A genetic features and gene interaction study for identifying the genes that cause hereditary spherocytosis

Jing Chen a, Yang Zhou b, Yaqi Gao c, Weijie Cao b, Hui Sun b, Yanfang Liu b and Chong Wang b

aNursing College of Zhengzhou University, Zhengzhou, China; bDepartment of Hematology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; cNursing College of Hebi Polytechnic, Hebi, China

ABSTRACT

Objective: Hereditary spherocytosis (HS) is a hemolytic disorder characterized by the presence of spherical-shaped red blood cells on the peripheral blood smear. Non-dominant HS cases are due to de novo mutations of the type associated with dominant inheritance or recessive genes. This study is aimed to identify HS-related biological mechanisms and predicting HS candidate genes.

Methods: We searched the known HS-related genes from the public databases. By analyzing the gene ontology (GO) and biological pathway of these genes, we extracted the optimal features to encode HS genes. Based on them, we predicted the HS-related genes from genes of whole genomes using the Random Forest classification. We used the gene interaction networks analysis to further identify the core regulatory genes that were related to HS.

Results: Forty-one known HS-related genes were found out and encoded. Three hundred and sixty-seven GO terms and ten biological pathway terms were identified as the optimal features for prediction. We subsequently predicted 150 novel HS-related genes and identified the core regulatory genes in the interaction network of predicted and known genes. These features and genes that we identified could complement the genetic features of HS.

KEYWORDS
Hereditary spherocytosis; gene ontology; gene interaction network; core regulatory genes

Introduction

Hereditary spherocytosis (HS) is a hemolytic disorder associated with a variety of mutations that characterized by the presence of spherical-shaped red blood cells on the peripheral blood smear [1,2]. HS usually is transmitted as an autosomal dominant trait [3], and it is relatively common in northern European populations with a high frequency of 1:5000 and most affected individuals have mild or only moderate hemolysis [4,5]. Although HS is encountered in all ethnic groups worldwide, its prevalence in other groups has not been established clearly. One systematic review demonstrated that the overall prevalence of HS in China is approximately 1.3 cases per 100,000 population [6]. In cases with a positive family history for HS and presence of spherocytes, it is sufficient to make the diagnosis by a confirmatory test. However, in cases with no family history or with atypical feature, molecular genetic analysis would help to determine whether inheritance is recessive or non-dominant [7].

About half of non-dominant HS cases are due to de novo mutations of the type associated with dominant inheritance; the others are assumed to be due to recessive genes. De novo mutations of ankyrin or other membrane protein genes is a frequent cause of non-dominant HS [8,9]. Some studies indicate that the mutations of the ankyrin-1 gene (ANK1) is one common cause in some patients of HS [10–12]. In a clinical study using human cells and transgenic mice, researchers discovered that upstream region of the ANK1 erythroid promoter acted as a barrier insulator and identify disruption of the barrier element as a potential pathogenetic mechanism of HS [13]. The clinical severity of HS varies considerably even within one family, reflecting the genetic heterogeneity of the disorder. Another study on a group of 190 patients suggested that HS was due to band three deficiency and all patients were heterozygous for the particular mutation in SLC4A1 with 15 novel mutations found [14]. Cases with co-existence of HS and UGT1A1 deficiency have been reported as well [15,16]. Mutations in HFE gene as the cause of the iron deposits, a synergistic effect of HS, was proposed in a report on a Spanish family in which three members of different generations were diagnosed with HS [17]. However, it still remains a challenge to identify reliable biomarkers of notable genetic changes for fully understanding genetic mechanism and effective diagnosis, in addition, novel therapeutic interventions such as gene targeting are urgently needed.

The gene ontology (GO) is community-based bioinformatics resource established with the aim to standardize description of the functions and classify gene
product function through the use of structured, controlled vocabularies in terms of their associated biological processes, cellular components (CCs), and molecular functions (MFs) [18]. GO consolidates specialized knowledge and expertise to ensure the data remain a key reference for up-to-date biological information. It has been demonstrated that GO annotations are effective predictors of disease causative genes. The Kyoto encyclopedia of genes and genomes (KEGG) is a large-scale data set that includes a collection of manually curated pathway maps for understanding the functions and utilities of cells and organisms from both high-level and genomic perspectives, such as genetic information processing, environmental information processing, cellular processes, human diseases, and drug development [19]. GO and KEGG are two main and most substantial functional annotations for investigators to understand biological meaning behind genes.

We could explore the mechanisms of diseases through reliable system biology approaches. Here we presented a novel systems biological measure to identify HS-related biological mechanisms and to predict HS candidate genes by analyzing the GO and KEGG materials in order to extract optimal features to encode HS genes. Subsequent analyses suggested that certain such genes were related directly or indirectly to HS formation or development. These predicted genes are promising module biomarkers which would play a vital role in prevention, prognostics, early diagnosis and effective treatment of HS and assist in the research progress of new molecular drug targets to treat this disease.

Results

The archived HS-related genes in the public databases

We used the ‘hereditary spherocytosis’ as the keyword to search the HS-related genes in the public databases including OMIM (Online Mendelian Inheritance in Man) [20], Gad (The Genetic Association Database) [21], and DisGeNET [22]. We finally found 6, 5, and 40 features in the prediction. We finally found 6, 5, and 40 genes in the whole genomes as the positive samples formed the six groups and the positive samples formed the six groups. The negatives had a large number and they were randomly divided into six groups. The six groups and the positive samples formed the six training datasets for selecting respective optimal features in the prediction.

The genetic features of these genes were represented by GO and KEGG enrichment scores. We constructed the feature vector of each genes with consistent length containing the 6242 GO terms and 214 KEGG terms. If the group of genes have no relationship with the given GO or KEGG, the value is zero.

The optimal features for predicting the HS-related genes

As there were lots of features in the training datasets, we firstly screened GO or KEGG terms to improve the efficiency and accuracy of the prediction for HS-related genes. We firstly removed the uncorrelated features and ranked the importance of them by minimum redundancy maximum relevance (mRMR) algorithm. We employed the incremental feature selection (IFS) to traverse each combination of the features and used the ten-fold cross validation to calculate the performance of the prediction model for the given combination of features in each training dataset. The six training datasets contained different negative samples and thus the optimal features for them were different (Table 1).

We defined the eventual optimal features if one feature appeared in at least four groups of optimal features of the six datasets. The final optimal features contained 367 GO terms and 10 KEGG terms. The top GO and KEGG terms are listed in Table 2.

The predicted HS-related genes

We employed the optimal features found above to encode all the genes in the whole genome and employed the same prediction model to classify them. In the classification, there were 188 genes that were predicted as HS-related genes. Among them, 150 genes were not archived in the databases we had searched.

We applied the functional annotation clustering tool provided by the database for annotation, visualization, and integrated discovery (DAVID) to the prediction list and the result demonstrated that many genes predicted were most likely associated with HS. GO term ‘cellular response to stress’, ‘response to oxidative stress’ and ‘oxidation reduction’ were enriched with adjusted P value 5.7E-18, 6.2E-6, and 5.4E-5

| Dataset | Total features in testing | Final optimal feature | Sn | Sp | Acc | Mcc |
|---------|--------------------------|-----------------------|----|----|-----|-----|
| 1       | 144                      | 0.82                  | 0.94 | 0.90 | 0.77 |
| 2       | 24                      | 0.74                  | 0.96 | 0.89 | 0.75 |
| 3       | 190                     | 0.82                  | 0.94 | 0.90 | 0.77 |
| 4       | 87                      | 0.84                  | 0.93 | 0.90 | 0.77 |
| 5       | 59                      | 0.87                  | 0.93 | 0.91 | 0.79 |
| 6       | 80                      | 0.84                  | 0.92 | 0.90 | 0.76 |

Note: Sn, sensitivity; Sp, specificity; Acc, accuracy; Mcc, Matthews’s correlation coefficient.
respectively. Thirty-eight genes in prediction list were annotated with ‘cellular response to stress’, 13 genes were annotated with ‘response to oxidative stress’, and 22 genes were annotated with ‘oxidation reduction’. Oxidative stress is believed to aggravate the symptoms of several anemias, including HS. Oxidative damage to erythrocyt cells plays a crucial role in hemolysis due to ineffective erythropoiesis in the bone marrow and short survival of red blood cells in the circulation [23]. Seven significantly-enriched GO terms were clustered in regulation of apoptosis and cell death, namely ‘regulation of apoptosis’, ‘regulation of programmed cell death’, and ‘induction of apoptosis by intracellular signals’, etc. Approximately 35 genes in prediction list were associated with programmed cell death. Eryptosis, the programmed cell death of erythrocytes, is characterized by erythrocyte shrinkage, blebbing, and phospholipid scrambling of the cell membrane [24]. The mechanism studies of faster erythrocyte clearance in HS showed that eryptosis may be a useful mechanism to remove defective erythrocytes prior to hemolysis [25,26].

**The core regulatory genes in known and predicted HS-related genes**

Given that the HS is caused by defects of groups of genes and the hubs in gene interaction networks might be the key regulators, we used the gene interaction information in the STRING database to construct the gene interaction network. We then applied the ‘k-core’ scores to identify those with highest networking degrees as the ‘key regulatory’ genes that may play pivotal roles in gene interactions and regulations. Thus 70 genes were chosen as the ‘key regulatory’ genes (Figure 1 and Supplementary Table 2).

The k-core scores also could depict the structure of the gene interaction for HS (Figure 1). The known HS-related genes were prone to be on the periphery. The predicted HS-related genes have tight connections with the known genes. It seemed that the core regulatory genes in HS had been predicted and we complemented the relationship of the genes related to the HS. We think these genes can be good candidate biomarkers for detection and targets for future therapies in future therapies.

**Discussion**

In this study, we predicted the HS-related genes by learning the features of known HS genes and selecting the optimal features for prediction. Because these optimal features performed best in the prediction for known HS genes, they could represent the genetic characteristics of HS. Analyzing these features will shed light on the pathogenesis of HS.

In accordance with the structure of the GO, we first classified GO terms in the optimal list into three types: biological process (BP), CC, and MF so as to illustrate the biological implications of the selected optimal features in HS pathogenesis (Figure 2).

**Biological process GO terms**

In the percentage of BP terms, the top five GO biological processes are GO:0023052: signaling (2.03%), GO:0051234: establishment of localization (2.03%), GO:0065007: biological regulation (1.75%), GO:0071840: CC organization or biogenesis (1.66%) and GO:0040007: growth (1.52%). These features associate closely with HS pathogenesis. For instance, a case report pointed out that HS erythroid progenitors with a Jak2(V617F) mutation, which constitutively activates downstream signaling independent of erythropoietin and its receptor could acquire a selective proliferation advantage over endogenous HS cells [27]. The peripheral membrane protein, protein 4.2, is one of the most abundant protein components of the erythrocyte membrane. One of cytoplasmic HS mutants, protein 4.2 fatty acylation mutant, G2A/C173A, had decreased plasma-membrane localization compared with wild-type protein 4.2, leaving protein 4.2 susceptible to loss during erythrocyte development [28]. It is revealed that regulation of erythrocyte volume homeostasis is critical for survival of the erythrocyte [29]. Inherited or acquired disorders that perturb this homeostasis jeopardize the erythrocyte, leading to diseases such as sickle cell disease, thalassemia, and HS.

**CC GO terms**

The top GO CC terms considering percentage are GO:0043226: organelle (3.48%), GO:0044422: organelle

---

**Table 2. The top features of the final optimal features.**

| Order | Features type | Name |
|-------|---------------|------|
| 1     | GO            | Regulation of protein stability |
| 2     | GO            | Positive regulation of mitochondrial membrane permeability |
| 3     | GO            | Organ regeneration |
| 4     | GO            | Positive regulation of sodium ion transmembrane transporter activity |
| 5     | GO            | Transferase activity, transferring acyl groups, acyl groups converted into alkyl on transfer |
| 6     | GO            | Benzodiazepine receptor activity |
| 7     | GO            | Ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity |
| 8     | GO            | Mitochondrial genome maintenance |
| 9     | GO            | Tropomysin binding |
| 10    | GO            | Viral genome replication |
| 11    | GO            | Hemoglobin binding |
| 12    | GO            | Response to toxic substance |
| 13    | GO            | Adiponectin binding |
| 14    | GO            | Intramolecular transferase activity |
| 15    | GO            | Gial cell projection |
| 27    | KEGG          | One carbon pool by folate |
| 42    | KEGG          | Porphyrin and chlorophyl metabolism |
| 68    | KEGG          | Lysine biosynthesis |
| 124   | KEGG          | O-Glycan biosynthesis |
| 189   | KEGG          | Pyrimidine metabolism |
part (3.14%), GO:0044464: cell part (2.78%),
GO:0005623: cell (2.61%), GO:0016020: membrane
(2.29%), GO:0044425: membrane part (2.28%),
GO:0030054: cell junction (2.17%). HS is usually
caused by mutations in genes relating to membrane
proteins that allow for the red blood cells to change
shape which involves multiple biological reactions in
different regions in cell. In HS, intrinsic defects in eryth-
rocyte membrane proteins result in RBC cytoskeleton
instability. The integrating protein ankyrin, most com-
monly defective in HS patients, is responsible for incor-
poration and binding of spectrin, thus in its dysfunction
cytoskeletal instabilities ensue [30]. The primary defect
in HS is a deficiency of membrane surface area. In red
blood cells, adducin heterotetramers localize to the
spectrin-actin junction of the peripheral membrane
skeleton. Deletion of beta-adducin in mice resulted in
osmotically fragile, microcytic erythrocytes, and a phe-
notype of HS [31].

**MF GO terms**
The top five GO MF terms of percentage are
GO:0016247: channel regulator activity (18.18%),
GO:0045182: translation regulator activity (16.67%),
GO:0060089: molecular transducer activity (3.49%),
GO:0005488: binding (3.17%), and GO:0016209: antiox-
didant activity (3.13%). It was documented how potas-
sium channel tetramerization domain containing 6
(KCTD6) represents a novel substrate adaptor for
cullin-3, effectively regulating protein levels of the
muscle small ANK1 isoform 5. The interaction of
KCTD6 with ANK1 may have implications beyond
muscle for HS [32]. Moreover, a translation initiation

---

**Figure 1.** The gene interaction network of the predicted and known HS genes. The deeper color indicates the higher k-core score. The known HS genes in the databases are marked by blue circles.
Codon mutation of the beta-spectrin gene (ATG → GTG) was tested to associate with HS and spectrin deficiency in a family in Brazil [33]. Band 3 is the most abundant protein in the erythrocyte membrane and forms the core of a major multiprotein complex with cytoskeletal binding domain required for multiprotein complex formation and retention during erythropoiesis. Specific mutations in the band 3 gene result in the erythrocyte diseases HS [34]. A group in Israel indicated that oxidative stress plays an important role in the pathophysiology of HS which can be ameliorated by an antioxidant such as fermented papaya preparation with known antioxidative properties [35].

**The KEGG pathways in the optimal set**

Among the KEGG pathway terms in the optimal set of features, two showed certain connection with HS. Lysine biosynthesis (hsa00300) is one feature and in previous study on implication in KCDT6’s interaction with ANK1 of HS, it showed binding of sAnk1.5 to KCTD6, and its subsequent turnover is regulated through posttranslational modification by nedd8, ubiquitin, and acetylation of C-terminal lysine residues [32]. The other feature is one carbon pool by folate (hsa00670). Dietary supplementation of folic acid is one of interventions that limit the severity of the disease [36]. EPB42-related HS is a chronic non-immune hemolytic anemia that is usually of mild to moderate severity. Treatment for mild EPB42-HS includes folate supplementation as needed for a hemolytic or aplastic crisis [37].

**Further selection of predicted HS-related genes using gene interaction network**

We constructed gene interaction network and selected genes with high connectivity as core genes. These predicted genes are considered to play important roles in HS pathology. We utilized DAVID for functional enrichment and clustering analysis of the selected candidates as well. Unsurprisingly, annotation cluster related to regulation of programmed cell death was assigned enrichment score of 8.02 and cluster associated with...
response to oxidative stress had enrichment score of 4.72. The results matched with functional annotation analysis we performed using preliminary prediction genes above and it demonstrated the reliability of our methods. Specifically, ENSG00000228849 (TNF) was annotated several functions in apoptosis regulation process. It was reported in a study on altered red cell and platelet adhesion in hemolytic diseases. RBC adhesion and TNF were augmented in HS [38]. It consisted with another HS patient clinical outcome study that concluded that HS patients presented with higher levels of neutrophils, TNF, IFN-γ, elastase, lactoferrin, and ferrit [36]. ENSG00000100292 (HMOX1) is the rate limiting enzyme in heme degradation, which results in the production of biliverdin, carbon monoxide, and free iron. A research group recently suggested that promoter polymorphisms in the UGT1A1 and HMOX1 genes could act synergistically to increase the risk of hyperbilirubinemia and gallstones in patients with HS [39]. Although there are non-dominant case reports of HS coexisting with Gilbert syndrome [40,41], authors implied systemic studies should have been done to look for the effect of promoter polymorphisms in the HMOX1 in patients with HS.

Methods

Features construction in prediction

In order to develop a powerful predictor for predicting related genes, the primary key is to formulate known related genes with an effective mathematical expression that can truly reflect its intrinsic correlation with the attribute to be predicted. In this study, the GO and KEGG enrichment score were utilized to formulate the features of genes. The GO [18] descripts the gene functions and gene products, which will facilitate computer-driven information retrieval and generation. The KEGG [42] also annotates the genetic features including the knowledge of molecular interaction and reaction networks of metabolism, genetic information processing, environmental information processing, cellular processes, human diseases, and drug development. If two genes shared a similar function, they will have similar annotations in the GO and KEGG pathway. Therefore, each gene was encoded as a numeric vector consisting of an enrichment score for each of the GO and KEGG terms. The GO or KEGG enrichment score of a gene is defined as the −log10 of the hyper-geometric test P value [43,44] of its direct neighbors in gene interaction network.

Feature reduction

The Cramer’s V coefficient between features and sample variables was utilized to measure association between the large number of features and sample variables. The features with Cramer’s V coefficient smaller than 0.1 were removed just to release the burden of calculation in selecting the optimal features.

mRMR method

mRMR algorithm simultaneously guarantees max-relevance and min-redundancy. It guarantees the features that will be selected is the ones giving most contribution to the classification and ones excluding the features whose classification ability has already been covered by selected features. The mRMR program developed by Peng et al. [45] is available at http://penglab.janelia.org/proj/mRMR/. The mRMR was utilized to rank the importance of the features. The rank of features benefits the following examination of the prediction performance.

Prediction model

In the prediction, the random forest (RF) algorithm was adopted as the classifier. The RF method comprises many decision trees and it determined a final classification label according to the class with the most votes from all trees. The RF algorithm had been successfully applied to several studies [46,47]. In this study, the software Weka [48] provide the RF algorithm and the classification or prediction was performed with default parameters.

Incremental feature selection and ten-fold cross validation

Based on the features’ rank derived from the mRMR Algorithm, features were added one by one to form the combination of the feature sets and to determine the optimal number of features according to the examination results. This approach is called IFS. For each of the feature combination, the RF classifier was performed and ten-fold cross-validation test examine the performance of classification. In the ten-fold cross-validation test, the dataset was randomly divided into 10 equal parts, of which each of the 10 parts was used in turn as test data, the other 9 as training data in RF classification. This process was repeated for 10 times and finally the prediction accuracy was calculated. The Matthews’s correlation coefficient (MCC) based on the true positive rate, true negative rate, the false positive rate, and false negative rate was employed to evaluate the performance of the classification. When the maximum of MCC value firstly appeared, the combination of the features were regarded as the optimal features set.

Gene interaction networks analysis

The known and predicted related genes were selected to build genes interaction network according to the
relationship between the genes in the STRING database. Within the network analysis, degree centrality is the simplest and most important measures of the centrality of a gene within a network that determines the relative importance. Degree centrality is defined as:

\[ \text{Degree Centrality} = \text{Number of edges} \]

In the network analysis, k-core scoring was correlated with the link numbers one node has to the other. In the relative importance, Degree centrality is defined as the simplest and most important measures of the centrality of a gene within a network that determines the degree centrality of a gene indicates its hub or nodal status with connection to 'k' other genes in a sub-network. Accordingly, the genes with largest k-core scores were identified as 'key regulatory genes' in the network.

Disclosure statement
No potential conflict of interest was reported by the authors.

References

[1] Bolton-Maggs PH, Langer JC, Iolascon A, et al. Haematology general haematology task force of the British committee for standards in, guidelines for the diagnosis and management of hereditary spherocytosis – 2011 update. Br J Haematol. 2012;156(1):37–49. PMID:22055020.

[2] Guitton C, Garcon L, Cynober T, et al. Hereditary spherocytosis: guidelines for the diagnosis and management in children. Arch Pediatr. 2009;16(6):556–558. PMID:19541083.

[3] Miraglia del Giudice E, Nobili B, Francese M, et al. Clinical and molecular evaluation of non-dominant hereditary spherocytosis. Br J Haematol. 2001;112(1):42–47. PMID:11167781.

[4] Wang C, Cui Y, Li Y, et al. A systematic review of hereditary spherocytosis reported in Chinese biomedical journals. Ann Hematol. 2009;88(1):53–58. PMID:19889093.

[5] Perrotta S, Gallagher PG, Mohandas N. Hereditary spherocytosis. Lancet. 2008;372(9647):1411–1426. PMID:18940465.

[6] Wang C, Cui Y, Li Y, et al. A systematic review of hereditary spherocytosis reported in Chinese biomedical journals from 1978 to 2013 and estimation of the prevalence of the disease using a disease model. Intractable Rare Dis Res. 2015;4(2):76–81. PMID:25984425.

[7] Bianchi P, Ferro E, Vercellati C, et al. Diagnostic power of laboratory tests for hereditary spherocytosis: a comparison study in 150 patients grouped according to molecular and clinical characteristics. Haematologica. 2012;97(4):516–523. PMID:22052813.

[8] Maciag M, Plochocka D, Adamowicz-Salach A, et al. Novel beta-spectrin mutations in hereditary spherocytosis associated with decreased levels of mRNA. Br J Haematol. 2009;146(3):326–332. PMID:19538529.

[9] Nakanishi H, Kanzaki A, Yawata A, et al. Ankyrin gene mutations in Japanese patients with hereditary spherocytosis. Int J Hematol. 2001;73(1):54–63. PMID:11372755.

[10] Gustafson S, Eber S, Heep A. A new ankyrin mutation (ANK1 E9X) causing severe hereditary spherocytosis in the neonatal period. Ann Hematol. 2011;90(2):231–232. PMID:20512576.

[11] Rank G, Sutton R, Marshall V, et al. Novel roles for erythroid ankyrin-1 revealed through an ENU-induced null mouse mutant. Blood. 2009;113(14):3352–3362. PMID:19179303.

[12] Eber SW, Gonzalez JM, Lux ML, et al. Ankyrin-1 mutations are a major cause of dominant and recessive hereditary spherocytosis. Nat Genet. 1996;13(2):214–218. PMID:8640229.

[13] Gallaghér PG, Steiner LA, Liem RL, et al. Mutation of a barrier insulator in the human ankyrin-1 gene is associated with hereditary spherocytosis. J Clin Invest. 2010;120(12):4453–4465. PMID:21099109.

[14] Van Zwieten R, Francois JJ, Van Leeuwen K, et al. Hereditary spherocytosis due to band 3 deficiency: 15 novel mutations in SLC4A1. Am J Hematol. 2013;88(2):159–160. PMID:23255290.

[15] Berardi A, Lugli L, Ferrari F, et al. Kernicterus associated with hereditary spherocytosis and UGT1A1 promoter polymorphism. Biol Neonate. 2006;90(4):243–246. PMID:16735790.

[16] Rocha S, Costa E, Ferreira F, et al. Hereditary spherocytosis and the (TA)n(TA) polymorphism of UGT1A1 gene promoter region – a comparison of the bilirubin plasmatic levels in the different clinical forms. Blood Cells Mol Dis. 2010;44(2):117–119. PMID:19931474.

[17] Montes-CanO MA, Rodriguez-Munoz F, Franco-Osorio R, Nunez-Roldan A, Gonzalez-Escribano MF. Hereditary spherocytosis associated with mutations in HFE gene. Ann Hematol. 2003;82(12):769–772. PMID:12961032.

[18] Shaw DR, Ashbumer M, Blake JA, et al. Gene Ontology: a controlled vocabulary to describe the function, biological process and cellular location of gene products in genome databases. Am J Hum Genet. 1999;65(4):A414. PMID:WOS:000082879802373.

[19] Kancheisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30. PMID:10592173.

[20] Rashbass J. Online mendelian inheritance in man. Trends Genet. 1995;11(7):291–292. PMID:10592173.

[21] Becker KG, Barnes KC, Bright TJ, et al. The genetic association database. Nat Genet. 2004;36(5):431–432. PMID:WOS:000211830000002.

[22] Bauer-Mehren A, Rautschka M, Sanz F, et al. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics. 2010;26(22):2924–2926. PMID:WOS:0002839198000027.

[23] Fibach E, Rachmilowitz E. The role of oxidative stress in hemolytic anemia. Curr Mol Med. 2008;8(7):609–619. PMID:18991647.

[24] Basu S, Banerjee D, Chandra S, et al. Erythropoiesis in hereditary spherocytosis and thalassemia: role of glyco-conjugates. Glycoconj J. 2010;27(9–10):717–722. PMID:19757027.

[25] Lang E, Qadri SM, Lang F. Killing me softly – suicidal erythocyte death. Int J Biochem Cell Biol. 2012;44(8):1236–1243. PMID:22561748.

[26] Lang F, Qadri SM. Mechanisms and significance of eryptosis, the suicidal death of erythrocytes. Blood Purif. 2012;33(1–3):125–130. PMID:22269222.

[27] Fleischmann RA. Polychromatemia vera and the Jak2(V617F) mutation in a case of hereditary spherocytosis. Am J Med Sci. 2013;346(4):328–330. PMID:23588264.

[28] Bustos SP, Reithmeier RA. Protein 4.2 interaction with hereditary spherocytosis mutants of the cytoplasmic domain of human anion exchanger 1. Biochem J. 2011;433(2):313–322. PMID:21039340.
[29] Gallagher PG. Disorders of red cell volume regulation. Curr Opin Hematol. 2013;20(3):201–207. PMID:23519154.
[30] Salomao M, Chen K, Villalobos J, et al. Hereditary spherocytosis and hereditary elliptocytosis: aberrant protein sorting during erythroblast enucleation. Blood. 2010;116(2):267–269. PMID:20339087.
[31] Sahre K, Lambert AJ, Cicciotto SL, et al. Targeted deletion of the gamma-adducin gene (Add3) in mice reveals differences in alpha-adducin interactions in erythroid and nonerythroid cells. Am J Hematol. 2009;84(6):354–361. PMID:19425068.
[32] Lange S, Perera S, Teh P, et al. Obscurin and KCTD6 regulate cullin-dependent small ankyrin-1 (sAnk1.5) protein turnover. Mol Biol Cell. 2012;23(13):2490–2504. PMID:22573887.
[33] Basseres DS, Vicentim DL, Costa FF, et al. Beta-spectrin Promiss-ao: a translation initiation codon mutation of the beta-spectrin gene (ATG → GTG) associated with hereditary spherocytosis and spectrin deficiency in a Brazilian family. Blood. 1998;91(1):368–369. PMID:9414314.
[34] Satchwell TJ, Hawley BR, Bell AJ, et al. The cytoskeletal binding domain of band 3 is required for multiprotein complex formation and retention during erythropoiesis. Haematologica. 2015;100(1):133–142. PMID:25344524.
[35] Ghoti H, Fibach E, Dana M, et al. Oxidative stress contributes to hemolysis in patients with hereditary spherocytosis and can be ameliorated by fermented papaya preparation. Ann Hematol. 2011;90(5):509–513. PMID:21063708.
[36] Rocha S, Costa E, Rocha-Pereira P, et al. Erhthropoiesis versus inflammation in hereditary spherocytosis clinical outcome. Clin Biochem. 2011;44(13):1317–1343. PMID:21704613.
[37] Kalfa TA, Connor JA, Begtrup AH. EPB42-related hereditary spherocytosis. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. Source GeneReviews® [Internet]. Seattle (WA): University of Washington; 2014. p. 1993–2016.
[38] Sakamoto TM, Canalli AA, Traina F, et al. Altered red cell and platelet adhesion in hemolytic diseases: hereditary spherocytosis, paroxysmal nocturnal hemoglobinuria and sickle cell disease. Clin Biochem. 2013;46(18):1798–1803. PMID:24060729.
[39] Warang P, Devendra R, D’Silva S, et al. Do UGT1A1 and HMOX1 gene promoter polymorphisms increase the risk of hyperbilirubinemia and gallstones in patients with hereditary spherocytosis? Ann Hematol. 2015;94(1):169–171. PMID:24947795.
[40] Garg PK, Kumar A, Teckchandani N, et al. Hereditary spherocytosis coexisting with Gilbert’s syndrome: a diagnostic dilemma. Singapore Med J. 2008;49(11):e308–e309. PMID:19037536.
[41] Kumar D, Parakh A, Sharma S. Gilbert syndrome increasing unconjugated hyperbilirubinemia in a child with hereditary spherocytosis. J Pediatr Hematol Oncol. 2012;34(1):54–56. PMID:22134611.
[42] Kancheisa M, Goto S, Sato Y, et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue): D199–D205. PMID:24214961.
[43] Carmona-Saez P, Chagoyen M, Tirado F, et al. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol. 2007;8(1):R3. PMID:17204154.
[44] Yang J, Chen L, Kong X, et al. Analysis of tumor suppressor genes based on gene ontology and the KEGG pathway. PLoS One. 2014;9(9):e107202. PMID:25207935.
[45] Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005;27(8):1226–1238. PMID:16119262.
[46] Mao R, Raj Kumar PK, Guo C, et al. Comparative analyses between retained introns and constitutively spliced introns in Arabidopsis thaliana using random forest and support vector machine. PLoS One. 2014;9(8):e104049. PMID:25110928.
[47] Diouf M, Filleron T, Pointet AL, et al. Prognostic value of health-related quality of life in patients with metastatic pancreatic adenocarcinoma: a random forest methodology. Qual Life Res. 2016;25(7):1713–1723. PMID:26615615.
[48] Bouckaert Remco R, Frank Eibe, Hall Mark A, et al. WEKA – experiences with a Java open-source project. J Mach Learn Res. 2010;10:2533–2541.
[49] Barabasi AL, Oltvai ZN. Network biology: understanding the cell’s functional organization. Nat Rev Genet. 2004;5(2):101–113. PMID:14735121.
[50] Ravasz E, Somera AL, Mongru DA, et al. Hierarchical organization of modularity in metabolic networks. Science. 2002;297(5586):1551–1555. PMID:12202830.