# A Family Based Whole Exome Sequence Study to Indentify Modifier Genes for Phenotype Heterogeneity Between Severe and Non-Severe Thalassemia Patients

ANUPAM BASU *, Dipankar Saha, Prosanto Kumar Chowdhury, , Debashis Pal, Kaustav Nayek, Gispati Chakrabortye, Surupa Basu

Posted Date: 3 January 2023

doi: 10.20944/preprints202103.0088.v2

Keywords: WES; Thalassemia; phenotype

*Article*

# A Family Based Whole Exome Sequence Study to Identify Modifier Genes for Phenotype Heterogeneity Between Severe and Non-Severe Thalassemia Patients

**Dipankar Saha [1#], Debashis Pal [1#], Prosanto Kumar Chowdhury [1,2] Kaustav Nayek [3], Gispati Chakraborty [1], Surupa Basu [2] and Anupam Basu [1,4] ***

[1] Department of zoology, The University of Burdwan, Purbo-Barddhaman, India,
[2] Institute of Child Health, Kolkata, India ,
[3] Department of Pediatric Medicine, Burdwan Medical College and Hospital, PurboBarddhaman, India.
[1] Burdwan University Health Centre, The University of Burdwan, PurboBarddhaman, India,
[4] National Institute of Biomedical Genomics, Kalayni, India
* Corresponding author: Email: ab3@nibmg.ac.in, abasu@zoo.buruniv.ac.in.
# Equal Contribution

**Abstract:** Thalassemia is a common monogenic autosomal disorder prevalent in India. HbE beta thalassemia is a compound heterozygous state of two different beta globin mutations (HBB), predominant in the Eastern India. In HbE-beta thalassemia ($\beta^+/\beta^\circ$) patients, one HBB mutation does not produce any functional protein ($\beta^\circ$); another mutation produces a structural altered haemoglobin, variant E ($\beta^+$). It has been observed that in HbE-beta thalassemia patients with same type of HBB mutations, there exist diverse phenotypic presentation ranging from very severe, with regular transfusion-dependence to clinically non severe requiring very less transfusion or no transfusion. To explore this phenotypic mystery, we performed trio based expanded whole exome sequencing (WES) of two extreme phenotypes of HbE beta thalassemia subjects with similar HBB genotype and their parents. For WES library preparation, Agilent Sure select V6+UTR kit was used and sequencing was done through the Illumina HiSeq 2500 platform. This study aimed to search for de novo or inherited variants associated with the clinical severity. Six significant inherited variants were found in the genes, ATOH8, TTLL4, P2RX7, PRSS54, SLC9C2 and VRK2 in severe subject, however only one significant inherited variant in MTRR gene was found in non-severe subject. Bioinformatic analysis also confirms that each variant significantly changes in respective protein structure. The identified genes were either associated with iron absorption, oxidative stress, hematopoietic stem cell differentiation, structural organization of RBC cytoskeleton or regulating anemia. No significant de novo variant were identified associated with clinical severity. The finding of this study indicates that the combined effect of mutation load in major pathways may responsible for severe phenotype.

**Keywords:** WES; Thalassemia; phenotype

## Introduction

Thalassemia is an inherited monogenic disease caused by the mutation in the beta globin (*HBB*) gene[1]. Although it's monogenic origin, clinical severity of beta thalassemia (β-thalassemia) varies from very severe to mild. It has been noticed that a group of subjects present the disease at very early age and need regular transfusion for survival, on the other hand, cases also reported with late onset and do not require regular transfusion[2]. Recently, we reported the role of HBB gene mutation and alpha globin gene (HBA) mutation has a role in the β-thalassemia clinical severity which ultimately regulate the globin synthesis[3,4]. Other studies also reveal that genetic variants which regulate the fetal hemoglobin (HbF) production may also causes the β-thalassemia clinical severity.

The present knowledge of genotyped for thalassemia severity, either on globin chain or globin synthesis centric. However, a huge percentage of β-thalassemia subject, in spite of having

 doi:10.20944/preprints202103.0088.v2



same globin genotype, but possess extreme different phenotype severity. Thus, it may be involvement of non-globin genes, which might play crucial role for such phenotype heterogeneity. To decipher this phenotypic mystery, we recruited two index cases with same HBB mutant of β+/β0 genotype [compound heterozygous IVS1-5(G>C) and CD26(G>A)], with *HBA* wild type.

## Subjects and Methods

Along with the index cases, their parents also recruited for trio-based study. Although they posses same mutant HBB genotype, but clinically they were of extreme variations (Table 1). This study was approved by the ethics committee of The University of Burdwan.

**Table 1.** Clinical findings of the selected subjects.

| Clinical parameter | Severe | Non-severe |
|---|---|---|
| Present age | 13 year | 16 year |
| Sex | Female | Female |
| Age of presentation | 10 months | 14 years |
| Baseline hemoglobin | 6.1 gm/dl | 8 gm/dl |
| Transfusion Interval | 2 months | No transfusion yet |
| Spleen size | 7 cm | 6 cm |
| MCV (fl) | 61.1 | 64.4 |
| MCH (pg) | 18.1 | 19.3 |
| RDW% | 28.5 | 29.8 |
| HPLC impression | HbE-β thalassemia | HbE-β thalassemia |
| HbA0 (%) | 3.1 | 6.6 |
| HbF (%) | 50.3 | 31.5 |
| HbA2+E (%) | 42.6 | 57.4 |

Subject 1, was presented the disease at 10 months of age and found to have severe anemia (steady state hemoglobin level was 6.1 g/dl). She undergone a regular blood transfusion at an interval of 2 months for the last 13 years (Table: 1). The sanger sequencing of *HBB* gene confirms the β thalassemia disease with [IVS1-5(G>C) and CD26(G>A)]. Both the parents are found to be carrier for β thalassemia (Figure 1A). She representing as severe, transfusion dependent thalassaemic subject

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202103.0088.v2



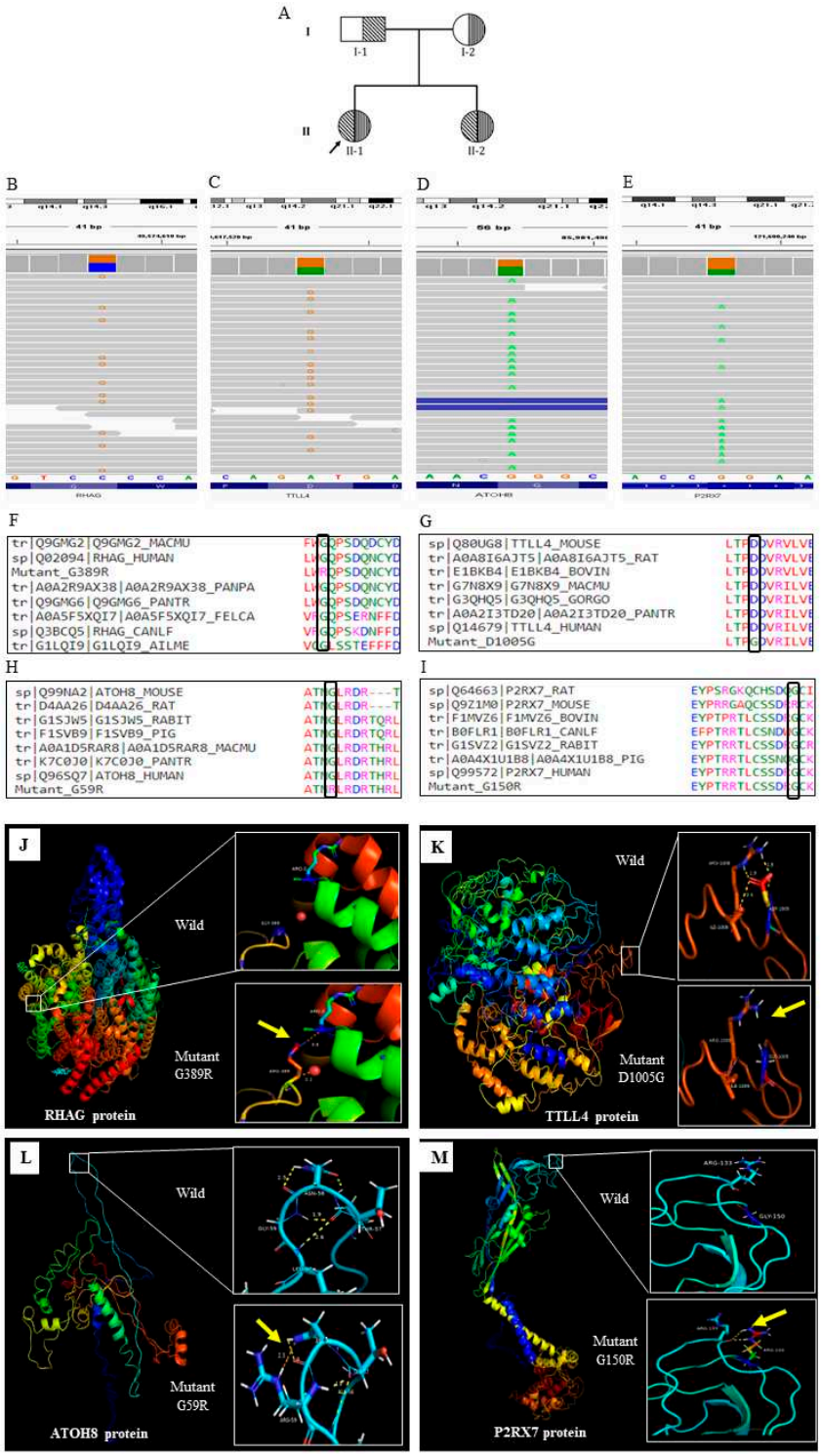


**Figure 1.** (**A)** Pedigree of family- 1, showing both the parents are carriers (I-1, I-2) and both the children are affected. II-1 is index case, the severe subject (marked by arrow) **(B, C, D, E)** IGV view of sequence BAM file, shows that all the selected variants are in heterozygous condition with high sequence coverage. **B;** *RHAG* (rs553258732) Reference base 'C' altered to G. **C;** *TTLL4* (rs116116034); **D;** *ATOH8* (rs558361422); **E;** *P2RX7* (rs28360447). **(F, G, H, I)** Conservation analysis of mutated amino acid by Multiple sequence alignment by Clustal Omega software, mutated amino acid sequence compare with human and other 6 mammalian ortholog species, conserved amino acid position marked in box, F. RHAG G. TTLL4, H. ATOH8 I. P2RX7 **(J, K, L, M)** Protein tertiary structure template PDB were either retrieved from RSCB PDB database or predicted by I-TASSER software and overall impact of missense mutations, were predicted by Comparing wild and mutated proteins using

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202103.0088.v2



PyMOL software. **J;** 3D structure of RHAG protein based on Protein Data Bank (PDB) template 7V0S. Substitution of p389 Gly>Arg in extended strand region introducing two new unfavorable interactions with the neighboring water molecule and with 2nd Arg amino acid. The water molecule is indicated by the red sphere and the unfavorable interaction indicated as orange dotted line. **K;** Predicted 3D structure of TTLL4 protein by I-TASSE, in wild type, 1005th Asp amino acid forms 2 hydrogen bonds with 1008th Arg, and is involved in 1 hydrogen bond with the 1009th Ile, but in mutant p1005 Asp> Gly all the 3-hydrogen bonds (yellow dotted line) are abolished, indicated with yellow arrow, which might leads to the disruption of the tight packaging of TTLL4 protein. **L;** Predicted 3D structure of ATOH8 protein by I-TASSER, mutation from Gly to Arg at 59th position involving a steric clash (orange dotted line) with Asn at 58th position indicated with yellow arrow **M;** Predicted 3D structure of P2RX7 protein by I-TASSER, mutation at 150th position from Gly to Arg creating 2 new hydrogen bonds (yellow dotted line) with neighbouring 133rd Arg amino acid which might introduce more rigidity to the protein conformation.

Subject 2, was presented the disease at 14 years of age with anaemia during regular blood test (Figure 2A). She did not take any transfusion for the last 16 years and maintained a steady state hemoglobin level of 8g/dl. She had a splenomegaly of 6 cm from the left costal margin. (Table 1). The sanger sequencing of the *HBB* gene of this subject also confirms the β thalassemia disease with [IVS1-5(G>C) and CD26(G>A)]. Both the parents are found to be carrier for β thalassemia. She representing as a non-severe thalassemia subject.

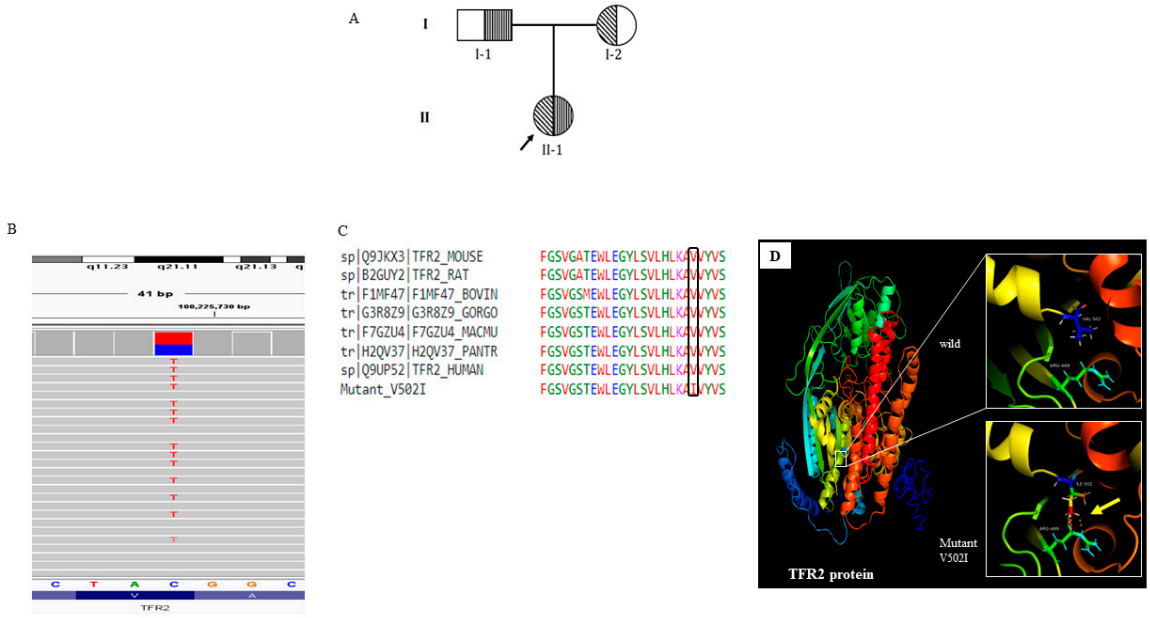


**Figure 2. A;** Pedigree of family-2, index II-1 (marked by arrow) is non-severe and both the parents are carriers. **B;** IGV view of sequence BAM file, shows that the *TFR2* (rs150618729) variant are in heterozygous condition with high sequence coverage **C;** Multiple sequence alignment by Clustal Omega, of TFR2 gene, and the conserved amino acid position marked in box. **D;** Predicted 3D structure of human TFR2 protein by I-TASSER, substitution from Val> Ile at 502nd position creating steric clash (orange dotted line) with Arg at 469th position at adjacent coil region might destabilize the protein structure.

We performed trio based extended whole exome sequencing (WES) of six subjects , two index cases as well as, father, mother of each one. For WES, DNA from the peripheral blood was extracted from the subjects. Agilent Sure select XT Human all Exome V6+UTR kit (Agilent, USA) was used for library preparation. Sequence run was performed in Illumina HiSeq 2500 platform. The sequence data was aligned using human reference genome, hg19 and variant calling file (VCF) was generated and annotated with VarAFT tool[5]. To find out the significant genetic variants we perform

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 January 2023 doi:10.20944/preprints202103.0088.v2



both de novo variant and inherited variants searching (Supplementary Fig-1, Supplementary Fig-2). Among the total variants, only the non-synonymous, exonic, splice site, indel, stop-gain, stop-loss variants were considered. Variants were further filtered MAF < 0.01as per 1000genome data base[6]. Variants with damaging prediction, either by SIFT, PolyPhen2, and Mutation tester were retained. These variants were compared between both severe and non-severe patient and only unique variants in both patients were further selected, as per the guidelines of the American College of Medical Genetics (ACMG) 2015[7]. Only the uncertain significant, pathogenic, and likely pathogenic variants were considered for further functional prediction /structural analysis. Function and disease association of each gene of respective variants were obtained from exhaustive literature survey as well as the public database, Gene Cards[8] and OMIM[9]. The impact of the variants on the gene product were predicted using SOPMA[10] and I-Mutant 2.0[11] tools. The tertiary protein structure were obtained from RSCB PDB database[12]. The 3D stricture were generated by I-TASSER[13,14] . To compare the alteration between wild and mutant proteins, PyMol was used[15]. Multiple sequence alignment was performed using Clustal Omega[16] tools to examine the sequence conservativeness of the translated protein sequence.

**Result and Discussion**

In case of severe subject (Subject 1), we identified 4 significant inherited variants. We don't find any significant de novo variants in this study

The identified 4 inherited variants in the severe subject are of in gene *RHAG* (rs553258732), *TTLL4* (rs116116034), *ATOH8* (rs558361422), *P2RX7* (rs28360447) (Fig;1.B-E). All the variants were disease causing and present in the conserved region (Figure 1 F-I) We further check the protein level changes of these variants using the bioinformatic tools and found that all the variants were significantly changes in the protein secondary.

structure and stability (Table 2). The tertiary structure of the altered protein was also significantly changes (Figure 1.J-M).

*RHAG* encodes the rhesus associated glycoprotein. Mutated RhAG leads to Overhydrated hereditary Stomatocytosis (OHST). Previously, it was reported that reduced or null expression of RhAG is associated with osmotically fragile erythrocyte and exhibit cation leak which leads to mild to moderate macrocytic hemolytic anemia[17].

TTLL4 is a member of the tubulin ligase like family protein and abundantly expressed in the hematopoietic tissue of bone marrow and maintains the structural organization of RBC cytoskeleton[18]. A recent study reported that *TTLL4* knock out mice have greater average diameter of RBCs than that of the wild type and showed greater vulnerability to oxidative stress leading to erythrolysis[19]. Mutation in these 2 major membrane proteins might destabilize the structural integrity of RBC membrane, making it much more fragile which can lead to hemolysis thus exacerbating anaemic state and enforcing severe condition.

**Table 2.** Selected Variants and their effect on the protein structure.

| Subject | Gene | db SNP ID | Predictive Effect on the Secondary structure of the protein** |
|---|---|---|---|
| Severe | RHAG | rs553258732 | Alpha helix: Increases. Extended strand: Decreases. Beta turn: Increases. Random coil: Decreases |
| | TTLL4 | rs116116034 | Alpha helix: No change, Extended strand: No change, Beta turn: Decreases. Random coil: Increases |
| | ATOH8 | rs558361422 | Alpha helix: Increases, Extended strand: Decreases. Beta turn: Decreases. Random coil: Decreases |
| | P2RX7 | rs28360447 | Alpha helix: Decreases, Extended strand: No change. Beta turn: No change. Random coil: Increases |

| | | | |
|---|---|---|---|
| Non-Severe | TFR2 | rs150618729 | Alpha helix: Increases, Extended strand: Decreases, Beta turn: Decreases, Random coil: Increases |

*ATOH8* gene, is a positive regulator of hepcidin transcription in the liver, which maintains iron homeostasis through regulating its own synthesis[20]. It has been reported that, hepcidin and ATOH8 expressions are downregulated in β thalassemia mice[21]. Mutation in this gene may causes the iron overload in the subject. Further evaluation reports that liver iron concentration is 6.1 mg/g in dry tissue in this subject, which is far more than the normal level.

P2RX7 is an ATP gated cation channel, previously it has been reported that, P2RX7 mediates ROS production in mature RBCs[22]. Mutation in the *P2RX7* gene may lead to various consequences that disrupt cellular homeostasis, and may induce severe phenotype.

In case 2 (Non -sever subject), only single inherited variant in the gene *TFR2* (rs150618729) was identified. (Figure 2. B; Supplementary table-2). Protein modelling confirms the tertiary structure changes (Figure 2.). *TFR2* encodes a transmembrane glycoprotein of transferrin receptor like family, which regulate iron homeostasis[23]. A recent study reported that, *TFR2* knockout β thalassemia mice sustained amelioration of anemia through improvement of the RBC count. Haploinsufficiency of TFR2 triggers erythroid cells to became more susceptible to erythropoietin (EPO) stimulation which helps erythropoiesis and improved anemic condition[24]. Thus this variant may help in the non-severe subject as better clinical outcome.

In the present whole exome based pilot study, it has been addressed the phenotype mystery or extreme clinical variability with same globin genotype (Beta globin, as well as alpha globin). We have sequenced the parental members of the index cases, anticipating the effect of any sporadic mutation for this clinical mystery. But didn't get any responsible variants out of sporadic mutation.

To the best of our knowledge, this is the first report in the world, explaining diverse clinical presentation of thalassemia with same globin mutations. From our study, it has been hypothesized that if the patient co associated with severe mutation in the gene/s responsible for RBC membrane, cytoskeleton formation, iron homeostatic, ROS generations, may be presented with severe clinical outcome, having beta globin genotype β +/ β 0 genotype.

In conclusion, hypothesis generated out of this study need to be validated, through the multi-omics study with large sample number. Generally, β 0/ β 0 genotype, would be severe clinical outcome, by the loss of two beta globin allele for β o mutation, but β +/ β 0, would have non -severe type. Thus, clue of the present study can helps to formulate suppurative therapy for β +/ β 0 thalassemia of severe subjects.

**Acknowledgments:** The authors are grateful to the Department of Biotechnology, Govt of India for funding this work [Sanc.No-BT/PR26461/MED/12/821/2018]. Authors are also express due acknowledgment to National Institute of Biomedical Genomics (NIBMG), Kalyani for providing necessary WES work.

**Authorship:** DS and DP performed the data collection, analysis prepared the initial version of manuscript, PKC did the patient management, KN and GS did the clinical confirmation, SB did data collection. AB did the overall planning and data validation and finalizing the manuscript.

## References

1. Giardine B, Borg J, Viennas E, et al. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res. 2014 Jan;42(Database issue):D1063-9. doi: 10.1093/nar/gkt911.. http://globin.cse.psu.edu/hbvar/menu.html [Accessed May 19, 2020] http://globin.bx.psu.edu/cgi-bin/hbvar/query_vars3
2. Sharma V, Kumar B, Kumar G, et al. Alpha globin gene numbers: an important modifier of HbE/beta thalassemia. *Hematology*.2009;14(5):297-300.doi:10.1179/102453309X446126
3. Saha D, Chowdhury PK, Panja A, et al. Effect of deletions in the α-globin gene on the phenotype severity of β-thalassemiaassemia. Hemoglobin. 2022;46(2):118-123.
4. Perera S, Allen A, Silva I, et al Genotype-phenotype association analysis identifies the role of α globin genes in modulating disease severity of β thalassaemia intermedia in Sri Lanka. Sci Rep. 2019 Jul; 12;9(1):10116. doi: 10.1038/s41598-019-46674-y.

5. Jean-Pierre Desvignes, Marc Bartoli, Valérie Delague, et al. VarAFT: a variant annotation and filtration system for human next generation sequencing data, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W545–W553, https://doi.org/10.1093/nar/gky471
6. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
7. Richards S, Aziz N, Bale S, et al; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30.
8. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics*. 2016;54:1.30.1-1.30.33. Published 2016 Jun 20. doi:10.1002/cpbi.5
9. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005;33(Database issue): D514-D517. doi:10.1093/nar/gki033
10. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995 Dec;11(6):681-4. doi: 10.1093/bioinformatics/11.6.681.
11. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res.2005 Jul 1;33(Web Server issue):W306-10. doi: 10.1093/nar/gki375.
12. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235
13. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinform 2008; 9(40).
14. The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D480–D489, https://doi.org/10.1093/nar/gkaa1100
15. PyMOL: The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Web site. http://www.pymol.org
16. Sievers F, Wilm A, Dineen D, et al. and Higgins D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 2011; 7:539
17. Jamwal M, Aggarwal A, Sachdeva MUS, et al. Overhydrated stomatocytosis associated with a complex RHAG genotype including a novel de novo mutation. J Clin Pathol. 2018; 71(7):648-652. doi:10.1136/jclinpath-2018-205018
18. Ijaz F, Hatanaka Y, Hatanaka T, et al. Proper cytoskeletal architecture beneath the plasma membrane of red blood cells requires *Ttll4*. Mol Biol Cell. 2017 Feb 15;28(4):535-544. doi: 10.1091/mbc. E16-02-0089
19. Ijaz F, Hatanaka Y, Hatanaka T, et al. Proper cytoskeletal architecture beneath the plasma membrane of red blood cells requires *Ttll4*. Mol Biol Cell. 2017 Feb 15;28(4):535-544. doi: 10.1091/mbc. E16-02-0089
20. Park CH, Valore EV, Waring AJ, Ganz T. Hepcidin, a urinary antimicrobial peptide synthesized in the liver. J Biol Chem. 2001 Mar 16;276(11):7806-10. doi: 10.1074/jbc.M008922200.
21. Upanan S, McKie AT, Latunde-Dada GO, et al. Hepcidin suppression in β-thalassemiaassemia is associated with the down-regulation of atonal homolog 8. Int J Hematol. 2017 Aug;106(2):196-205.
22. Wang B, Sluyter R. P2X7 receptor activation induces reactive oxygen species formation in erythroid cells. Purinergic Signal. 2013 Mar;9(1):101-12. doi: 10.1007/s11302-012-9335-2.
23. Silvestri L, Nai A, Pagani A, Camaschella C. The extrahepatic role of TFR2 in iron homeostasis. *Front Pharmacol*. 2014;5:93. Published 2014 May 7. doi:10.3389/fphar.2014.00093
24. Artuso I, Lidonnici MR, Altamura S, et al. Transferrin receptor 2 is a potential novel therapeutic target for β-thalassemiaassemia: evidence from a murine model [published correction appears in Blood. 2019 Jul 4;134(1):94]. *Blood*. 2018;132(21):2286-2297. doi:10.1182/blood-2018-05-852277