Unannotated single nucleotide polymorphisms in the TATA box of erythropoiesis genes show in vitro positive involvements in cognitive and mental disorders

Mikhail Ponomarenko1,2*, Ekaterina Sharypova1, Irina Drachkova1, Irina Chadaeva1, Olga Arkova3, Olga Podkolodnaya1, Petr Ponomarenko1, Nikolay Kolchanov1 and Ludmila Savinkova1

From 11th International Young Scientists School "Systems Biology and Bioinformatics" – SBB-2019
Novosibirsk, Russia. 24-28 June 2019

Abstract

Background: Hemoglobin is a tetramer consisting of two α-chains and two β-chains of globin. Hereditary aberrations in the synthesis of one of the globin chains are at the root of thalassemia, one of the most prevalent monogenic diseases worldwide. In humans, in addition to α- and β-globins, embryonic zeta-globin and fetal γ-globin are expressed. Immediately after birth, the expression of fetal Aγ- and Gγ-globin ceases, and then adult β-globin is mostly expressed. It has been shown that in addition to erythroid cells, hemoglobin is widely expressed in nonerythroid cells including neurons of the cortex, hippocampus, and cerebellum in rodents; embryonic and adult brain neurons in mice; and mesencephalic dopaminergic brain cells in humans, mice, and rats. Lately, there is growing evidence that different forms of anemia (changes in the number and quality of blood cells) may be involved in (or may accompany) the pathogenesis of various cognitive and mental disorders, such as Alzheimer’s and Parkinson’s diseases, depression of various severity levels, bipolar disorders, and schizophrenia. Higher hemoglobin concentrations in the blood may lead to hyperviscosity, hypovolemia, and lung diseases, which may cause brain hypoxia and anomalies of brain function, which may also result in cognitive deficits.

Methods: In this study, a search for unannotated single-nucleotide polymorphisms (SNPs) of erythroid genes was initially performed using our previously created and published SNP-TATA_Z-tester, which is a Web service for computational analysis of a given SNP for in silico estimation of its influence on the affinity of TATA-binding protein (TBP) for TATA and TATA-like sequences. The obtained predictions were finally verified in vitro by an electrophoretic mobility shift assay (EMSA).

(Continued on next page)
Background
According to World Health Organization estimates, ~7% of the world population have hereditary aberrations in the synthesis of hemoglobin, meaning that these are the most prevalent monogenic disorders [1]. Alpha- and β-thalassemia and sickle cell anemia are the most prevalent autosomal recessive diseases caused by mutations in the genes of α- and β-globin. β-Thalassemia was first documented in the Mediterranean, Middle-Eastern, and Asian regions, whereas sickle cell anemia was first documented in Central Africa. As a result of increasing population migration, these diseases are currently global and therefore represent a growing problem for health care in many countries. The spread of these diseases is also promoted by natural selection favoring carriers of hemoglobin synthesis aberrations who are relatively resistant to the malaria caused by Plasmodium falciparum [2].

It is known that human hemoglobin, the protein that carries oxygen from the lungs to various tissues, is a tetramer consisting of two α-subunits and two β-subunits of globin. Aberration one of the globin chains are the reason of thalassemia that is one of the most prevalent monogenic diseases all over worldwide [3]. Patients with severe thalassemia have serious problems related to life-long blood transfusion dependence; besides, they develop severe anemia, chronic hemolysis, and iron overload of various organs as a result of inefficient erythropoiesis, which launches compensatory mechanisms leading to, among other problems, elevated iron absorption in the heart and gastrointestinal tract [4], hepatosplenomegaly, and such complications as cardiopathy and endocrine disturbances [4]. Patients with mild thalassemia may not experience the disease manifestations and may not have anemia. α-Thalassemia, caused by deletion of one or both α-globin genes located on chromosome 16, is characterized by an almost asymptomatic clinical phenotype. The deletion of three or four genes of α-globin (Bart disease is caused by the deletion of all four genes) is characterized by a severe or lethal form of hemolytic anemia, respectively [5], where α-globin is not synthesized at all, and this condition causes fetal death (prenatal or at birth) [6]. α-Thalassemia is associated with a fairly rare syndrome of intellectual disability, ATR-X (α-thalassemia X-linked mental retardation), which features moderate or severe intellectual disability, delayed development, and other problems [7].

In addition to α- and β-globins, there are embryonic zeta-globin and fetal γ-globin in humans. Immediately after birth, the expression of fetal Aγ- and Gγ-globin stops, and then adult β-globin is expressed mostly [8].

It has been reported that in addition to erythroid cells, hemoglobin is widely expressed in nonerythroid cells including neurons of the cortex, hippocampus, and cerebellum in rodents [9]; embryonic and adult brain neurons in mice [10]; and mesencephalic dopaminergic brain cells in humans, mice, and rats [11]. Lately, there is growing evidence that different forms of anemia may be involved in or accompany the pathogenesis of various cognitive and mental disorders, such as Alzheimer’s and Parkinson’s diseases, depression of various severity levels, and bipolar disorders [12–14]. Severe depressive disorders are the third leading cause of disability worldwide [15]. Bipolar disorder is a severe psychiatric illness that starts as mild depression and brief hypomania and progresses to the acute phase with manic episodes [16].

It is known that the brain responds to a lowered hemoglobin level during anemia with vasodilation increasing cerebral blood flow. Nonetheless, when hemoglobin concentration drops substantially, the aforementioned compensatory mechanism is not triggered, and cerebral blood flow becomes insufficient, leading to brain function deficits and ischemia [17]. Both clinical [18, 19] and subclinical ischemia [20] are risk factors of mild cognitive impairment and dementia. The associations with anemias, in particular with microcytic anemia, which is most often iron deficient, may be related to oxidative stress in the brain, as a consequence of increasing population migration, these diseases are currently global and therefore represent a growing problem for health care in many countries. The spread of these diseases is also promoted by natural selection favoring carriers of hemoglobin synthesis aberrations who are relatively resistant to the malaria caused by Plasmodium falciparum [2].

It is known that human hemoglobin, the protein that carries oxygen from the lungs to various tissues, is a tetramer consisting of two α-subunits and two β-subunits of globin. Aberration one of the globin chains are the reason of thalassemia that is one of the most prevalent monogenic diseases all over worldwide [3]. Patients with severe thalassemia have serious problems related to life-long blood transfusion dependence; besides, they develop severe anemia, chronic hemolysis, and iron overload of various organs as a result of inefficient erythropoiesis, which launches compensatory mechanisms leading to, among other problems, elevated iron absorption in the heart and gastrointestinal tract [4], hepatosplenomegaly, and such complications as cardiopathy and endocrine disturbances [4]. Patients with mild thalassemia may not experience the disease manifestations and may not have anemia. α-Thalassemia, caused by deletion of one or both α-globin genes located on chromosome 16, is characterized by an almost asymptomatic clinical phenotype. The deletion of three or four genes of α-globin (Bart disease is caused by the deletion of all four genes) is characterized by a severe or lethal form of hemolytic anemia, respectively [5], where α-globin is not synthesized at all, and this condition causes fetal death (prenatal or at birth) [6]. α-Thalassemia is associated with a fairly rare syndrome of intellectual disability, ATR-X (α-thalassemia X-linked mental retardation), which features moderate or severe intellectual disability, delayed development, and other problems [7].

In addition to α- and β-globins, there are embryonic zeta-globin and fetal γ-globin in humans. Immediately after birth, the expression of fetal Aγ- and Gγ-globin stops, and then adult β-globin is expressed mostly [8].

It has been reported that in addition to erythroid cells, hemoglobin is widely expressed in nonerythroid cells including neurons of the cortex, hippocampus, and cerebellum in rodents [9]; embryonic and adult brain neurons in mice [10]; and mesencephalic dopaminergic brain cells in humans, mice, and rats [11]. Lately, there is growing evidence that different forms of anemia may be involved in or accompany the pathogenesis of various cognitive and mental disorders, such as Alzheimer’s and Parkinson’s diseases, depression of various severity levels, and bipolar disorders [12–14]. Severe depressive disorders are the third leading cause of disability worldwide [15]. Bipolar disorder is a severe psychiatric illness that starts as mild depression and brief hypomania and progresses to the acute phase with manic episodes [16].

It is known that the brain responds to a lowered hemoglobin level during anemia with vasodilation increasing cerebral blood flow. Nonetheless, when hemoglobin concentration drops substantially, the aforementioned compensatory mechanism is not triggered, and cerebral blood flow becomes insufficient, leading to brain function deficits and ischemia [17]. Both clinical [18, 19] and subclinical ischemia [20] are risk factors of mild cognitive impairment and dementia. The associations with anemias, in particular with microcytic anemia, which is most often iron deficient, may be related to oxidative stress in the brain, as a consequence

| Keywords: Erythropoiesis, Cognitive disorder, In silico prediction, TBP/TATA affinity, kinetics of TBP-TATA binding in vitro, TATA-binding protein, SNP marker |
of poor or ineffective iron assimilation [21]. Elevated hemoglobin levels can cause blood hyperviscosity, hypervolemia, and lung diseases, which may result in brain hypoxia and anomalies of brain function, which may also cause cognitive deficits [22]. The incidence and prevalence of anemia and hyperhemoglobinemia increase with age [23]. This state of affairs has prompted more targeted research into hematopoiesis disturbances to elucidate their mechanisms in relation to possible cognitive deficits.

Previously, by annotating experimental results from various authors, we have compiled a list of single-nucleotide polymorphism (SNP)-containing TATA boxes of human genes [24] associated with diverse hereditary monogenic diseases. The link between SNPs and diseases has been proven by those authors both molecularly and clinically. By means of the stepwise interaction of TATA-binding protein (TBP) with a TATA box [25], the sequences of TATA boxes with the surrounding nucleotides have been analyzed, and in silico predictions have been made regarding the influence of each SNP on TBP–TATA affinity. Subsequently, experimental in vitro verification of these in silico predictions has been conducted [26, 27] revealing a coefficient of correlation \( r \); between the prediction and experiment) of 0.822 at \( p < 10^{-7} \).

After the completion of the “1000 Genomes” project [28], more than 10 thousand individual human genomes have been successfully sequenced [29], and \( 8.6 \times 10^9 \) possible SNPs have been deposited in the dbWGFP database [30]. Undoubtedly, these numerous SNPs require preliminary filtration in silico before costly and labor-intensive studies in vitro, ex vivo, and in vivo. With a high degree of certainty, it can be assumed that the number of identified SNPs in TATA boxes has increased substantially since 2009. These observations motivated us to update the list of previously documented TATA box SNPs in the promoters of human hemoglobin genes. Using publicly available online databases, in this work we selected unannotated SNPs of the promoters of genes associated with erythropoiesis, which were analyzed by means of our public Web service SNP_TATA_Z-tester [31], developed within the framework of the stepwise model for TBP binding to TATA-like boxes [25], namely: (i) TBP slides along DNA ↔ (ii) TBP stops at a TBP-binding site [TATA-box] ↔ (iii) the TBP–promoter complex is fixed by a 90° bend of DNA, – as in vitro independently observed experimentally in solution [32]. The Web service has been used by us previously to predict candidate SNP markers in TATA boxes in the context of real-world promoters of genes related to reproductive potential [33], aggressiveness [34], Alzheimer’s disease [35], obesity [36], chronopathologies [37], and autoimmune diseases [37].

In the present study, after in silico analysis of 161 unannotated SNPs, 45 candidate SNP markers were predicted for 15 out of the 25 genes participating in erythropoiesis and showing changes in TBP–TATA box affinity in the promoter. Some of them were chosen for experimental validation in vitro. A possible link between cognitive deficits and the influence of SNPs from TATA boxes and TATA-like sequences on the promoters of hematopoiesis genes is also discussed.

**Methods**

**DNA sequences**

We performed a search and retrieval of data from the databases of unannotated SNPs in the TATA boxes of erythroid genes. We employed Web service SNP_TATA_Z-tester [31], developed within the framework of the stepwise model for TBP binding to TATA-like boxes [25]. Using this software, we analyzed unannotated substitutions in the TATA boxes of human hemoglobin genes (HBZ, HBB, HBD, and HBG1) and the genes encoding the enzymes participating in heme biosynthesis (ALASI, CAI, EPOR, and GYPC) and other proteins (involved in erythropoiesis): a total of 161 SNPs from 25 human genes. These data are publicly available thanks to Web service UCSC Genome Browser [38]. Computational analysis of \( 8.58 \times 10^9 \) possible SNPs throughout the whole human genome, predictions, experimental data, and the relevant clinical findings accumulated in database dbWGFP may accelerate and orient the clinical search for biomedical markers [30].

Genes directly participating in the processes of proliferation, differentiation, and maturation of erythrocytes in humans were retrieved from database ESGR-TRRD (erythroid-cell–specific regulated genes in TRRD format), created by Podkolodnaya O.A. and Stepanenko I.L. ([http://wwwmgs.bionet.nsc.ru/mgs/papers/podkolodnaya/esg-trrd](http://wwwmgs.bionet.nsc.ru/mgs/papers/podkolodnaya/esg-trrd)) for humans, chickens, rabbits, mice, rats, and zebrafish. The numbers of promoters, their sequences in the region [–100; −1] relative to the transcription start site in the 22 erythroid genes under study as well as the presence or absence of TATA boxes in these genes were determined in the Eukaryotic Promotor Database (EPD). Physical coordinates of each promoter were found in database UCSC (build 38 or GRCh38/hg38) via the BLAST software (rapidly aligning sequences to the genome). The SNPs present in promoters were retrieved from the UCSC database (build 38 or GRCh38/hg38) via a resource called Browser (to download data from the Genome Browser database). Sequence data for the substitutions and their populational characteristics were obtained from database dbSNP [39] (Database of Single Nucleotide Polymorphisms) at the National Center for Biotechnology Information (NCBI; Bethesda, MD, USA).

**DNA sequence analysis in silico**

We analyzed DNA sequences between nucleotide positions –100 and –1 upstream of the protein-coding regions in the human genes retrieved from the human
reference genome using our Web service SNP_TATA_Z-tester.

Synthetic double-helical deoxyoligonucleotides (ODNs)
The ODNs corresponding to ancestral and minor alleles of the selected SNPs were synthesized and purified (Biosset, Novosibirsk, Russia).

Labeling of the ODNs with $^{32}$P
The ODNs were 26 bp long; they were synthesized and additionally purified by polyacrylamide gel (PAAG) electrophoresis at Biosset (Novosibirsk). To prepare a labeled double-stranded ODN, both strands were labeled with $^{32}$P-ATP (Biosan, Novosibirsk) by means of T$^4$ poly nucleotide kinase (SibEnzyme, Novosibirsk), annealed at 95 °C (in the equimolar ratio), and were slowly cooled (no less than 3 h) to room temperature. The annealed duplexes were purified and analyzed by electrophoresis in a 15% PAAG under nondenaturing conditions with subsequent autoradiography on a Molecular Imager PharosFX Plus phosphorimager (Bio-Rad). Unlabeled duplexes were prepared in the same way and were used after the additional purification by electrophoresis.

Preparation and purification of recombinant full-length human TBP
In this study, recombinant human TBP (hTBP) was employed that was expressed in Escherichia coli BL21(DE3) cells from plasmid pAR3038-hTBP (kindly provided by Prof. B. Puhg, Center for Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA). Transformation of E. coli BL21(DE3) was performed as described elsewhere [41]. Expression and purification of TBP were carried out as described previously [42], except for isopropyl β-D-1-thiogalactopyranoside (IPTG) concentration (1.0 instead of 0.1 mM) and induction duration (3.0 instead of 1.5 h).

Determination of the association and dissociation rate constants of TBP–ODN complexes
Experiments on TBP–ODN binding were conducted at 25 °C in a buffer consisting of 20 mM HEPES-KOH (pH 7.6), 5 mM MgCl$_2$, 70 mM KCl, 1 mM dithiothreitol, 100 μg/ml BSA, 0.01% of NP-40, and 5% of glycerol with a fixed concentration of active TBP (usually 0.5 nM).

To determine the kinetic constants of the association and dissociation of complexes, at least three independent experiments were conducted. Each experiment on kinetics usually included 32 binding reactions (eight time points, each with four concentrations of an ODN). All four binding reactions (one time point) were started simultaneously by the addition of TBP into the reaction mixture containing an ODN, with immediate transfer to a thermostat at 25 °C. Upon completion of the binding reaction, all the reaction mixtures were applied to a PAAG concurrently at electric-field intensity of 10 V/cm. TBP–ODN complexes were separated from the unbound ODN by an electrophoretic mobility shift assay. The electrophoresis was performed in a native 5% PAAG based on Tris-glycine buffer (pH 8.3) for 40 min at 10°C and electric-field intensity 25 V/cm. The gels were dried and used for exposure of a screen, Imaging Screen-K (Kodak), on the Molecular Imager PhorosFX Plus phosphorimager (Bio-Rad). The screen was scanned on the phosphorimager, and the autoradiograph was analyzed quantitatively in the Quantity One software, version 4.5.0 (Bio-Rad). Velocity constants and equilibrium constants of dissociation of TBP–ODN complexes were calculated in GraphPad Prism 5 (GraphPad Software, San Diego, CA, USA).

Statistical analysis
A comparison of our predictions with the experimental values of changes in TBP–TATA affinity after the substitutions in TATA boxes was conducted by means of two options, “Multiple Regression” and “Nonparametrics,” in the standard software STATISTICA (Statsoft®, Tulsa, USA). Changes in $K_D$ ($\delta$) with an SNP in the TATA box were calculated as $-\ln [K_D,\text{TATA}_\text{mut}] – (-\ln [K_D,\text{TATA}])$.

Results and DISCUSSIONS
In silico analysis of human erythropoiesis genes
Tables 1 and 2 show unannotated SNPs of TATA boxes from the promoters of erythropoiesis genes; these SNPs were retrieved from publicly available databases. Predictions were made via Web service SNP_TATA_Z-tester, which is a modified version of the SNP_TATA_Z-tester Web service [31], as illustrated in the example presented in Fig. 1.

The cluster of human β-globin genes consists of five genes ($\delta^\epsilon$-γG-γA-δ-β-3′) and a locus control region [44]. Clinical manifestations of β-thalassemia (MIM 613985) mostly depend on the mutations in the regulatory and structural regions of the β-globin gene [45]. In the severe form of the disease, these manifestations consist of stunted growth, paleness, jaundice, hepatosplenomegaly, skeletal changes, and other problems as well as complications related to iron overload, including a wide range of endocrine disturbances (stunted growth, disorder of sexual maturity, diabetes mellitus and hypothyroidism, pituitary failure, and more rarely, adrenal insufficiency), dilated cardiomyopathy, and liver fibrosis and cirrhosis [46]. In patients with intermediate thalassemia, osteoporosis is present, as is the involvement of the spleen, liver, and lymph nodes as well as bone deformity, gallbladder stones, and higher susceptibility to thromboses. Mild thalassemia may be asymptomatic. It has been demonstrated that in the human genes encoding β- and δ-hemoglobins (HBB
Table 1 Predictions of the influence of SNPs in the promoters of the hemoglobin subunit genes on TBP–TATA affinity, as performed by SNP-TATA-Z-tester [31]

| Gene        | Alleles, WT or mut | DNA sequence, TATA-like box; S^2\,^3* | Prognosis, −ln K_D |
|-------------|--------------------|--------------------------------------|-------------------|
| HBB         | WT                 | cagggctggccatataaactcagggcagag        | 19.20             |
|             | rs62750400:G       | cagggctggccatataaactcagggcagag        | 19.12             |
|             | rs28186452:G       | cagggctggccatataaactcagggcagag        | 19.169            |
|             | rs28186452:G       | cagggctggccatataaactcagggcagag        | 17.849            |
|             | rs33980857:A       | cagggctggccatataaactcagggcagag        | 17.701            |
|             | rs33980857:C       | cagggctggccatataaactcagggcagag        | 18.172            |
|             | rs33980857:G       | cagggctggccatataaactcagggcagag        | 17.673            |
|             | rs33980857:G       | cagggctggccatataaactcagggcagag        | 18.553            |
|             | rs33980857:G       | cagggctggccatataaactcagggcagag        | 18.062            |
|             | rs34500389:G       | cagggctggccatataaactcagggcagag        | 20.175            |
|             | rs34500389:A       | cagggctggccatataaactcagggcagag        | 18.635            |
|             | rs33931746:C       | cagggctggccatataaactcagggcagag        | 18.341            |
| HBB _2      | WT                 | ggaccagcagccaaaggccagggcagag          | 19.29             |
|             | rs35518301:G       | ggaccagcagccaaaggccagggcagag          | 18.65             |
| HBZ         | WT                 | cagctctctgtataaaggccacctg             | 20.76             |
|             | rs113180943:T      | cagctctctgtataaaggccacctg             | 19.28             |
|             | rs1035033590:C     | cagctctctgtataaaggccacctg             | 20.773            |
| HBZ _3      | WT                 | ggagggtggtggcccctatctctctct           | 17.611            |
|             | rs559282746:G      | ggagggtggtggcccctatctctctct           | 17.605            |
| HBA1        | WT                 | gcgcaccagcataaaccctgggcgcgc           | 18.36             |
|             | rs571582665:T      | gcgcaccagcataaaccctgggcgcgc           | 18.62             |
| HBG1        | WT                 | agggtctccctttttättctttcctc            | 19.06             |
|             | rs57324152:G       | agggtctccctttttättctttcctc            | 17.54             |
| HBM         | WT                 | ggtgcggccagggctctataagggaggc          | 19.13             |
|             | rs90573014:8:A     | ggtgcggccagggctctataagggaggc          | 19.23             |
|             | rs937192507:C      | ggtgcggccagggctctataagggaggc          | 19.95             |
| HBE1        | WT                 | aggccaaaagagagctttgtgtaggc            | 18.31             |
|             | rs768296486:A      | aggccaaaagagagctttgtgtaggc            | 18.16             |
|             | rs1036491581:G     | aggccaaaagagagctttgtgtaggc            | 18.01             |

**Notes:** Ancestral (WT) and minor (mut) alleles; K_D, dissociation constant of the TBP–DNA complex [13]; Genes: HBB, β-globin; HBB _2, β-globin promoter 2; HBZ zeta-globin; HBZ _3 zeta-globin promoter 3; HBA1, α1-globin; HBG1, αy globin; HBM, Hemoglobin subunit μ; and HBE1, ε-globin.

and HBD, respectively), six of many regulatory SNP markers (rs33981098, rs33980857, rs34598529, rs3393 1746, rs397509430, and rs35518301) weaken the affinity between TBP and the promoter of these genes and cause both thalassemia and resistance to malaria [47]. By contrast, one SNP marker (rs34500389) that enhances this affinity was found only once during the screening of newborns in Spain without any symptoms of thalassemia, i.e., this SNP allele was assumed to be harmless [48].

The zeta-globin gene (HBZ) codes for an α-polypeptide that is synthesized in the yolk sac of an early embryo, whereas α-globin is produced throughout the lifespan of the fetus and adult person. HBZ is a member of the cluster of human α-globin genes, which includes five functional genes and two pseudogenes [49]. This cluster features deletions of long segments (the most prevalent ones are the so-called Mediterranean deletion, Southeastern deletion, and Thai deletion), including α-globin genes, thereby causing α-thalassemia [50, 51]. The Zeta2 gene is expressed at the embryonic stage and is also subject to deletion [51].

In hemoglobinopathies, mutations in DNA simultaneously influence the expression of several genes, e.g., γ-globin genes (HBG1 and HBG2), responsible for the synthesis of γ-chains, which are present in fetal hemoglobin (HbF) [52]. Regulatory sequences of the 5′ region of HBG1 and HBG2 promoters contain at least eight SNP markers, which are located outside the core promoter region studied by us [−100; −1]. The main function of promoter elements of HBG1 and HBG2 is the binding to transcription complexes, which inhibit the expression of fetal hemoglobin at an appropriate stage [53, 54]. Unannotated rs573241527, which was detected by us, replaces T by G in the TATA box of the HBG1 promoter (Table 1) and is predicted to lower the affinity of the interaction with TBP.

The 5′-aminolevulinate synthase 1 gene (ALAS1) encodes a mitochondrial matrix enzyme, which catalyzes the first, rate-limiting stage of heme biosynthesis [55]. The most well-known role of ALAS1 is catalysis of the condensation of glycine and succinyl-coenzyme A for formation of 5′-aminolevulinate; the latter is a universal precursor of tetrapyrrole compounds participating in
Carbonic anhydrases (CA) are zinc metalloenzymes that catalyze reversible hydration of carbon dioxide to the bicarbonate ion and proton [56, 57]. It is reported that there is substantial tissue-specific expression and heme-mediated (as feedback loop) regulation of enzymes in the heme biosynthesis pathway [8]. In particular, it is known that ALAS exists in the form of tissue-specific isozymes, that is, erythroid and nonspecific ALAS, which are encoded by two different genes. Unannotated rs1046254329 (G > A)—detected by us in the TATA box of the promoter—is predicted to slightly enhance TBP–TATA affinity, as readers can see in Table 2.

Carbonic anhydrases (CA) are zinc metalloenzymes that catalyze reversible hydration of carbon dioxide to the bicarbonate ion and proton [58, 59]. At present, there are 15 known CA isozymes belonging to the α class. CA isozymes have different distribution profiles among tissues and organs and different subcellular localizations, catalytic activities, and physiological functions [60]. The carbonic anhydride family (CA1, − 2, and − 3) performs a multitude of functions in various species. It is known that CA1 takes part in homeostasis, mitochondrial respiration, and erythroid differentiation and may be involved in such pathological processes as anemia, chronic acidosis, and diabetic macular edema [60]. Unannotated rs538698304 detected by us in three separate projects on population research is regarded as rare: its minor allele frequencies are 0.00006, 0.0001, and 0.0002. This substitution in the middle of the TATA box of the CA1 promoter worsens the TATA box sequence, and according to our prediction, lowers TBP–TATA affinity. Therefore, this SNP may be a candidate marker of anemia.

In addition to erythropoiesis regulation in hematopoietic tissues, erythropoietin (EPO) functions in other tissues, including the nervous system. To perform its function, EPO/EPOR system has a wide spectrum of nonhematopoietic effects [64]. Functional receptors encoded by EPOR have been discovered in nonerythroid blood cell lineages, such as lymphocytes, myeloid cells, and megakaryocytes. Besides, EPOR is widely expressed in

---

**Table 2** Predictions regarding the influence of SNPs in the promoters of erythroid genes (except the hemoglobin subunit genes) on TBP–TATA affinity, as performed by SNP–TATA-Z-tester [31]

| Gene         | Alleles: WT or mut | DNA sequence, TATA-like box; S'-3' | Prognosis, −ln K0 |
|--------------|--------------------|-----------------------------------|------------------|
| ALAS1        | WT                 | ctcctcggtattatattaaagggccgccc     | 20.11            |
|              | rs564394089A       | ctcACgtattatattaagggccgccc        | 20.23            |
|              | rs1046254329A      | ctcACgtatattataagggccgccc         | 20.40            |
|              | rs50053347T        | ctcACgtattatattaaagggccgccc       | 20.11            |
| CA1          | WT                 | ggaatgggcaagctttatagctcaggggg     | 19.81            |
|              | rs540950375T       | ggaatgggcaagctttatagctcaggggg     | 19.70            |
|              | rs991064314T       | ggaatgggcaagctttatagctcaggggg     | 19.82            |
| CA1_2        | WT                 | ctggggcattttttatagaaactact        | 20.22            |
|              | rs557418566G       | ctggggcattttttatagaaactact        | 20.16            |
|              | rs538698304C       | ctggggcattttttatagaaactact        | 19.61            |
| HOXB2        | WT                 | aggacctcagcgaatttacaggggaat       | 17.98            |
|              | rs762972656T       | aggacctcagcgaatttacaggggaat       | 17.92            |
| EPOR         | WT                 | cgtagcagcagaaaaatagtgacggtg       | 18.24            |
|              | rs1006576690T      | cgtagcagcagaaaaatagtgacggtg       | 19.39            |
|              | rs971717075C       | cgtagcagcagaaaaatagtgacggtg       | 18.11            |
|              | rs567946217A       | cgtagcagcagaaaaatagtgacggtg       | 19.33            |
|              | rs567946217C       | cgtagcagcagaaaaatagtgacggtg       | 18.51            |
| GYPB         | WT                 | gctcactagctgttatctttctcagcc       | 18.26            |
|              | rs3967038C         | gctcactagctgttatctttctcagcc       | 18.30            |
|              | rs3967038A         | gctcactagctgttatctttctcagcc       | 18.23            |
| GYPC         | WT                 | cattgggagtttctctgactctt           | 16.97            |
|              | rs970970552A       | cattgggagtttctctgactctt           | 17.83            |
|              | rs911469201A       | cattgggagtttctctgactctt           | 17.13            |
|              | rs989175270G       | cattgggagtttctctgactctt           | 17.27            |
| EDRF1        | WT                 | tcgcgagatttaattggcagttacag        | 18.45            |
|              | rs946240545G       | tcgcgagatttaattggcagttacag        | 18.34            |
|              | rs774081749G       | tcgcgagatttaattggcagttacag        | 17.76            |
|              | rs895690311G       | tcgcgagatttaattggcagttacag        | 18.18            |

**Notes:** See “Notes” under Table 1. Genes: ALAS1 Aminolevulinate synthase 1; CA1 Carbonic anhydrase 1; CA1_2 Carbonic anhydrase promoter 2; HOXB2 Homeobox B2; EPOR Erythropoietin receptor; GYPB Glycophorin B; GYPC Glycophorin C; and EDRF1 Erythroid differentiation regulatory factor 1
the central nervous system (brain) of embryos, fetuses, and adults [65]. The EPO–EPOR complex has an influence on the recovery and nutrition of brain tissues after injury [66].

The GYP C gene codes for two sialoglycoproteins on the erythrocyte surface in humans: glycophorin C (GPC) and glycophorin D (GPD) [67]. Parasite P. falciparum, which is a causative agent of malaria in humans, employs GPC to enter human erythrocytes. GYP C is widely expressed in the large intestine, esophagus, and 23 other organs or tissues. Studies on the molecular evolution of GYP C among hominids (great and smaller apes) and on polymorphism patterns at the respective locus [67] have revealed an excess of nonsynonymous between-species divergences, which apparently are related to accelerated evolution of GYP C in the Homo sapiens lineage.

DNA methylation is a hereditary epigenetic modification that stably changes the expression of genes involved in such a complex mental disorder as schizophrenia. Some researchers [68] have demonstrated that the
methylation profiles of \textit{GYP\textbf{C}}, which contributes to schizophrenia pathogenesis, significantly differ between female and male patients and between patients with schizophrenia and controls. Unannotated rs970970552, identified by us as a candidate SNP marker via a prediction made by means of SNP\textunderscore TATA\textunderscore Z-tester, increases the affinity between TBP and the TATA box of such a sequence.

**Experimental in vitro validation of the predictions about the influence of SNPs on TBP–TATA affinity**

For experimental in vitro validation, we chose SNP-containing TATA boxes of human genes \textit{HBB}, \textit{HBZ}, and \textit{EPOR}. Table 3 presents the results of in vitro verification of our predictions regarding the influence of substitutions in TATA boxes on TBP–TATA affinity. For instance, TBP–TATA affinity in the \textit{Hbz} promoter decreases 6.6-fold as a result of a substitution in the TATA box (rs113180943). SNPs in the TATA-like sequence of the promoter of \textit{EPOR} (−27\textgreater\textless A, −31C\textgreater\textless A, and −31C\textgreater\textless G) enhancing the TBP–TATA affinity 15.5-fold, 8.1-fold, and 2.1-fold, respectively. Additionally, Table 3 shows negative natural logarithms of the predicted and experimentally determined \(K_D\) values and changes in \(K_D\) (\(\delta\)) in the presence of the SNP in the TATA box: \(\delta = −\ln [K_{D,Mut}] − (−\ln [K_{D,WT}])\), and values of association and dissociation rate constants \(k_a\) and \(k_d\), respectively for TBP–TATA complexes, reflecting the velocities of their formation and dissociation. The results were obtained in an electrophoretic mobility shift assay (Fig. 2): dependences of reaction rates on deoxyoligonucleotide (hereinafter: ODN) concentrations — Fig. 2a; electropherograms, from which kinetic curves were derived — Fig. 2b, as well as one can see Additional file 1: Supplementary Electropherogram.

Experimental validation of the influence of the unannotated SNP (rs34500389 located in \textit{HBB}) changing the TATA box to a more canonical form (C\textunderscore 32ATAAAA AG → T\textunderscore 32ATAAAAAG) indicated, as predicted, an increase in TBP–TATA affinity (Table 3). This affinity for the minor allele was 1.56-fold stronger (Table 3; \(K_D = 50\text{ nM}\) for the ancestral allele and 32 nM for the minor allele) because of the greater (also 1.5-fold) velocity of formation of the complex. The dissociation velocity and lifetime of the complex barely changed. Perhaps the relatively small increase in affinity is explained by the presence of CATAAA in both flanks, i.e., a string of “AT” base pairs, which weakens the TBP–TATA binding.

The unannotated SNP (rs113180943, detected by us in the TATA box of the \textit{Hbz} promoter), which replaces the third T base (having the highest weight among the nucleotides in the TATA box sequence [69]) with “A,” according to the prediction, should weaken TBP–TATA affinity. Experimental in vitro validation suggested that this substitution indeed attenuates TBP–TATA affinity 6.6-fold (\(K_D = 6\text{ nM}\) for the ancestral allele and \(K_D = 40\text{ nM}\) for -29T\textgreater\textless A). The reason is that the velocity of

**Table 3** Kinetic and thermodynamic characteristics of TBP–TATA interactions

| Gene | ODN | Alleles: | Allele | Sequences (5\textsuperscript{\textprime} – 3\textsuperscript{\textprime} strands), 26 bp | Prediction | Experiments |
|------|-----|---------|-------|--------------------------------|------------|-------------|
|      |     | WT or mut |       |                                | -ln\(K_D\) | \(\delta\) | \(K_D\) nM | \(k_a\) M\textsuperscript{-1} s\textsuperscript{-1} | \(k_d\) s\textsuperscript{-1} | \(t_{1/2}\), min |
| HBB  | 1   | WT      | -32c > T | aggctgctggcataaaagtcagggcag | 19.20 +    | 0.98 | 16.81 + | 32 ± 3 (1.4 ± 0.110\(^4\)) | 7.1 ± 0.710\(^4\) | 16 ± 2 |
|      | 2   | rs34500389T |       | agggctgctggTataaaagtcagggcag | 20.18 +    | 0.98 | 17.26 ± 0.45 | 32 ± 3 (2.1 ± 0.110\(^4\)) | 6.6 ± 0.510\(^4\) | 17 ± 2 |
| HbZ  | 3   | WT      | -29 t > A | agctccctgtataaaagggacctgt | 20.76 −1.48 | 19.28 | 18.93 −1.90 | 6 ± 1 (4.5 ± 0.410\(^6\)) | 3 ± 0.610\(^4\) | 38 ± 4 |
|      | 4   | rs113180943A |       | agctccctgtatataaaagggacctgt | 17.03 + 4 | 15.59 | 17.30 + 4 | 40 ± 20 (2.3 ± 0.110\(^4\)) | 1.4 ± 0.410\(^6\) | 115 ± 15 |
| EPOR | 5   | WT      |       | cacgtcatctatTTTTTgctgctacg | 18.24 + 30 | 15.59 | 17.03 + 30 | 170 ± 30 (2.8 ± 0.103) | 4.6 ± 0.610\(^4\) | 25 ± 3 |
|      | 6   | rs1006576690A | -27 t > A | cacgtcatctatatTTTTTgctgctacg | 19.39 + 1.15 | 18.33 + 2.74 | 11 ± 3 (9.3 ± 0.710\(^3\)) | 1.0 ± 0.310\(^3\) | 15 ± 0 |
|      | 7   | rs971717705G | -29a > G | cacgtcatctatatTTTTTgctgctacg | 18.11 −0.13 | 15.71 + 0.12 | 150 ± 30 (2.4 ± 0.310\(^3\)) | 3.6 ± 0.510\(^4\) | 32 ± 4 |
|      | 8   | rs567946217A | -31c > A | cacgtcatctatatTTTTTgctgctacg | 19.33 + 1.09 | 17.68 + 2.09 | 21 ± 6 (6.8 ± 0.710\(^3\)) | 1.4 ± 0.410\(^4\) | 82 ± 6 |
|      | 9   | rs567946217G | -31c > G | cacgtcatctatatTTTTTgctgctacg | 18.51 + 0.27 | 16.34 + 0.75 | 80 ± 10 (3.6 ± 0.310\(^3\)) | 3.0 ± 0.510\(^4\) | 38 ± 4 |

\textit{ODN} Oligodeoxyribonucleotide; WT Ancestral allele; mut Minor allele; \(K_D\) Equilibrium dissociation constant, which is presented as mean ± standard deviation; \(\delta\), the difference between the affinity of TBP for an ODN with and without the SNP in its TATA box expressed as natural logarithms; \(\delta = −\ln [K_{D,Mut}] − (−\ln [K_{D,WT}])\); \(k_a\) and \(k_d\) association and dissociation rate constants, respectively; \(t_{1/2} = \ln2/k_d\) the average half-life of the TBP/ODN complexes; \textit{HBB} \&-\textbeta-globin gene; \textit{HBZ} Zeta-globin gene; and \textit{EPOR} Erythropoietin receptor gene.
complex formation was 19.5-fold lower, dissociation velocity 3-fold lower, whereas the lifetime of the complex was 3-fold longer.

The promoter of the erythropoietin receptor gene (EPOR) contains a TATA-like sequence, TATTTTTG, with the weightiest "T" at the third position [69] but with the flanks containing AT pairs. This arrangement results in weak TBP–TATA affinity: $K_D = 170 \pm 30$ nM. The -27'T > A substitution strengthens this affinity 15.5-fold. Of note, this enhancement of affinity is accompanied by only 3.3-fold greater velocity of formation of the complex and a 4.6-fold decrease in the velocity of its dissociation. The lifetime of the complex was more than 4-fold longer in comparison with the reference allele.

In case of rs567946217 (−31C > A), adjacent to a TATA-like sequence upstream, affinity was 8.1-fold stronger, while the velocity of complex formation was 2.4-fold greater, the velocity of dissociation 3.3-fold lower, and complex lifetime was >3-fold greater. The -29A > G substitution (rs971717705), barely affecting TBP–TATA affinity ($K_D$ strengthened by 13%), caused insignificant changes in the velocities of formation and dissociation of the complex (Table 3) and most likely can be categorized as a polymorphism causing "noise." Rs567946217 (−31C > G), replacing C (adjacent to the TATA box) with G, enhanced the affinity of the TBP–TATA binding by 2-fold, raising the velocity of formation 1.3-fold and decreasing the dissociation velocity 1.5-fold. The complex’s half-life also increased relative to the reference allele and was 38 min.

**Fig. 2** Measuring the kinetics of TBP binding to TATA-containing ODN N2 identical to the HBB promoter in healthy people. Legend: a) Dependences of reaction rates on ODN N2 concentrations; b) electropherograms from which these curves were derived. TBP concentration was 0.4 nM in all the experiments; the concentrations of a TATA-containing ODN that we used are indicated. TBP–TATA-associated isotherms and $k_a$ and $k_d$ values were inferred from the electropherograms using GraphPad Prism 5 software.
Significant correlations between our predictions in silico and measurements in vitro

As presented in Fig. 3, our predicted and experimental values significantly correlate with each other. As depicted in the graphs in the figure, for predicted \( K_D \) and its experimental values expressed in ln units (−ln \( [K_D] \); Fig. 3a, and predictions of measurements −ln \( [K_D] \), δ; Fig. 3b) when the SNP was present in TATA boxes, whose values are given in Table 3, the coefficient of linear correlation \( r \) was 0.87 at \( p < 0.0025 \) and \( r = 0.91 \) (\( p < 0.001 \)), respectively, whereas \( \delta \) was −ln \( [K_{D_{Mut}}] \) − (−ln \( [K_{D_{WT}}] \)). Finally, as readers can see in Fig. 3, two considered reliable linear correlations between our predictions in silico and measurements in vitro expressed in both absolute (a) and relative (b) scales were successfully verified using three independent correlation criteria, namely: the Goodman–Kruskal generalized correlation (γ), and both Spearman’s (R) and Kendall’s (τ) rank correlations.

A comparison between the in vitro measurements of this work with those of other authors

It should be noted that the authors of some studies have investigated the influence of a substitution of a first “T” with “C” in a TATA box sequence. For instance, in a work dealing with the TBP of the yeast *Saccharomyces cerevisiae* [70], it was found that this substitution has almost no influence on the binding. The same authors [71] have demonstrated that a promoter containing sequence CATAAAA is 40-fold less effective at inducing transcription in vivo than the classic sequence TATA AAA. Other researchers [72] have shown that transcription in HeLa cells decreases 3-fold after a substitution of the first “T” with “C” in the TATA box (TATA AA→CATAAA). These examples reveal that the substitution of T with C at the first position of the TATA box has different effects on transcription and therefore on the TBP–TATA interaction. One possible reason is differences in the sequences flanking the TATA box, because in some studies [73, 74], it has been reported that the affinity of TBP for oligonucleotides differing in the number of AT pairs in the sequences flanking the TATA box can differ 25- to 30-fold.

Our previously analyzed data on the relation of clinically confirmed diseases with substitutions in the TATA boxes of the promoter of β-globin genes indicate that the phenotype of intermediate thalassemia is associated with changes in TBP–TATA affinity causing an imbalance of chains, on average by 2.0- to 2.6-fold [26]. Consequently, carriers of the rs34500389 allele featuring 1.5-fold stronger TBP–TATA affinity may be susceptible to mild thalassemia because of the elevated amount of synthesized β-chains and can lead a normal lifestyle.

It is noteworthy that zeta-globin synthesis starts in the yolk sac in the third week of embryogenesis. By the fifth week, its level drops, and fetal ε-globin synthesis is initiated. Downregulation of *HbZ* after the substitution in the TATA box (rs113180943) can cause a stage-dependent disruption of “switching” events for genes of hemoglobins and the development of erythroblastosis [75]. In our study, experimental in vitro verification revealed that substitution -29T > A (rs113180943) indeed weakens TBP–TATA affinity 6.6-fold, and accordingly, downregulates the gene. Thus, rs113180943 may be a candidate SNP marker of severe hemoglobinopathies.
comorbid with cognitive and mental disorders associated with disturbances of cerebral blood flow [14, 76–80]. It should be noted that the experiment coincides well with the prognosis, which is due to the TATA box sequence close to consensus. And this is especially true for "T" in the third position of TATA box, which has the maximum weight among the nucleotides of the TATA box.

As a result of studying molecular and cellular programs regulating zeta-globin synthesis, some authors [81–83] have demonstrated its good potential as a new therapeutic agent for α-thalassemia and sickle cell anemia. They found that after therapeutic reactivation, a silent but structurally intact gene of embryonic zeta-globin starts expression, and zeta-globin chains are assembled with adult β-globin into a heterotetrameric complex, Hb z2β2, fully physiologically consistent with normal adult hemoglobin [84]. This approach to α-thalassemia treatment has been illustrated on a pathological phenotype of mice with homozygous embryonically lethal deactivation of α-globin genes [82]; transgenic human zeta-globin makes these mice fully viable. Carriers of the minor allele of rs113180943 may have hemoglobinopathy in addition to the higher risk of mental disorders (more often bipolar disorder) associated with hypoxia or hyperoxia of the central nervous system [8, 14].

There is evidence that glycoprotein hormone EPO is the main regulator of erythrocyte formation and has protective properties during cerebral ischemia [85]. To perform its function, EPO binds to its receptor, EPOR, which then dimerizes and activates, thereby inducing a cascade of genes responsible for the proliferation, survival, and differentiation of erythroid progenitor cells [86]. Although in the blood, EPOR regulates erythrocyte differentiation, in the brain, it protects several types of neurons from death, including A9 dopaminergic neurons of the substantia nigra [87], and promotes oxygen storage under hypoxia; this property of EPOR is especially important for neurons with heightened needs for energy [11]. Several studies suggest that brain ischemia upregulates EPO and EPOR for repair of the damage [88, 89]. Accordingly, it can be hypothesized that SNPs in the TATA-like sequence of the promoter of EPOR (-27 T > A, -31C > A, and -31C > G) enhancing the TBP–TATA affinity 15.5-fold, 8.1-fold, and 2.1-fold, respectively, in agreement with the expression levels of the gene, may indicate some degree of hereditary ischemic damage to the brain in the carriers of these alleles and may be candidate markers of this disorder. Substitutions -27 T > A and -31C > A most likely may serve as SNP markers of a rare disease, i.e., an erythropoiesis disorder [14] caused by an excessive number of erythrocytes.

The results of some experiments [77, 78] point to the participation of hemoglobins in mitochondrial respiration and metabolism. Therefore, both hypoxia and hyperoxia can damage brain tissues either by blocking ATP production or by increasing the formation of reactive oxygen species. Consequently, hemoglobins involved in mitochondrial metabolism and oxidative phosphorylation can have a major influence on neuron integrity and neurodegenerative events, which play a role in the pathogenesis of cognitive and mental disorders [14, 90]. On the other hand, SNPs -27 T > A and -31C > A, which cause a greater enhancement of TBP–TATA affinity and therefore upregulation of EPOR, may be candidate markers of increased risk of cognitive and mental disorders. This notion is supported by many studies indicating that both lower and higher levels of hemoglobin are related to an accelerated cognitive impairment [12, 13, 87–92].

**Conclusion**

It is known that genes of hemoglobins are expressed in the astrocytes of the cortex and hippocampus and in oligodendroglia that is located in almost all brain regions, including the striatum, corpus collosum, and medulla [11]. There are many reports about the correlation between aberrations in the metabolism of hemoglobins and symptoms of mental disorders [12, 88–93]. Many investigators believe that cognitive deficits are a cause not a consequence of erythropoiesis disturbances [12, 91–97].

Thus, from previous literature data as well as our computational and experimental results, it can be concluded that the identified candidate SNP markers of erythropoiesis disturbances may also be studied in cohorts of patients with cognitive and/or mental disorders with comorbid erythropoiesis diseases in comparison to conventionally healthy volunteers. Research on the regulatory mechanisms—by which the uncovered SNP markers contribute to the development of hemoglobinopathies and the comorbid cognitive deficits—will help physicians to not only take timely and adequate measures against hemoglobinopathies but also to implement strategies preventing or slowing the progression of cognitive and mental disorders. This new knowledge is an additional resource for biomedical research, personalized medicine, diagnostics, and the development of therapeutics.

Finally, because we have already verified our in silico predictions in question using ex vivo transfections of the pGL4.10 plasmid with LUC gene-reporter (Promega, USA) into both cancer cell cultures (e.g., HCT116) [35] and fibroblast ones (e.g., hTERT-BJ1) [98], a recent cell-reprogramming technologies based on the induced pluripotent stem cell lines aimed to simulate neurodegenerations (e.g., ICGi007-A [99]) would in future give a chance for more-or-less adequate testing ex vivo the candidate SNP markers of neuropathogenesis as soon as it could be free-available and well-studied.
Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12881-020-01106-x.

Additional file 1: Supplementary Electropherogram. The original, raw, unfiltered, uncropped, unedited electropherogram used for Fig. 2b in the cases of the minor allele T of the unannotated SNP rs34500389 of the human HBG gene promoter under this study.

Abbreviations

EMSA: Electrophoretic mobility shift assay; $k_0$ and $k_d$: Association and dissociation rate constants, respectively; $K_d = k_d / k_0$: Equilibrium dissociation constant; In: Natural logarithm; mut: Minor allele; ODN: Oligodeoxynucleotide; SNP: Single-nucleotide polymorphism; TBP: TATA-binding protein; TBP-binding site; $r = k_0 / k_d$: The average half-life of the TBP/ODN complexes; WT: Wild type (norm); $\Delta$: The difference between the affinity of TBP for an ODN with and without the SNP in its TATA box expressed as natural logarithms; ALAS1: Aminolevulinate synthase 1; CA1: Carbonic anhydrase 1; CA1: Carbonic anhydrase promoter 2; EDRFI: Equilibrium dissociation regulatory factor 1; EPOR: Erythropoietin receptor; GYPB: Glycoprotein B; GYPC: Glycoprotein C; HBA1: $alpha$-globin; HBB: $beta$-globin; HBB: $beta$-globin promoter 2; HBG1: $gamma$-globin; HBG1: $gamma$-globin promoter 2; Homeobox B2.

Acknowledgments

We are grateful to Shevchuk Editing (Brooklyn, NY, USA:http://www.shevchuk-editing.com) for English translation and editing.

About this supplement

This article has been published as part of BMC Medical Genetics Volume 21 Supplement 1: 2020: Selected Topics in “Systems Biology and Bioinformatics” – 2019: medical genetics. The full contents of the supplement are available online at https://bmcmedgenet.biomedcentral.com/articles/supplements/volume-21-supplement-1.

Authors’ contributions

NK and LS conceived of and supervised the study. ID, ES, and IC conducted the in vitro experiments. PP, OP, and MP designed, developed, maintained, adapted, and tuned the software for sequence analysis. OA analyzed the data in silico. MP and LS interpreted the computer-based predictions in comparison with experimental data. LS wrote the manuscript. All the coauthors read and approved the final version of the manuscript.

Funding

Manuscript writing was supported by project #0324–2019-0042-C-01 from the Russian Government Budget (for LS, OA, and ID). The data compilation was supported by project #0324–2019-0040-C-01 from the Russian Government Budget (for IC, PP, and OP). The in vitro experiment was supported by the project “Investigation, analysis and complex independent expertise of projects of the National technological initiatives, including the accompanying projects of the road map “NeuroNet,” which is executed within the framework of state assignment #28.12487.2018/12.1 (for ES) and “S–100" (for MP) from the Ministry of Science and Higher Education of the Russian Federation.

Publication costs have been funded by the Russian Federal Science & Technology Program for the Development of Genetic Technologies. All the authors declare that these funding bodies did not play roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

Web service SNP_TACETA_Z-tester is publicly available (http://www.mgs.bionet.nsc.ru/cgi-bin/mgs/tatacan_zx/start.pl).

Due to two publicly available Web services UCSC Genome Browser (URL = http://genome.ucsc.edu/) [38] and NCBI dbSNP database build No.151 (URL = http://www.ncbi.nlm.nih.gov/snp/) [39]; we predicted SNPs, which can reliably cause over- and underexpression of the 13 human erythroid genes HBB, HBZ, HBA1, HBG1, HBME1, ALAS1, CA1, HOXB2, EPOR, GYPB, GYPC, EDRFI and IDs of which are listed, as follows: rs1006576690, rs103503390, rs1036491581, rs1046254329, rs1131180873, rs150246890, rs281864525, rs33913746, rs33980857, rs33991098, rs34500389, rs345858299, rs35518301, rs3967038, rs38896304, rs540950375, rs557418569, rs5979927846, rs564934089, rs567946217, rs571522665, rs573241527, rs637550400, rs762972655, rs768296486, rs74801749, rs856503911, rs905305347, rs905730148, rs911469201, rs91712507, rs942640545, rs970970552, rs971717705, rs981975270, and rs991064314.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, 10 Lavrentyev Ave, Novosibirsk 630090, Russia. 2Novosibirsk State University, 1 Pirogova Street, Novosibirsk 630090, Russia. 3Institute of Gene Biology Russian Academy of Sciences, 34/5 Vavilova Street, Moscow 119334, Russia.

Published: 22 October 2020

References

1. 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;381:142–51.
2. Galanello R, Origa R. Beta-thalassemia. Orphanet J Rare Dis. 2010;5:11.
3. Traeger-Synodinos J, Vrettou C, Kanavakis E. Rapid detection of fetal Mendelian disorders: thalassemia and sickle cell syndromes. Methods Mol Biol. 2008;444:133–45.
4. Rivella S. Ineffective erythropoiesis and thalassemias. Curr Opin Hematol. 2009;16:87–94.
5. Vichinsky EP. Clinical manifestations of-/-thalassemia. Cold Spring Harb Perspect Med. 2013;3:a011742.
6. Wu H, Wang H, Lan L, Zeng M, Guo W, Zheng Z, et al. Invasive molecular prenatal diagnosis of alpha and beta thalassemia among Hakka pregnant women. Medicine (Baltimore). 2018;97:e13557.
7. O’Shea LC, Fair T, Hensley C. Linked alpha-thalassemia with mental retardation is downstream of protein kinase A in the meiotic cell cycle signaling cascade in Xenopus oocytes and is dynamically regulated in response to DNA damage. Biol Reprod. 2019;102:1238–49.
8. Cavazzana M, Antoniani C, Miccio A. Gene therapy for beta-hemoglobinopathies. Mol Ther. 2017;25:1142–54.
9. Schelhorn DW, Schneider A, Kuschinsky W, Weber D, Kruger C, Dittgen T, et al. Expression of hemoglobin in rodent neurons. J Cereb Blood Flow Metab. 2009;29:585–95.
10. Ohyagi Y, Yamada T, Goto I. Hemoglobin as a novel protein developmentally regulated in neurons. Brain Res. 1994;635:323–7.
11. Biagioli M, Pinto M, Ceselli D, Zaninello M, Lazarovic D, Roncalia P, et al. Unexpected expression of alpha- and beta-globin in mesencephalic dopaminergic neurons and glial cells. Proc Natl Acad Sci U S A. 2009;106:15454–9.
12. Faux NG, Rembach A, Wiley J, Ellis KA, Ames D, Fowler CJ, et al. Anemia of Alzheimer's disease. Mol Psychiatry. 2014;19:1227–34.
13. Shen SH, Bu XL, Jin WS, Shen LL, Wang J, Zhuang ZQ, et al. Expression of hemoglobin in rodent neurons. J Cereb Blood Flow Metab. 2009;29:585–95.
14. Altnoz MA, Ince B. Hemoglobinins emerging roles in mental disorders. Metabolical, genetical and immunological aspects. Int J Dev Neurosci. 2017;51:63–78.
15. Collins PY, Joestl SS, March D, Inse TR, Collins PY, et al. Grand challenges in global mental health. Nature. 2011;475:27–30.
16. Cassidy F, Zhao C, Badger J, Claffey E, Dobrin S, Roche S. Genome-wide scan of bipolar disorder and investigation of population stratification effects on linkage: support for susceptibility loci at 4q21, 7q36, 9p21, 12q24, 14q24, and 16p13. Am J Med Genet B Neuropsychiatr Genet. 2007;144B:791–801.
17. Kramer AH, Zygun DA. Anemia and red blood cell transfusion in neurocritical care. Crit Care. 2009;13:89.

18. Ihle-Hansen H, Thommesen B, Wyller T, Engedal K, Oksenberg AR, Engedal K, et al. Incidence and subtypes of MCI and dementia 1 year after first-ever stroke in patients without pre-existing cognitive impairment. Dement Geriatr Cogn Disord 2011; 32: 401–407.

19. Troncoso JC, Zonderman AB, Resnick SM, Crain B, Breiter MM. Effect of infarcts on dementia in the Baltimore longitudinal study of aging. Ann Neurol. 2008;64:176–88.

20. Vermeer SE, Prins ND, den Heijer T, Hofman A, Koudstaal PJ, Breteler MM. Effect of infarcts on dementia in the Baltimore longitudinal study of aging. Aging Cell. 2007;6:361–70.

21. Shah RC, Wilson RS, Tang Y, Dong X, Murray A, Bennett DA. Relation of oxidative stress and aberrant signaling in aging and cognitive decline. J Nutr. 2020, 150(9):1515–21.

22. Droge W, Schipper HM. Oxidative stress and aberrant signaling in aging and cognitive decline. Aging Cell. 2007;6:361–70.

23. Deadmond MA, Smith-Gagen JA. Changing incidence of myeloproliferative neoplasms: trends and subgroup risk profiles in the USA, 1973–2011. J Cancer Res Clin Oncol. 2015;141:231–8.

24. Savinova LV, Ponomarenko MP, Ponomarenko PM, Drachkova IA, Lysyova MV, Aristova TV, et al. TATA box polymorphisms in human gene promoters and associated hereditary pathologies. Biochemistry (Moscow). 2009;74:117–29.

25. Ponomarenko PM, Savinova LV, Drachkova IA, Lysyova MV, Aristova TV, Ponomarenko MP, et al. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. Dokl Biochem Biophys. 2008;419:88–92.

26. Savinova L, Drachkova I, Aristova T, Ponomarenko P, Ponomarenko M, Kolchavon N. An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein. PLoS One. 2013;8:e54626.

27. Drachkova I, Savinova L, Aristova T, Ponomarenko M, Petlekh S, Kolchavon N. The mechanism by which TATA-box polymorphisms associated with human hereditary diseases influence interactions with the TATA-binding protein. Hum Mutat. 2014;35:601–8.

28. Colonna V, Ayub Q, Chen Y, Pagoni L, Luisi P, Pybus M, et al. Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. Genome Biol. 2014;15:R88.

29. Teltent A, Pierce LC, Biggs WH, di Iulio J, Wong EH, Fabani MM, et al. Deep sequencing of 10,000 human genomes. Proc Natl Acad Sci USA. 2016;113:11901–6.

30. Wu J, Wu M, Li L, Liu Z, Zeng W, Jiang R. dbWGFP: a database and web tool for human gene promoters. BMC Genomics. 2016;17:995.

31. Chen Y, Wu M, Li L, Liu Z, Zeng W, Jiang R. dbWGFP: a database and web tool for human gene promoters. BMC Genomics. 2016;17:995.

32. Ponomarenko PM, Rasskazov D, Suslov V, Sharypova E, Savinova L, Podkolodnaya O, et al. Candidate SNP markers of charcotopathologies are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. Biomed Res Int. 2016;2016:842703.

33. Haessler M, Raney B, Hinichi AS, Clavson H, Zwaag AS, Karolchik D, et al. Navigating protected genomics data with UCSC genome browser in a box. Bioinformatics. 2015;31:764–6.

34. Sherry ST, Ward MH, Kholodov M, Blegen WM, Jia L, Babbitt CP, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.

35. Drachkova IA, Repkova MN, Lysova MV, Aristova TV, Savinova LK. Interaction of proteins from general transcription complex RNA polymerase II with oligoribonucleotides. Mol Biol (Mosk). 2005;39:133–46.

36. Peterson MG, Tanese N, Pugh BF, Tjian R. Functional domains and upstream activation properties of cloned human TATA binding protein. Science. 1990; 248:1625–30.

37. Pugh BF. Purification of the human TATA-binding protein, TBP. Methods Mol Biol. 1995;37:359–67.

38. Waardenberg AJ, Barset SD, Bouweret R, Harvey RP. CompGO: an R package for comparing and visualizing gene ontology enrichment differences between DNA binding experiments. BMC Bioinformatics. 2015;16:275.

39. Leung PP, Bungert J. The human beta-globin locus control region. Eur J Biochem. 2002;269:1589–99.

40. Higgs DR, Engel JD, Stamatinopoulos G. Thalassaemia. Lancet. 2012;379:373–83.

41. Thein SL. Genetic association studies in β-hemoglobinopathies. Hematology Am Soc Hematol Educ Program. 2011;2013:534–61.

42. Martinez JA, Cerami A, Slater AF. Inhibition of hemozoin formation in plasmodium falciparum falciparum trophozoite extracte by heme analogs: possible implication in the resistance to malaria conferred by the beta-thalassemia trait. J Biol Chem. 1996;271:2346–46.

43. Eng B, Walker L, Nakamura LM, Hoppie C, Azimi M, Lee H, Wayne JS. Three new beta-globin gene promoter mutations identified through newborn screening. Hemoglobin. 2007;31:129–34.

44. He Z, Russell JE. Expression, purification, and characterization of human hemoglobins Gower-1 (zeta (2) epsilon (2)), Gower-2 (alpha (2) epsilon (2)), and Portland-2 (zeta (2) beta (2)) assembled in complex transgenic–knockout mice. Blood. 2001;97:1099–105.

45. Sabath DE, Detter JC, Tait JF. A novel deletion of the entire a globin locus causing a -thalassemia-1 in a northern European family. Am J Clin Pathol. 1994;102:650–6.

46. Ko TM, Tseng LH, Kao CH, Lin YW, Hwa HL, Hsu PM, et al. Molecular characterization and PCR diagnosis of Thailand deletion of alpha-globin gene cluster. Am J Hematol. 1998;57:124–30.

47. Langlois KD, Oner C, Dimovski AJ, Gu YC, Huisman TH. Sequence variations in the 5' flanking and IVS-II regions of the G-gamma- and a-gamma-globin genes of beta S chromosomes with five different haplotypes. Blood. 1991;77:4886–96.

48. Winter S, Kinney TR, Ware RE. Gallbladder sludge in children with sickle cell disease. J Pediatr. 1994;125:747–9.

49. Guberman AS, Scassa ME, Canepe ET. Repression of 5-aminolevulinate synthase gene by the potent tumor promoter, TPA, involves multiple signal transduction pathways. Arch Biochem Biophys. 2005;436:285–96.

50. Thunell S, Porphyria. porphyries metabolism and porphyrias. I. Update. Scand J Clin Lab Invest. 2000;60:509–54.

51. Sassa S, Nagai T. The role of heme in gene expression. Int J Hematol. 1996; 63:167–78.

52. Cox EH, McDonald GL, Morel FMM, Lane TW, Prince RC, Pickering U, et al. The active site structure of Thalassosira weissflogii carbonic anhydrase. Biochemistry. 2000;39:12128–30.

53. Vulic R, Tycicakova S, Dubrovakova M, Skutevly L, Lakota J. Silencing of CA1 mRNA in tumour cells does not change the gene expression of the extracellular matrix proteins. J Cell Mol Med. 2018;22:695–9.

54. Lokal N, Akocak S, Rajesh S, Sanku R, Supuran C. Discovery of new ureido benzene sulfoxonamides incorporating 1,3,5-triazine moieties as carbonic anhydrase I, II, IX and XII inhibitors. Bioorg Med Chem. 2019;15(27):1588–94.

55. Guberman AS, Scassa ME, Canepe ET. Repression of 5-aminolevulinate synthase gene by the potent tumor promoter, TPA, involves multiple signal transduction pathways. Arch Biochem Biophys. 2005;436:285–96.

56. Thunell S. Porphyria. porphyiris metabolism and porphyrias. I. Update. Scand J Clin Lab Invest. 2000;60:509–54.

57. Sassa S, Nagai T. The role of heme in gene expression. Int J Hematol. 1996; 63:167–78.

58. Cox EH, McDonald GL, Morel FMM, Lane TW, Prince RC, Pickering U, et al. The active site structure of Thalassosira weissflogii carbonic anhydrase. Biochemistry. 2000;39:12128–30.

59. Vulic R, Tycicakova S, Dubrovakova M, Skutevly L, Lakota J. Silencing of CA1 mRNA in tumour cells does not change the gene expression of the extracellular matrix proteins. J Cell Mol Med. 2018;22:695–9.
62. Kouny MJ, Bondurant MC. Erythropoietin retards DNA breakdown and prevents programmed death in erythroid progenitor cells. Science. 1990; 248:378–81.
63. Bento C. Genetic basis of congenital erythrocytosis. Int J Lab Hematol. 2018; 40:62–7.
64. Vocanec D, Prijateli T, Debeljak N, Kurej T. Genetic variants of erythropoietin (EPO) and EPO receptor genes in familial erythrocytosis. Int J Lab Hematol. 2019;41:162–7.
65. Dame C, Juul SE, Christensen RD. The biology of erythropoietin in the central nervous system and its neurotrophic and neuroprotective potential. Biol Neonate. 2001;79:228–35.
66. Marti HH, Wengher RH, Rivas LA, Straumann U, Dicigaylioglu M, Henn V, et al. Erythropoietin gene expression in human, monkey and murine brain. Eur J Neurosci. 1996;8:666–76.
67. Wilder JA, Hewett EK, Gansner ME. Molecular evolution of GYPC: evidence for recent structural innovation and positive selection in humans. Mol Biol Evol. 2009;26:2679–87.
68. Ruksa B, Stanova E, Hadjidekova S, Stamenov G, Milanova 5th, Toncheva D. Rukova B. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol. 1990;212:563–78.
69. Stewart JJ, Stargell LA. The stability of the TFIIA-TBP-DNA complex dependent on the sequence of the TATAAA element. Biol Chem. 2001;276:3070–8.
70. Wobbe CR, Struhl K. Yeast and human TATA-binding proteins have nearly identical DNA binding properties with nearly identical DNA sequence requirements for transcription in vitro. Mol Cell Biol. 1990;10:3859–67.
71. Fager H, Ivanchenko M, Cohen I, Haran TE. TBP ranking sequences: asymmetry of binding, long-range effects and consensus sequences. Nucleic Acids Res. 2006;34:104–19.
72. Savinova KV, Drachkova IA, Ponmarenko MP, Lysova MV, Arshavina T, Kolchanov NA. Interaction between the recombinant TATA-binding protein and the TATA-boxes of the mammalian gene promoters. Ecological Genetic. 2007;4:44–9.
73. Chen RL, Chou YC, Lan YJ, Huang TS, Shen ON. Developmental silencing of human zeta-globin gene expression is mediated by the transcriptional repressor RREB1. J Biol Chem. 2010;285:10189–97.
74. Murayama H, Shinkai S, Nishi M, Taniguchi Y, Amano H, Seino S, et al. Erythropoietin gene expression in human, monkey and murine brain. Eur J Neurosci. 1996;8:666–76.
75. Maziade M, Roy MA, Chagnon YC, Cliche D, Fournier JP, Montgrain N, et al. Intranasal erythropoietin protects ca1 hippocampal cells, modulated by specific time pattern molecular changes after ischemic damage in rats. J Mol Neurosci. 2019;68:590–602.
76. Winter SS, Howard T, Ware RE. Regulation of expression of the human erythropoietin receptor gene. Blood Cells Mol Dis. 1996;22:214–24.