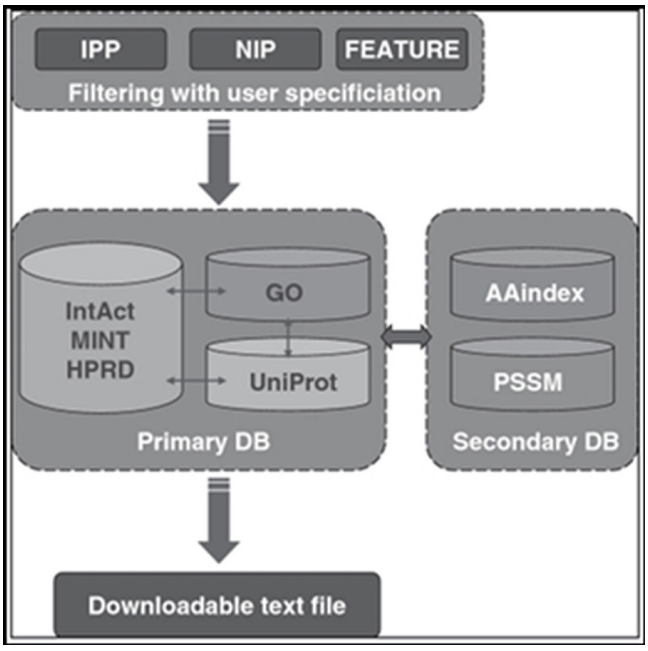

Figure 1. Flowchart of the KUPS software.

Statistical analyses

The mean values and standard deviations were calculated from technical triplicates in experiments. The Student *t*-test (unpaired and 2 tailed) was used to test for statistical significance of the differences in differential gene expression between control and iron-overload samples. A $P < 0.05$ and $|logFC|>1$ were set as thresholds of screening for DEGs.

RESULTS

Specific gene expression screening

After data pretreatment, the normalized data (Figure 2) of the gene expression profiles were used in comparative analysis, which identified 5 DEGs in myocardial muscle samples and 6 DEGs that met the difference threshold in skeletal muscle samples. Two genes, *S100a8* and *S100a9*, which were differentially expressed in both tissues, were identified as common DEGs (see Table 1).

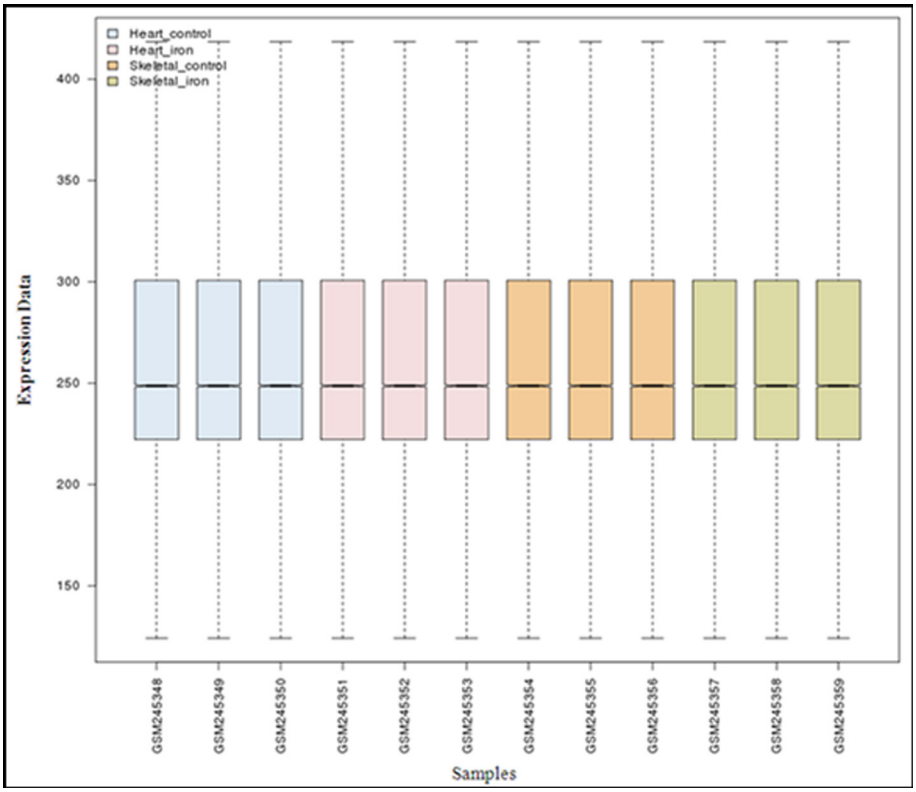

Figure 2. Gene expression value chart after standardizing. Blue, pink, orange, and green boxes indicate normal heart, iron overload heart, normal skeletal muscle and iron overload skeletal muscle samples. The black line in box is the median of every data. We can determine data standardization by its distribution. In this figure, the median line is almost on the same line, this indicate that standardization degree is very good.

Table 1. Differentially expressed gene list of heart muscle (a) and skeletal muscle (b) samples.

(a)			
ID	Gene symbol	P value	logFC
1090041	Dnajb1	2.43E-07	-1.30968
1850315	Fos	1.47E-05	-1.01542
70112	S100a8	5.71E-07	1.36072
7050528	S100a9	8.43E-08	1.101616
1230605	Slc25a25	2.91E-06	-1.20058
(b)			
ID	Gene symbol	P value	logFC
840538	Acta1	0.00009199	-1.25467
6760593	Angptl4	0.00005426	1.683493
1690017	Hsph1	0.00008896	-1.15994
6400300	Pdk4	0.00017326	1.187162
70112	S100a8	0.0040153	1.293363
7050528	S100a9	0.00273445	1.029499

The mean values and standard deviations were calculated from technical triplicates experiments. $P < 0.05$ and $|logFC|>1$ were regarded as significant.

Interaction prediction of the common DEGs

The KUPS software was used to search for interactions of the 2 common DEGs, *S100a8* and *S100a9*, with other genes, and an interaction network was constructed. As shown in Figure 3, the *S100a8/a9* genes interacted with *TLR4*, *RAGE*, *CD68*, *CD14*, and other genes.

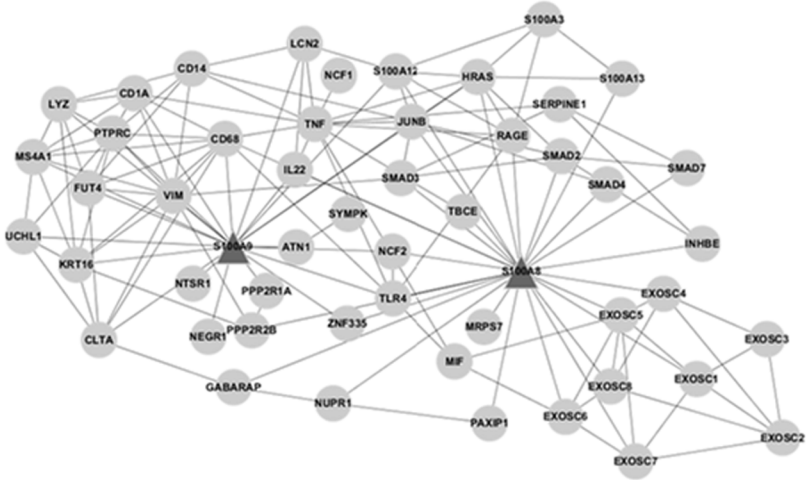

Figure 3. Interaction networks of common differences in gene expression between S100a8 and S100a9 predicted. In this figure, triangle nodes represent the *S100a8* and *S100a9* genes, the other round nodes represent the interactive objects predicted by software.

Gene functional enrichment analysis in the interaction network

All genes were entered into the EASE software, and 3 significantly enriched functions for the *S100a8* and *S100a9* genes were obtained. These functions included defense and trauma

responses and inflammatory reaction. Among these 3 functions, the most significant function was associated with immune defense responses (see Table 2).

Table 2. List of gene enrichment function on network.

Term	Function	P value	FDR
GO:0006952	Defense response	8.47E-07	0.00135311
GO:0009611	Response to wounding	1.45E-06	0.00232245
GO:0006954	Inflammatory response	1.46E-06	0.00232438

The mean values and standard deviations were calculated from technical triplicates experiments, $P < 0.05$ and FDR (False Discovery Rate) <0.05 are considered significant.

Module function analysis of the interaction network

After visualizing the network by using Cytoscape, the whole network function module was classified and annotated by the plugins Mcode and Bingo on the basis of the hypergeometric distribution ($P < 0.05$), and models containing the *S100a8* and *S100a9* genes were obtained (see Figure 4). The located modules for *S100a8* and *S100a9* had 5 and 2 significant functions, respectively (see Table 3).

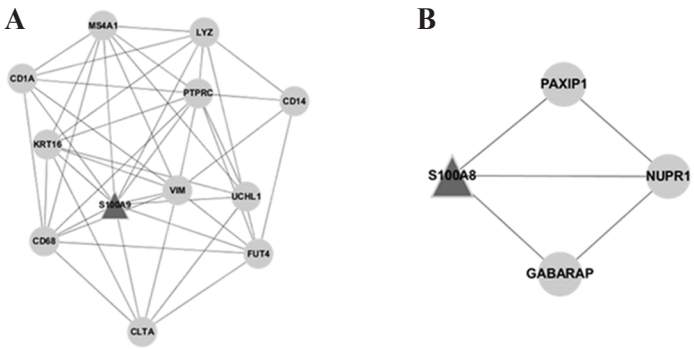

Figure 4. Gene *S100a8* and *S100a9* located module. **A.** Gene *S100a9* located module **B.** Gene *S100a9* located module. In this figure, triangle nodes represent the *S100a8* and *S100a9* genes, the other round nodes represent the interactive objects predicted by software.

Table 3. Function list of the *S100a8* (a) and *S100a9* (b) genes located module.

(a)			
GO-ID	P value	Corr. P value	Description
2376	3.33E-06	2.88E-04	Immune system process
9987	2.38E-06	2.88E-04	Cellular process
50896	9.30E-06	3.88E-04	Response to stimulus
51179	1.86E-03	1.05E-02	Localization
19222	1.94E-02	3.98E-02	Regulation of metabolic process
(b)			
GO-ID	P value	Corr. P value	Description
50896	2.49E-03	2.40E-02	Response to stimulus
42221	8.02E-03	3.58E-02	Response to chemical stimulus

S100a8 and *S100a9* located modules had 5 and 2 significant functions respectively. The mean values and standard deviations were calculated from technical triplicates experiments, $P < 0.05$ was regarded as significant difference.

DISCUSSION

Maintaining iron homeostasis is crucial for normal metabolic functions in humans. Loss of function of proteins involved in these regulatory mechanisms can cause HHC. HHC is a common genetic disorder involving dysregulation of iron absorption. HHC results in iron overload characterized by high transferrin saturation, low iron content in macrophages, and deposition of iron in several organs, including the liver, heart, and pancreas. Causative mutations for HHC have been described for several genes, namely, *HFE*, *TFR2* (encoding transferrin receptor 2), *HJV* (encoding hemojuvelin), and *HAMP* (encoding hepcidin) (Camaschella et al., 2000; Roetto et al., 2003). It has been proposed that these mutations cause deficient hepcidin synthesis (Papanikolaou et al., 2004; Nemeth et al., 2005). However, few studies have used effective high-throughput screening such as gene microarrays to screen for gene mutations.

HHC encompasses genetic disorders of iron overload characterized by deficient expression or function of the iron-regulatory hormone hepcidin, and the heart muscle is one of the target tissues affected by iron overload, resulting in cardiomyopathy in some HHC patients (Rodriguez et al., 2007). Therefore, to better understand the mechanisms behind the pathological effects of iron overload in skeletal muscle and heart muscle, we have performed a genome-wide gene expression analysis in skeletal and heart muscle of mice with or without iron loading.

Among the upregulated genes showing the highest fold changes in expression in several studies are 2 genes, *S100a8* (calgranulin A) and *S100a9* (calgranulin B), which encode calcium- and zinc-binding proteins. The S100A8 and S100A9 proteins have molecular masses of 10.8 and 13.2 kDa and are 93 and 114 amino acids long, respectively (Schafer and Heizmann, 1996; Salama et al., 2008). The S100A8 and S100A9 proteins became the focus of intensive research because of their association with numerous human disorders, including acute and chronic inflammatory conditions, autoimmune diseases, cancer, atherosclerosis, cardiomyopathies, and neurodegenerative diseases (Healy et al., 2006), and they also play crucial roles in normal physiological processes within cells. Some authors have reported that the transcripts of both *S100a8* and *S100a9* are strongly upregulated in skeletal muscle, heart, and liver of iron-loaded mice; however, the levels of the *S100a8* transcript in skeletal muscle are very low and, in general, both genes are weakly expressed. It is noteworthy that the *S100a8* and *S100a9* transcripts show a very similar pattern of upregulation in these tissues, which is consistent with the observation that the 2 proteins form a heterodimer (Rodriguez et al., 2007).

To date, increasing evidence suggests that both S100A8 and S100A9 are involved in many intra- and extracellular biological processes. This functional diversity lies in their ability to interact with various protein targets and to modulate the structures and functions of these targets. Many researchers have screened for the extracellular function of the S100A8/A9 protein complex and identified a number of interaction partners or receptors. For example, Kerkhoff et al. (2001) identified the FAT/CD36 receptor as being involved in facilitating the uptake of arachidonic acid bound to the S100A8/A9 protein. Furthermore, S100A8 has been shown to interact with TLR4, amplifying phagocyte activation during sepsis (Vogl et al., 2007). The receptor for advanced glycation end products (RAGE) has been proposed to serve as a cellular receptor of S100A8/A9, thereby mediating some of the activities described (Herold et al., 2007). Bode et al. (2008) have presented evidence that annexin A6 interacts with S100A8/A9 and that this interaction takes place in membranous structures. Moreover, the presence of annexin A6 in the cell membrane may be necessary but not sufficient for

the calcium-induced cell surface exposition of S100A8/A9. Recently, Li et al. (2003), have reported that oncostatin induces S100A9 via a STAT3-dependent pathway in MCF-7 cells and that at least in MCF-7 cells oncostatin M-mediated growth inhibition is ablated by the repression of S100A9 expression.

The gene set enrichment analysis essentially investigates whether the expression of a list of genes in a microarray data set is positively or negatively associated with one of the two experimental groups (e.g., in mice with and without iron overload). In this study, analysis of the microarray data sets identified several genes whose expression were altered in skeletal muscle and heart because of iron overload. By comparing gene expression in normal and iron-overload samples, we observed that both *S100a8* and *S100a9* were common DEGs in the heart and skeletal muscle. We obtained the feature genes *S100a8* and *S100a9*, which are related to HHC (iron overload), and the gene-set enrichment analysis indicated an enrichment of genes involved in iron overload.

Over the past 20 years, the protein complexes of S100A8 and S100A9 have emerged as very potent biomarkers of a wide range of inflammatory processes (Healy et al., 2006; Bresolin et al., 2012). Therefore, we speculate that S100A8 and S100A9 and the proteins with which they interact not only serve as useful markers of inflammation, but also play critical roles in the pathogenesis of inflammatory disorders.

According to the relationship between early stage of MDS and the immune response, we hypothesize that expression of the *S100a8* and *S100a9* genes represent diagnostic features of early-stage MDS and that the 2 genes could be used as targets for clinical treatment of MDS. Our results might have revealed novel links between iron overload and pathological manifestations of HHC. Further investigation of the *S100A8* and *S100A9* genes may lead to a better understanding of how iron overload contributes to these common HHC manifestations.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Clinic Scientific Research Foundation of the Third Military Medical University (#2007XG073).

REFERENCE

Arisi I, D'Onofrio M, Brandi R, Felsani A, et al. (2011). Gene expression biomarkers in the brain of a mouse model for Alzheimer's disease: mining of microarray data by logic classification and feature selection. *J. Alzheimers Dis* 24: 721-738.

Beck D, Ayers S, Wen J, Brandl MB, et al. (2011). Integrative analysis of next generation sequencing for small non-coding RNAs and transcriptional regulation in Myelodysplastic Syndromes. *BMC Med. Genomics* 4: 19.

Bejar R, Levine R and Ebert BL (2011). Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J. Clin. Oncol.* 29: 504-515.

Benza AT, Calvert S and McQuown CB (2010). Prevention BINGO: reducing medication and alcohol use risks for older adults. *Aging Ment. Health* 14: 1008-1014.

Bode G, Luken A, Kerkhoff C, Roth J, et al. (2008). Interaction between S100A8/A9 and annexin A6 is involved in the calcium-induced cell surface exposition of S100A8/A9. *J. Biol. Chem.* 283: 31776-31784.

Bresolin S, Trentin L, Zecca M, Giordan M, et al. (2012). Gene expression signatures of pediatric myelodysplastic syndromes are associated with risk of evolution into acute myeloid leukemia. *Leukemia* 26: 1717-1719.

Camaschella C, Roetto A, Cali A, De Gobbi M, et al. (2000). The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. *Nat. Genet.* 25: 14-15.

Chen XW, Jeong JC and Dermyer P (2011). KUPS: constructing datasets of interacting and non-interacting protein pairs with associated attributions. *Nucleic Acids Res.* 39: D750-D754.

Fujita A, Sato JR, Rodrigues Lde O, Ferreira CE, et al. (2006). Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7: 469.

Healy AM, Pickard MD, Pradhan AD, Wang Y, et al. (2006). Platelet expression profiling and clinical validation of myeloid-related protein-14 as a novel determinant of cardiovascular events. *Circulation* 113: 2278-2284.

Herold K, Moser B, Chen Y, Zeng S, et al. (2007). Receptor for advanced glycation end products (RAGE) in a dash to the rescue: inflammatory signals gone awry in the primal response to stress. *J. Leukoc. Biol.* 82: 204-212.

Hosack DA, Dennis G Jr, Sherman BT, Lane HC, et al. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4: R70.

Jadersten M and Hellstrom-Lindberg E (2009). Myelodysplastic syndromes: biology and treatment. *J. Intern. Med.* 265: 307-328.

Kerkhoff C, Sorg C, Tandon NN and Nacken W (2001). Interaction of S100A8/S100A9-arachidonic acid complexes with the scavenger receptor CD36 may facilitate fatty acid uptake by endothelial cells. *Biochemistry* 40: 241-248.

Li C, Shridhar K and Liu J (2003). Molecular characterization of oncostatin M-induced growth arrest of MCF-7 cells expressing a temperature-sensitive mutant of p53. *Breast Cancer Res. Treat.* 80: 23-37.

Majore S, Ricerca BM, Radio FC, Binni F, et al. (2013). Type 3 hereditary hemochromatosis in a patient from sub-Saharan Africa: is there a link between African iron overload and TFR2 dysfunction? *Blood Cells Mol. Dis.* 50: 31-32.

Nemeth E, Roetto A, Garozzo G, Ganz T, et al. (2005). Hepcidin is decreased in TFR2 hemochromatosis. *Blood* 105: 1803-1806.

Papanikolaou G, Samuels ME, Ludwig EH, MacDonald ML, et al. (2004). Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat. Genet.* 36: 77-82.

Pietrangelo A (2010). Hereditary hemochromatosis: pathogenesis, diagnosis, and treatment. *Gastroenterology* 139: 393-408.

Rodriguez A, Hilvo M, Kytomaki L, Fleming RE, et al. (2007). Effects of iron loading on muscle: genome-wide mRNA expression profiling in the mouse. *BMC Genomics* 8: 379.

Roetto A, Papanikolaou G, Politou M, Alberti F, et al. (2003). Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. *Nat. Genet.* 33: 21-22.

Salama I, Malone PS, Mihaimeed F and Jones JL (2008). A review of the S100 proteins in cancer. *Eur. J. Surg. Oncol.* 34: 357-364.

Schafer BW and Heizmann CW (1996). The S100 family of EF-hand calcium-binding proteins: functions and pathology. *Trends Biochem. Sci.* 21: 134-140.

Shannon P, Markiel A, Ozier O, Baliga NS, et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498-2504.

Smyth GK, Michaud J and Scott HS (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21: 2067-2075.

Sredoja Tišma V, Bulimbašić S, Jaganjac M, Stjepandić M, et al. (2012). Progressive pigmented purpuric dermatitis and alopecia areata as unusual skin manifestations in recognizing hereditary hemochromatosis. *Acta Dermatovenerol. Croat.* 20: 181-186.

Troyanskaya O, Cantor M, Sherlock G, Brown P, et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520-525.

Vogl T, Tenbrock K, Ludwig S, Leukert N, et al. (2007). Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nat. Med.* 13: 1042-1049.

Xiong M, Fang X and Zhao J (2001). Biomarker identification by feature wrappers. *Genome Res.* 11: 1878-1887.