


RESEARCH

Open Access



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

Mikhail Ponomarenko^{1,2*} , 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 γ - and ζ -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)

* Correspondence: pon@bionet.nsc.ru

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

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Results: On the basis of these experimental in vitro results and literature data, we studied TATA box SNPs influencing both human erythropoiesis and cognitive abilities. For instance, TBP–TATA affinity in the *HbZ* promoter decreases 6.6-fold as a result of a substitution in the TATA box (rs113180943), thereby possibly disrupting stage-dependent events of “switching” of hemoglobin genes and thus causing erythroblastosis. Therefore, rs113180943 may be a candidate marker of severe hemoglobinopathies with comorbid cognitive and mental disorders associated with cerebral blood flow disturbances.

Conclusions: The literature data and experimental and computations results suggest that the uncovered candidate SNP markers of erythropoiesis anomalies 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 into the regulatory mechanisms by which the identified SNP markers contribute to the development of hemoglobinopathies and of the associated cognitive deficits will allow physicians not only to take timely and adequate measures against hemoglobinopathies but also to implement strategies preventing cognitive and mental disorders.

Keywords: Erythropoiesis, Cognitive disorder, In silico prediction, TBP/TATA affinity, kinetics of TBP-TATA binding in vitro, TATA-binding protein, SNP marker

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 $A\gamma$ - and $G\gamma$ -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 poor or ineffective iron assimilation [21]. Elevated hemoglobin levels can cause blood hyperviscosity, hypovolemia, 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 8.6×10^9 possible SNPs have been deposited in the dbWGF 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 three-step biophysical 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], chromopathologies [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 (*ALAS1*, *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 8.58×10^9 possible SNPs throughout the whole human genome, predictions, experimental data, and the relevant clinical findings accumulated in database dbWGF 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://www.mgs.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 T4 polynucleotide 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 PharoFX Plus phosphorimager (Bio-Rad) [40]. 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 $\mu\text{g}/\text{ml}$ 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 PharoFX 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 (δ) with an SNP in the TATA box were calculated as $-\ln [K_{D,TATA_{Mut}}] - (-\ln [K_{D,TATA}])$.

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 (5'- ϵ - γ G- γ A- 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; 5'-3'	Prognosis, $-\ln K_D$
<i>HBB</i>	WT	cagggctgggcataaaaactcagggcagag	19.20
	rs63750400:G	cagggctgggcataaaaGtcagggcagag	19.12
	rs281864525:T	cagggctgggcataaaaTctcagggcagag	19.169
	rs34598529:G	cagggctgggcatGaaactcagggcagag	17.849
	rs33980857:A	cagggctgggcaAaaaactcagggcagag	17.701
	rs33980857:C	cagggctgggcaCaaaactcagggcagag	18.172
	rs33980857:G	cagggctgggcaGaaaactcagggcagag	17.673
	rs33981098:C	cagggctgggcCtaaaactcagggcagag	18.553
	rs33981098:A	cagggctgggcGtaaaactcagggcagag	18.062
	rs34500389:T	cagggctgggTataaaactcagggcagag	20.175
	rs34500389:A	cagggctgggActaaaactcagggcagag	18.635
	rs33931746:C	cagggctgggcataCaactcagggcagag	18.341
	<i>HBB_2</i>	WT	ggaccagcacaagggcagggcagag
rs35518301:G		ggaccagcacGaaagggcagggcagag	18.65
<i>HBZ</i>	WT	cagctccctgtaataaggggaccctg	20,76
	rs113180943:T	cagctccctgtaTtaaggggaccctg	19,28
	rs1035033590:C	cagctccctgtaataagggCgaccctg	20,773
<i>HBZ_3</i>	WT	ggaggggtggggccctatctctcct	17.611
	rs559282746:G	ggaggggtggggGcctatctctcct	17.605
<i>HBA1</i>	WT	gcgccccaagcataaacctggcgcg	18.36
	rs571582665:T	gcgccccaagcataaacTtggcgcg	18.62
<i>HBG1</i>	WT	agggtgcttccttttattcttcatcc	19.06
	rs573241527:G	agggtgcttccttttCttcttcatcc	17.54
<i>HBM</i>	WT	ggtgcccgaggctctataaggaggc	19.13
	rs905730148:A	ggtgcccgaggAccctataaggaggc	19.23
	rs937192507:C	ggtgcccgaggcccCataaggaggc	19.95
<i>HBE1</i>	WT	aggcaaaaagagagcttgtgtagagc	18.31
	rs768296486:A	aggcaaaaagagagAttgtgtagagc	18.16
	rs1036491581:G	aggcaaaaagagagaGtgtgtagagc	18.01
	rs150246840:C	aggcaaaaagagaggttgtgtagCgc	18.16

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*, α 1-globin; *HBG1* A γ globin; *HBM* Hemoglobin subunit μ ; and *HBE1* ϵ -globin

and *HBD*, respectively), six of many regulatory SNP markers (rs33981098, rs33980857, rs34598529, rs33931746, 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'-aminolevulinic 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'-aminolevulinic; the latter is a universal precursor of tetrapyrrole compounds participating 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; 5'-3'	Prognosis, $-\ln K_D$
<i>ALAS1</i>	WT	ctcccgtgtatattaaggcgcggc	20.11
	rs564394089:A	ctacAgctgtatattaaggcgcggc	20.23
	rs1046254329:A	ctaccgctAtatattaaggcgcggc	20.40
	rs905035347:T	ctaccgctatatattaaggTgcccggc	20.11
<i>CA1</i>	WT	ggaatgggcagcttatgtacagggg	19.81
	rs540950375:T	ggaatgggcaTcttatgtacagggg	19.70
	rs991064314:T	ggaatgggTagcttatgtacagggg	19.82
<i>CA1_2</i>	WT	cttgggcatttttatagaaacttact	20.22
	rs557418569:G	cttgggcGtttttatagaaacttact	20.16
	rs538698304:C	cttgggcattttCatagaaacttact	19.61
<i>HOXB2</i>	WT	aggactccagcgaattacagggaaat	17.98
	rs762972656:T	aggactccaTcgaattacagggaaat	17.92
<i>EPOR</i>	WT	cgtagcagacaaaaatagatgacgtg	18.24
	rs1006576690:T	cgtagcagacaaaTatagatgacgtg	19.39
	rs971717705:C	cgtagcagacaaaaCagatgacgtg	18.11
	rs567946217:A	cgtagcagacaaaaaAatgacgtg	19.33
	rs567946217:C	cgtagcagacaaaaaCatgacgtg	18.51
<i>GYPB</i>	WT	gcctactagctgttatcttccaggcc	18.26
	rs3967038:C	gcctacCagctgttatcttccaggcc	18.30
	rs3967038:A	gcctacAagctgttatcttccaggcc	18.23
<i>GYPC</i>	WT	cattggggagttttccctgcactcct	16.97
	rs970970552:A	cattAgggagttttccctgcactcct	17.83
	rs911469201:A	cattggggagttttccctAcactcct	17.13
	rs989175270:G	cattggggagttttccctgGactcct	17.27
<i>EDRF1</i>	WT	tcgcgagatttaatggcgcgagtcacag	18.45
	rs946240545:G	tcgcgagGtttaatggcgcgagtcacag	18.34
	rs774081749:G	tcgcgagattGaatggcgcgagtcacag	17.76
	rs895690311:G	tcgcgagatttaatggcgcGgtcacag	18.18

Notes: See “Notes” under Table 1. Genes: *ALAS1* Aminolevulinic 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

various reactions, including heme biosynthesis, single-electron transfer, and catalysis of redox reactions [56, 57]. It is reported that there is substantial tissue-specific expression and heme-mediated (as a 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 anhydrase 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 uses a homodimeric receptor (EPOR), which is also widely expressed in the nervous system [61]. The main role of EPOR is to stimulate the rapid spread of erythrocyte progenitors and to promote their survival [62]. Mutations of *EPOR* cause primary hereditary erythrocytosis [63] and have been detected in 15% of the cases [64]. It is well proven that the 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

methylation profiles of *GYPC*, 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_TATA_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 *HBB*, *HBZ*, and *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 *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 *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. Additionally, Table 3 shows negative natural logarithms of the predicted and experimentally determined K_D values and changes in K_D (δ) 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 *HBB*) changing the TATA box to a more canonical form (C₋₃₂ATAAAAAG → T₋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; $K_D = 50$ 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 weaken the TBP–TATA binding.

The unannotated SNP (rs113180943, detected by us in the TATA box of the *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 ($K_D = 6$ nM for the ancestral allele and $K_D = 40$ nM for –29T > A). The reason is that the velocity of

Table 3 Kinetic and thermodynamic characteristics of TBP–TATA interactions

Gene	ODN	Alleles: WT or mut	Allele	Sequences (5' – 3' strands), 26 bp	Prediction		Experiments					
					–lnK _D	δ	–lnK _D	δ	K _D , nM	k _a , M ⁻¹ s ⁻¹	k _d , s ⁻¹	t _{1/2} , min
<i>HbB</i>	1	WT	–32c > T	agggctgggcataaaagtcaggggcag	19.20	+	16.81	+	50 ± 7	(1.4 ± 0.7)*10 ⁴	(7.1 ± 0.7)*10 ⁻⁴	16 ± 2
	2	rs34500389:T		agggctgggTataaaagtcaggggcag	20.18	0.98	17.26	0.45	32 ± 3	2.1 ± 0.1*10 ⁴	(6.6 ± 0.5)*10 ⁻⁴	17 ± 2
<i>HbZ</i>	3	WT	–29t > A	agctccctgtatataaggggaccctg	20.76	–1.48	18.93	–1.90	6 ± 1	(4.5 ± 0.4)*10 ⁴	3.0 ± 0.6*10 ⁻⁴	38 ± 4
	4	rs113180943: A		agctccctgtaAataaggggaccctg	19.28		17.03		40 ± 20	(2.3 ± 0.2)*10 ³	(1.0 ± 0.4)*10 ⁻⁴	115 ± 15
	5	WT		cacgtcatctatttttgtctgctacg	18.24		15.59		170 ± 30	(2.8 ± 0.3)*10 ³	(4.6 ± 0.6)*10 ⁻⁴	25 ± 3
<i>EPOR</i>	6	rs1006576690: A	–27 t > A	cacgtcatctatAttgtctgctacg	19.39	+	18.33	+	11 ± 3	(9.3 ± 0.7)*10 ³	(1.0 ± 0.3)*10 ⁻⁴	115 ± 15
	7	rs971717705: G	–29a > G	cacgtcatctGtttttgtctgctacg	18.11	–0.13	15.71	+	150 ± 30	(2.4 ± 0.3)*10 ³	(3.6 ± 0.5)*10 ⁻⁴	32 ± 4
	8	rs567946217: A	–31c > A	cacgtacatAttttttgtctgctacg	19.33	+	17.68	+	21 ± 6	(6.8 ± 0.7)*10 ³	(1.4 ± 0.4)*10 ⁻⁴	82 ± 6
	9	rs567946217: G	–31c > G	cacgtcatGtatttttgtctgctacg	18.51	+	16.34	+	80 ± 10	(3.6 ± 0.3)*10 ³	(3.0 ± 0.5)*10 ⁻⁴	38 ± 4

ODN Oligodeoxyribonucleotide; WT Ancestral allele; mut Minor allele; K_D Equilibrium dissociation constant, which is presented as mean ± standard deviation; δ, 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; *HBB* β-globin gene; *HBZ* Zeta-globin gene; and *EPOR* Erythropoietin receptor gene

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 δ 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 TATAAAA. 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 (TATAAA→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

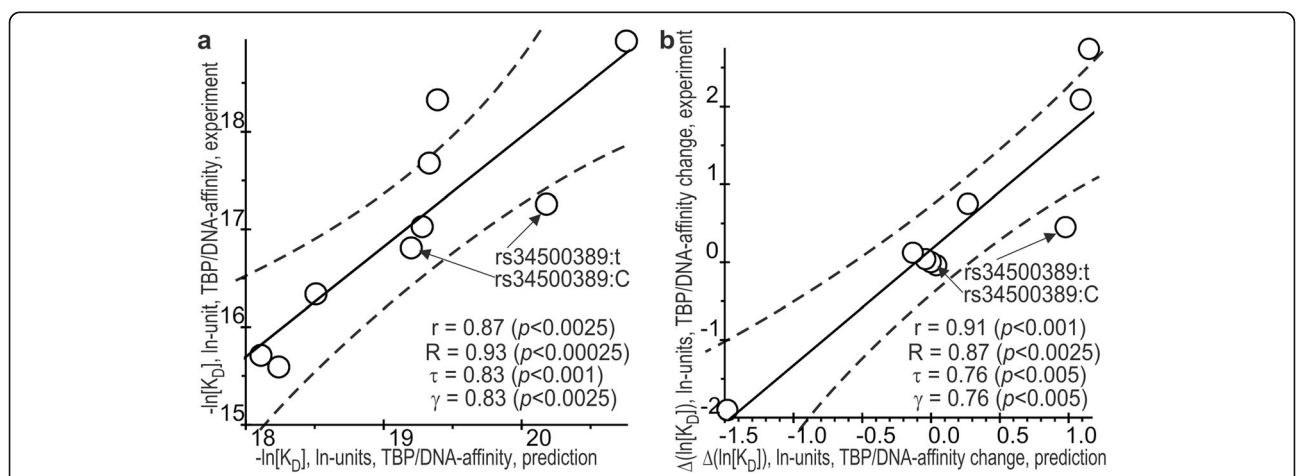


Fig. 3 The significant correlations between the in silico predicted and in vitro experimentally measured values. *Legend:* **a** TBP–DNA affinity and **b** the TBP–DNA affinity change caused by the minor allele of the analyzed SNPs with respect to the norm, $-\ln [K_D]$ and $\Delta \ln [K_D]$, respectively. Solid and dashed lines denote the linear regression and boundaries of its 95% confidence interval, calculated by means of software package STATISTICA (Statsoft™, USA); arrows pinpoint the ancestral rs34500389:C and minor rs34500389:t alleles of the SNP being studied, an analysis of which is depicted in Fig. 1 as an example of the application of our Web service SNP_TATA_Z-tester [31] in this work; r , τ , γ , and p are coefficients of Pearson’s linear correlation, Spearman’s rank correlation, Kendall’s rank correlation, and Goodman–Kruskal generalized correlation and their p values, respectively

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 z2 β 2, 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 *HBB* gene promoter under this study.

Abbreviations

EMSA: Electrophoretic mobility shift assay; k_a and k_d : Association and dissociation rate constants, respectively; $K_D = k_d / k_a$: Equilibrium dissociation constant; ln: Natural logarithm; mut: Minor allele; ODN: Oligodeoxyribonucleotide; SNP: Single-nucleotide polymorphism; TBP: TATA-binding protein; TBP-site: TBP-binding site; $t_{1/2} = \ln 2 / k_d$: The average half-life of the TBP/ODN complexes; WT: Wild type (norm); δ : The difference between the affinity of TBP for an ODN with and without the SNP in its TATA box expressed as natural logarithms; *ALAS1*: Aminolevulinatase synthase 1; *CA1*: Carbonic anhydrase 1; *CA1_2*: Carbonic anhydrase promoter 2; *EDRF1*: Erythroid differentiation regulatory factor 1; *EPOR*: Erythropoietin receptor; *GYPB*: Glycophorin B; *GYPC*: Glycophorin C; *HBA1*: α 1-globin; *HBB*: β -globin; *HBB_2*: β -globin promoter 2; *HBG1*: γ globin; *HBE1*: ϵ -globin; *HBM*: Hemoglobin μ ; *HBZ*: Zeta-globin; *HBZ_3*: Zeta-globin promoter 3; *HOXB2*: 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://bmcmcdgenet.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 "5–100" (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_TATA_Z-tester is publicly available (http://www.mgs.bionet.nsc.ru/cgi-bin/mgs/tatascan_fox/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, HBM, HBE1, ALAS1, CA1, HOXB2, EPOR, GYPB, GYPC, EDRF1 and IDs of which are listed, as follows: rs1006576690, rs1035033590,

rs1036491581, rs1046254329, rs113180943, rs150246840, rs281864525, rs33931746, rs33980857, rs33981098, rs34500389, rs34598529, rs35518301, rs3967038, rs538698304, rs540950375, rs557418569, rs559282746, rs564394089, rs567946217, rs571582665, rs573241527, rs63750400, rs762972656, rs768296486, rs774081749, rs895690311, rs905035347, rs905730148, rs911469201, rs937192507, rs946240545, rs970970552, rs971717705, rs989175270, 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

- 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.
- Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis*. 2010;5:11.
- 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.
- Rivella S. Ineffective erythropoiesis and thalassemias. *Curr Opin Hematol*. 2009;16:187–94.
- Vichinsky EP. Clinical manifestations of α -thalassemia. *Cold Spring Harb Perspect Med*. 2013;3:a011742.
- 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.
- O'Shea LC, Fair T, Hensy C. X-linked α -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;100:1238–49.
- Cavazzana M, Antoniani C, Miccio A. Gene therapy for β -hemoglobinopathies. *Mol Ther*. 2017;25:1142–54.
- Schelshorn 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.
- Ohyagi Y, Yamada T, Goto I. Hemoglobin as a novel protein developmentally regulated in neurons. *Brain Res*. 1994;635:323–7.
- Biagioli M, Pinto M, Cesselli D, Zaninello M, Lazarevic D, Roncaglia 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.
- Faux NG, Rembach A, Wiley J, Ellis KA, Ames D, Fowler CJ, et al. An anemia of Alzheimer's disease. *Mol Psychiatry*. 2014;19:1227–34.
- Chen SH, Bu XL, Jin WS, Shen LL, Wang J, Zhuang ZQ, et al. Altered peripheral profile of blood cells in Alzheimer disease: A hospital-based case-control study. *Medicine (Baltimore)*. 2017;96:e6843.
- Altinöz MA, Ince B. Hemoglobins emerging roles in mental disorders. Metabolic, genetical and immunological aspects. *Int J Dev Neurosci*. 2017; 61:73–85.
- Collins PY, Pate V, Joestl SS, March D, Inse TR, Collins PY, et al. Grand challenges in global mental health. *Nature*. 2011;475:27–30.
- 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:R89.
18. Ihle-Hansen H, Thommessen B, Wyller T, Engedal K, Oksengard 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, Pletnikova O, O'Brien RJ. Effect of infarcts on dementia in the Baltimore longitudinal study of aging. *Ann Neurol*. 2008;64:168–76.
20. Vermeer SE, Prins ND, den Heijer T, Hofman A, Koudstaal PJ, Breteler MM. Silent brain infarcts and the risk of dementia and cognitive decline. *N Engl J Med*. 2003;348:1215–22.
21. Droge W, Schipper HM. Oxidative stress and aberrant signaling in aging and cognitive decline. *Aging Cell*. 2007;6:361–70.
22. Shah RC, Wilson RS, Tang Y, Dong X, Murray A, Bennett DA. Relation of hemoglobin to level of cognitive function in older persons. *Neuroepidemiology*. 2009;32:40–6.
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;141:2131–8.
24. Savinkova LK, Ponomarenko MP, Ponomarenko PM, Drachkova IA, Lysova MV, Arshinova TV, et al. TATA box polymorphisms in human gene promoters and associated hereditary pathologies. *Biochemistry (Mosc)*. 2009;74:117–29.
25. Ponomarenko PM, Savinkova LK, Drachkova IA, Lysova MV, Arshinova 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. Savinkova L, Drachkova I, Arshinova T, Ponomarenko P, Ponomarenko M, Kolchanov 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, Savinkova L, Arshinova T, Ponomarenko M, Peltek S, Kolchanov 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, Pagani 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. Telenti 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. dbWGF: a database and web server of human whole-genome single nucleotide variants and their functional predictions. *Database (Oxford)*. 2016;2016:baw024.
31. Sharypova E, Drachkova I, Kashina E, Rasskazov D, Ponomarenko P, Ponomarenko M, et al. An experimental study of the effect of rare polymorphisms of human HBB, HBD and F9 promoter TATA boxes on the kinetics of interaction with the TATA-binding protein. *Vavilovskii Zhurnal Genetiki i Selektcii*. 2018;22:145–52.
32. Delgadillo R, Whittington J, Parkhurst L, Parkhurst L. The TBP core domain in solution variably bends TATA sequences via a three-step binding mechanism. *Biochemistry*. 2009;48:1801–9.
33. Chadaeva IV, Ponomarenko PM, Rasskazov DA, Sharypova EB, Kashina EV, Zhechev DA, et al. Candidate SNP markers of reproductive potential are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics*. 2018;9:0.
34. Chadaeva IV, Ponomarenko MP, Rasskazov DA, Sharypova EB, Kashina EV, Matveeva MY, et al. Candidate SNP markers of aggressiveness-related complications and comorbidities of genetic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics*. 2016;17:995.
35. Ponomarenko P, Chadaeva I, Rasskazov DA, Sharypova E, Kashina EV, Drachkova I, et al. Candidate SNP markers of familial and sporadic Alzheimer's diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front Aging Neurosci*. 2017;9:231.
36. Arkova OV, Ponomarenko MP, Rasskazov DA, Drachkova IA, Arshinova TV, Ponomarenko PM, et al. Obesity-related known and candidate SNP markers can significantly change affinity of TATA-binding protein for human gene promoters. *BMC Genomics*. 2015;16:S5.
37. Ponomarenko P, Rasskazov D, Suslov V, Sharypova E, Savinkova L, Podkolodnaya O, et al. Candidate SNP markers of chronopathologies are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Biomed Res Int*. 2016;2016:8642703.
38. Haeussler M, Raney BJ, Hinrichs AS, Clawson H, Zweig AS, Karolchik D, et al. Navigating protected genomics data with UCSC genome browser in a box. *Bioinformatics*. 2015;31:764–6.
39. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29:308–11.
40. Drachkova IA, Repkova MN, Lysova MV, Arshinova TV, Savinkova LK. Interaction of proteins from general transcription complex RNA polymerase II with oligoribonucleotides. *Mol Biol (Mosk)*. 2005;39:139–46.
41. 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.
42. Pugh BF. Purification of the human TATA-binding protein, TBP. *Methods Mol Biol*. 1995;37:359–67.
43. Waardenberg AJ, Basset SD, Bouveret R, Harvey RP. CompGO: an R package for comparing and visualizing gene ontology enrichment differences between DNA binding experiments. *BMC Bioinformatics*. 2015;16:275.
44. Levings PP, Bungert J. The human beta-globin locus control region. *Eur J Biochem*. 2002;269:1589–99.
45. Higgs DR, Engel JD, Stamatoyannopoulos G. Thalassemia. *Lancet*. 2012;379: 373–83.
46. Thein SL. Genetic association studies in β -hemoglobinopathies. *Hematology Am Soc Hematol Educ Program*. 2013;2013:354–61.
47. Martiny JA, Cerami A, Slater AF. Inhibition of hemozoin formation in plasmodium falciparum trophozoite extract by heme analogs: possible implication in the resistance to malaria conferred by the beta-thalassemia trait. *Mol Med*. 1996;2:236–46.
48. Eng B, Walker L, Nakamura LM, Hoppe C, Azimi M, Lee H, Wayne JS. Three new beta-globin gene promoter mutations identified through newborn screening. *Hemoglobin*. 2007;31:129–34.
49. 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.
50. Sabath DE, Dettler 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.
51. 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.
52. Lanclos 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-globin genes of beta S chromosomes with five different haplotypes. *Blood*. 1991; 77:2488–96.
53. Winter SS, Kinney TR, Ware RE. Gallbladder sludge in children with sickle cell disease. *J Pediatr*. 1994;125:747–9.
54. Guberman AS, Scassa ME, Cánepa ET. Repression of 5-aminolevulinic synthase gene by the potent tumor promoter, TPA, involves multiple signal transduction pathways. *Arch Biochem Biophys*. 2005;436:285–96.
55. Thunell S. Porphyrins, porphyrin metabolism and porphyrias. I. Update. *Scand J Clin Lab Invest*. 2000;60:509–54.
56. Sassa S, Nagai T. The role of heme in gene expression. *Int J Hematol*. 1996; 63:167–78.
57. Cox EH, McLendon GL, Morel FMM, Lane TW, Prince RC, Pickering IJ, et al. The active site structure of *Thalassiosira weissflogii* carbonic anhydrase. *Biochemistry*. 2000;39:12128–30.
58. Vulić R, Tyciakova S, Dubrovackova M, Skultety 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.
59. Lolak N, Akocak S, Rajesh S, Sanku R, Supuran C. Discovery of new ureido benzenesulfonamides incorporating 1,3,5-triazine moieties as carbonic anhydrase I, II, IX and XII inhibitors. *Bioorg Med Chem*. 2019;15(27):1588–94.
60. Gambhir KK, Ornasir J, Headings V, Bonar A. Decreased total carbonic anhydrase esterase activity and decreased levels of carbonic anhydrase 1 isozyme in erythrocytes of type II diabetic patients. *Biochem Genet*. 2007;45: 431–9.
61. Ostrowski D, Heinrich R. Alternative erythropoietin receptors in the nervous system. *J Clin Med*. 2018;7:E24.

62. Koury 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, Prijatelj T, Debeljak N, Kunej 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, Wenger RH, Rivas LA, Straumann U, Digicaylioglu 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. Rukova B, Staneva R, Hadjidekova S, Stamenov G, Milanova 5th, Toncheva D. Genome-wide methylation profiling of schizophrenia Balkan. *J Med Genet*. 2015;17:15–23.
69. Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol*. 1990;212:563–78.
70. Stewart JJ, Fischback JA, Chen X, Stargell LA. Non optimal TATA element exhibits diverse mechanistic consequences. *J Biol Chem*. 2006;281:22665–73.
71. Stewart JJ, Stargell LA. The stability of the TFIIA-TBP-DNA complex dependent on the sequence of the TATAAA element. *Biol Chem*. 2001;276: 30078–84.
72. Wobbe CR, Struhl K. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. *Mol Cell Biol*. 1990;10:3859–67.
73. Faiger H, Ivanchenko M, Cohen I, Haran TE. TBP flanking sequences: asymmetry of binding, long-range effects and consensus sequences. *Nucleic Acids Res*. 2006;34:104–19.
74. Savinkova LK, Drachkova IA, Ponomarenko MP, Lysova MV, Arshinov 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.
75. Chen RL, Chou YC, Lan YJ, Huang TS, Shen CK. Developmental silencing of human zeta-globin gene expression is mediated by the transcriptional repressor RREB1. *J Biol Chem*. 2010;285:10189–97.
76. Murayama H, Shinkai S, Nishi M, Taniguchi Y, Amano H, Seino S, et al. Albumin, hemoglobin, and the trajectory of cognitive function in community-dwelling older Japanese: a 13-year longitudinal study. *J Prev Alzheimers Dis*. 2017;4:93–9.
77. Ince B, Guloksuz S, Altınbaş K, Oral ET, Alpkan LR, Altinoz MA. Minor hemoglobins HbA2 and HbF associate with disease severity in bipolar disorder with a likely protective role of HbA2 against post-partum episodes. *J Affect Disord*. 2013;151:405–8.
78. Mangin O. High oxygen affinity hemoglobins. *Rev Rev Med Interne*. 2017;38: 106–12.
79. Marvasti VE, Dastoori P, Karimi M. Is beta-thalassemia trait a risk factor for developing depression in young adults. *Ann Hematol*. 2006;85:873–4.
80. Maziade M, Roy MA, Chagnon YC, Cliche D, Fournier JP, Montgrain N, et al. Shared and specific susceptibility loci for schizophrenia and bipolar disorder: a dense genome scan in eastern Quebec families. *Mol Psychiatry*. 2005;10: 486–99.
81. He Z, Russell JE. Anti-sickling effects of an endogenous human alpha-like globin. *Nat Med*. 2004;10:365–7.
82. He Z, Lian L, Asakura T, Russell JE. Functional effects of replacing human [alpha] and [beta] globins with their embryonic globin homologues in defined hemoglobin heterotetramers. *Br J Haematol*. 2000;109:882–90.
83. Russell JE, Liebhaber SA. Reversal of lethal alpha- and beta-thalassemias in mice by expression of human embryonic globins. *Blood*. 1998;92:3057–63.
84. Macias-Velez RJ, Fukushima-Diaz de León L, Beas-Zárate C, Rivera-Cervantes MC. 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.
85. Winter SS, Howard T, Ware RE. Regulation of expression of the human erythropoietin receptor gene. *Blood Cells Mol Dis*. 1996;22:214–24.
86. Marcuzzi F, Zucchelli S, Bertuzzi M, Santoro C, Tell G, Carninci P, Gustincich S. Isoforms of the erythropoietin receptor in dopaminergic neurons of the Substantia Nigra. *J Neurochem*. 2016;139:596–609.
87. Gao Y, Mengana Y, Cruz YR, Munoz A, Teste IS, Garcia JD, et al. Different expression patterns of Ng2 and EPOR in the cerebral cortex and hippocampus revealed distinctive therapeutic effects of intranasal delivery of neuro-EPO for ischemic insults to the gerbil brain. *J Histochem Cytochem*. 2011;59:214–27.
88. Tsai PT, Ohab JJ, Kertesz N, Groszer M, Matter C, Gao J, et al. A critical role of erythropoietin receptor in neurogenesis and post-stroke recovery. *J Neurosci*. 2006;26:1269–74.
89. Sousa NDS, Menezes TN, Silva NA, Eulalio MDC, Paiva AA. Prevalence of anemia and correlation between the concentration of hemoglobin and cognitive factors among the elderly. *Cien Saude Colet*. 2018;23:935–44.
90. Andro M, Le Squere P, Estivin S, Gentric A. Anaemia and cognitive performances in the elderly: a systematic review. *Eur J Neurol*. 2013;20: 1234–40.
91. Ferrer I, Gomez A, Carmona M, Huesa G, Porta S, Riera-Codina M, et al. Neuronal hemoglobin is reduced in Alzheimer's disease, argyrophilic grain disease, Parkinson's disease, and dementia with Lewy bodies. *J Alzheimers Dis*. 2011;23:537–50.
92. Bocchetta A, Del Zompo M. Bipolar affective disorder and heterozygous beta-thalassaemia. *Am J Psychiatr*. 1990;147:1094.
93. Bocchetta A. Heterozygous beta-thalassaemia as a susceptibility factor in mood disorders: excessive prevalence in bipolar patients. *Clin Pract Epidemiol Ment Health*. 2005;1:6.
94. Winchester LM, Powell J, Lovestone S, Nevado-Holgado AJ. Red blood cell indices and anaemia as causative factors for cognitivefunction deficits and for Alzheimer's disease. *Genome Med*. 2018;10:51.
95. Shah RC, Buchman AS, Wilson RS, Leurgans SE, Bennett DA. Hemoglobin level in older persons and incident Alzheimer disease: prospective cohort analysis. *Neurology*. 2011;77:219–26.
96. Hong CH, Falvey C, Harris TB, Simonsick EM, Satterfield S, Ferrucci L, et al. Anemia and risk of dementia in older adults: findings from the health ABC study. *Neurology*. 2013;81:528–33.
97. Atti AR, Palmer K, Volpato S, Zuliani G, Winblad B, Fratiglioni L. Anaemia increases the risk of dementia in cognitively intact elderly. *Neurobiol Aging*. 2006;27:278–84.
98. Ponomarenko M, Rasskazov D, Chadaeva I, Sharypova E, Ponomarenko P, Arkova O, Kashina E, Ivanisenko N, Zhechev D, Savinkova L, Kolchanov N. SNP_TATA_Comparator: genomewide landmarks for preventive personalized medicine. *Front Biosci (Schol Ed)*. 2017;9:276–306.
99. Grigor'eva EV, Malankhanova TB, Surumbayeva A, Minina JM, Morozov VV, Abramycheva NY, Illarioshkin SN, Malakhova AA, Zakian SM. Generation of induced pluripotent stem cell line, ICGi007-a, by reprogramming peripheral blood mononuclear cells from a patient with Huntington's disease. *Stem Cell Res*. 2019;34:101382.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

