RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

The Mechanism by which TATA-Box Polymorphisms Associated with Human Hereditary Diseases Influence Interactions with the TATA-Binding Protein

Irina Drachkova,[1] Ludmila Savinkova,[1]* Tatyana Arshinova,[1] Mikhail Ponomarenko,[1] Sergey Peltek,[1] and Nikolay Kolchanov[1,2]

[1] *Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, Russia;* [2] *Novosibirsk State University, Novosibirsk, Russia*

Communicated by Sergio Ottolenghi

Received 3 October 2013; accepted revised manuscript 20 February 2014.
Published online 25 February 2014 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22535

ABSTRACT: SNPs in **TATA** boxes are the cause of monogenic diseases, contribute to a large number of complex diseases, and have implications for human sensitivity to external and internal environmental signals. The aim of this work was to explore the kinetic characteristics of the formation of human **TBP** complexes with **TATA** boxes, in which the SNPs are associated with β–thalassemias of diverse severity, immunosuppression, neurological disorders, and so on. It has for the first time been demonstrated, using an electrophoretic mobility shift assay, that TBP interacts with SNP-containing **TATA** boxes with a significant (8–36-fold) decrease in **TBP/TATA** association rate constant (k_a) as compared with that in healthy people, a smaller decrease in dissociation rate constant (k_d) and changes in the half-lives of **TBP/TATA** complexes. Carriers of the **−24G** allele (rs 1800202T>G) in the TATA box of the triosephosphate isomerase gene promoter, associated with neurological and muscular disorders, were observed to have a 36-fold decrease in TBP/TATA association rate constant that are consistent with TPI deficiency shown for patients who carry this defective allele. The kinetic characteristics of **TBP/TATA** complexes obtained suggest that, at a molecular level, hereditary diseases are largely caused by changes in **TBP/TATA** association rates and these changes have a bearing on disease severity.
Hum Mutat 35:601–608, 2014. © 2014 Wiley Periodicals, Inc.

KEY WORDS: hereditary disease; TBP; TATA box; association rate; dissociation rate

Introduction

Although sequence data have become more affordable, data analysis is not trivial and requires comparative analysis of identified SNPs to separate disease-causing SNPs from common variants within a population and the study of the molecular mechanisms by which disease-associated SNPs influence the binding of target transcription factors (TFs). Understanding the impact of genetic variation on gene regulation remains a major challenge in deciphering the human transcriptional regulatory code. SNPs represent DNA sequence variants that can affect human sensitivity to external and internal signals. Unlike the SNPs that are found in coding regions and lead, as a result of nonsynonymous substitutions, to incorrect protein folding, phosphorylation, and changes in the properties of the encoded proteins, the SNPs that are found in noncoding regions (that is, regulatory SNPs) may result in modulation of recognition by sequence-specific TFs and altered gene expression [Gelinas et al., 1985; Rockman et al., 2002]. However, predicting the effects of DNA sequence variation in the large regulatory parts of the genome remains a largely unsolved problem. Regulatory SNPs as these, in particular, those in TATA boxes, represent a still poorly studied class of genetic variation. The understanding of the association between the genotype and phenotypic variation and the molecular mechanisms underlying changes in regulatory sequences is required for predicting the organism's responses to environmental effects, drug therapy, susceptibility to diseases, and for personalized medicine. However, the proportion of SNPs within regulatory DNA that have reproducible functional implications for regulatory factor binding is currently unknown, and our ability to predict such outcomes from the known rules of protein–DNA interactions is extremely limited. The major problem in identification and analysis of functional regulatory SNPs as these is associated with the lack of precise and experimentally verified computer-aided techniques. To improve the situation, we have performed a comprehensive experimental computer-aided study of TATA boxes, the SNPs in which are associated with increased risk of various human hereditary diseases. Based on original and literature data, we developed a step-by-step model of TBP binding to TATA boxes [Ponomarenko et al., 2008], and made in silico predictions for changes in TBP/TATA affinity. Next, we performed an experimental verification of the predicted values (the experimental values for equilibrium K_D characterizing the affinity TBP/TATA had already been known) and found a high correlation [Savinkova et al., 2013] between the predicted and experimental values (the coefficient of linear correlation, r, for the predicted equilibrium K_D values was 0.822). However, one question remains: What implications do a change in TBP/TATA affinity have for the process of complex formation? In this work, we have studied the molecular mechanisms of the SNPs effect on the kinetic parameters of the formation of TBP/TATA complexes. Our

Contract grant sponsors: Russian Foundation for Basic Research (14-04-00485 and Research schools-5278.2012.4); Siberian Branch of the Russian Academy of Sciences (136); Russian Academy of Sciences (B.30 (30.29)).

*Correspondence to: Ludmila Savinkova, Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences , Novosibirsk 630090, Russia. E-mail: savinkl@mail.ru

© 2014 WILEY PERIODICALS, INC.

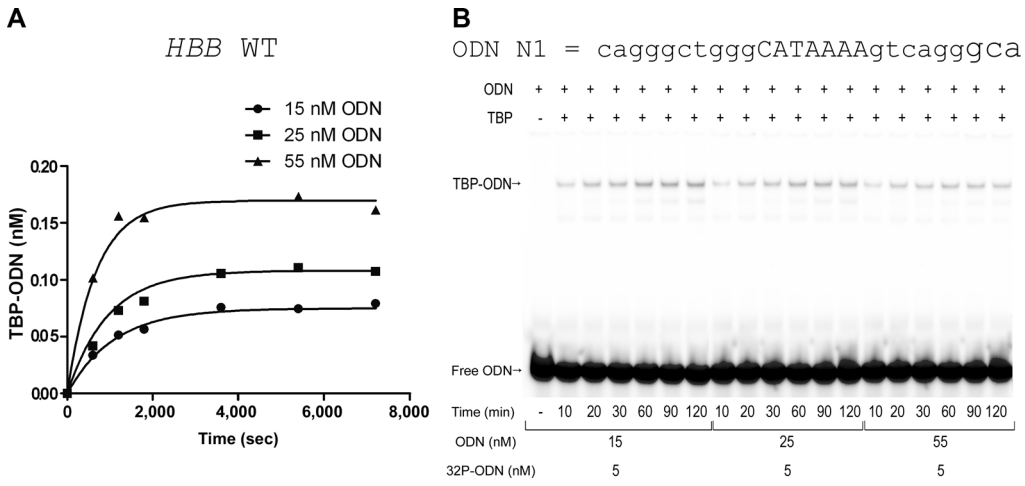

Figure 1. Measuring the kinetics of TBP binding with the TATA-containing ODN N1 identical to the *HBB* promoter in healthy people. **A:** dependences of reaction rates on ODN N1 concentrations. **B:** Electropherograms, from which these curves were derived. TBP concentration was 0.3 nM in all experiments; the concentrations of TATA-containing ODN were as indicated. TBP/TATA-associated isotherms, and k_a and k_d values were inferred from the electropherograms using GraphPad Prism 5 software.

results demonstrate that, at a molecular level, hereditary diseases are caused by changes in TBP/TATA association and dissociation rate constants, by changes in the lifetime of the complexes, and that these changes have a bearing on disease severity.

Materials and Methods

Protein Expression and Purification

Recombinant full-length human TBP containing only native amino acid sequences was overexpressed in *Escherichia coli* BL21(DE3) cells transformed with plasmid pAR3038-*hTBP* gene (*TBP*; MIM #600075), very much appreciated courtesy of Prof. B. Pugh, Pennsylvania State University. Expression of TBP was as described by Pugh (1995) with modifications (the isopropyl β-D-1-thiogalactopyranoside concentration was 1 mM instead of 0.1 mM and the induction time was 3 hr instead of 1.5 hr). TBP was purified to homogeneity using a three-step procedure involving polyethylenimine precipitation, phosphocellulose chromatography, and ammonium sulfate precipitation as described by Pugh (1995). Based on Coomassie Brilliant Blue R 250 stained SDS-PAGE analyses coupled with scanning densitometry, the purity of these TBP preparations was determined to be more than 98%. The total protein concentration was determined after Bradford (1976). The concentration of active TBP was determined by titrations of TBP against known concentrations of the AdML TATA box, which were well above the K_D value (50% of the total protein concentration).

Labeling Oligodeoxyribonucleotides with ^{32}P

Twenty-six base-pair TATA-containing oligodeoxyribonucleotides (ODNs) identical to the promoters of the wild-type (WT) and SNP-containing human genes were synthesized and additionally purified by electrophoresis in PAGE (Biosset, Novosibirsk, Russia). ODN quality was tested for with the use of MALDI TOF MS (Bruker Daltonics, Bremen, Germany). Labeled double-stranded ODNs were obtained and analyzed as described [Drachkova et al., 2005; Savinkova et al., 2013].

Determination of the Association and Dissociation Rate Constants for TBP/TATA Complexes

The association rate constant (k_a) and the dissociation rate constant (k_d) were determined for the complexes of TBP with TATA-containing double-stranded ODNs identical to WT and SNP-containing TATA-box variants. Association kinetic experiments were performed using multiple (typically four) ODN concentrations. A single best-fit estimate for k_a and a single best-fit estimate for k_d were obtained by global fitting of data to the association kinetic model. Experiments on TBP/ODN binding were performed at 25°C in a binding buffer containing 20 mM HEPES–KOH (pH 7.6), 5 mM $MgCl_2$, 70 mM KCl, 1 mM DTT, 100 μg/ml BSA, 0.01% NP-40, and 5% glycerol with a fixed amount of active TBP (typically 0.3 nM). Reaction mixtures containing the binding buffer and ODN, and TBP prediluted with the binding buffer to a concentration of 1.2 nM (that was much lower than the TBP dimerization K_D) were stored in ice. Prediluted TBP was stable in ice for 6 hr. Each association kinetic experiment typically includes 32 binding reactions (8 time points × 4 ODN concentrations). All the four binding reactions (four concentrations at each time point) were run simultaneously by adding TBP and immediately transferring material to a thermostat at 25°C. At the end of the binding reactions, all the reaction mixtures were loaded simultaneously into the running gel. The TBP-ODN complexes were separated from the unbound ODN using an electrophoretic mobility shift assay. Electrophoresis was performed using 5% PAGE in Tris–glycine buffer (pM 8.3) for 40 min at a temperature of 10°C and a field intensity of 25 V/cm. The gels were dried and exposed to an Imaging Screen-K (Kodak, Rochester, NY) for use with a Molecular Imager PharosFX Plus phosphorimager (Bio-Rad, Herts HP2 7DX, United Kingdom). The screen was scanned by the phosphorimager and the radioautographs were quantitated using Quantity One 4.5.0 software (Bio-Rad, Herts HP2 7DX, United Kindom). The association rate constant (k_a) and the dissociation rate constant (k_d) were determined using GraphPad Prism 5 software (Equation: Association kinetics (two ligand concentrations)). Figures 1A and B and 2A and B present examples of binding isotherms for TBP and the WT TATA box of β-globin (ODN N1), and for TBP and the TATA box with a –29A>G SNP

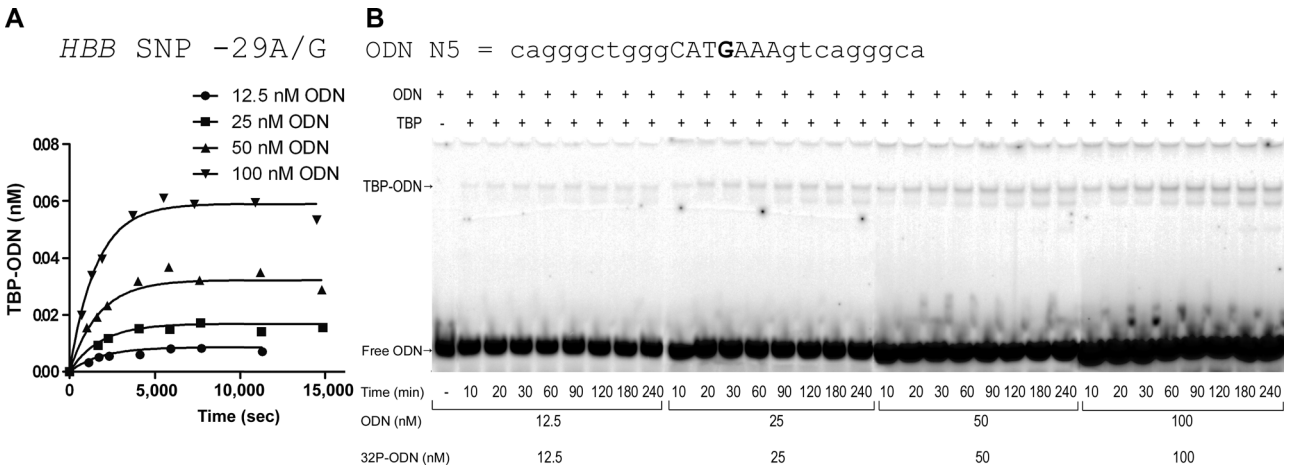

Figure 2. Measuring the kinetics of TBP binding with the TATA-containing ODN N5 identical to the *HBB*-globin promoter with the TATA box containing a SNP associated with β-thalassemia (the –29 G allele). **A:** Dependences of reaction rates on ODN N5 concentrations. **B:** Electropherograms, from which these curves were derived. TBP concentration was 0.3 nM in all experiments; the concentrations of TATA-containing ODN were as indicated. TBP/TATA-associated isotherms, and k_a and k_d values were inferred from the electropherograms using GraphPad Prism 5 software.

(ODN N5), respectively. [Note: the numbering of alleles, for example, "–29A>G" refers to the polymorphism name in original reports, for example, 5′-3′ DNA strand, 29 nucleotides upstream of +1 transcription start site.] The equilibrium dissociation constant of the TBP/TATA complexes, which characterizes TBP/TATA affinity, is $K_D = k_d/k_a$; the average half-life of the TBP/TATA complexes is $t_{1/2} = \ln 2/k_d$. The Gibbs free energy of binding, $\Delta G^0 = -RT\ln K_A$, where R is the universal gas constant, T is the absolute temperature, and K_A (the equilibrium constant) is k_a/k_d.

Results and Discussion

Table 1 presents the values of association rate constants (k_a) and dissociation rate constants (k_d) for the complexes formed by TBP and ODNs, which are identical to TATA boxes with the flanking regions of the promoters of the genes being studied in healthy individuals and patients with SNPs in the TATA boxes associated with clinical and experimentally confirmed hereditary α-, β-, and δ-thalassemias (MIM #604131, 613985, and 613985 respectively) of diverse severity, neurological (MIM #615512) and immune disorders (MIM #614372), myocardial infarction (MIM #608446), hemophilia B Leyden (MIM #613985), and lung cancer (MIM #211980). Importantly, these values have for the first time been obtained under the most standardized experimental conditions. Also, Table 1 presents the half-lives ($t_{1/2}$) of the TBP/TATA complexes, the values of the dissociation constants ($K_D = k_d/k_a$), which characterize TBP/TATA affinity, and the Gibbs free energy of binding (ΔG^0).

SNPs as an Influence on the Association and Dissociation Rates of Complexes of TBP with the TATA Boxes of the *HBB*, *HBD*, and *HBZP1* Genes

As is known [Steinberg et al., 1991; Flint et al., 1998; Levings et al., 2002; Thein, 2005; Lacerra et al., 2008; Schechter et al., 2008; Muncie et al., 2009; Galanello et al., 2010; Jain et al., 2012], SNPs in the TATA boxes of the promoters of the β–, δ–globin genes, and the $\varepsilon\psi$-globin pseudogene (*HBB*; MIM #141900, *HBD*; MIM #142000, and *HBZP*; HGNC:4836, respectively) lead to β–, δ–, and α–thalassemias of diverse severity because of the disruption of the balance in the synthesis of structurally normal globin chains, which compose normal hemoglobins: A ($\alpha_2\beta_2$, the commonest structural unit of hemoglobin in adult humans, HbA, with an amount of ~97%) and A2 ($\alpha_2\delta_2$, with an amount of ~3%) [Schechter, 2008]. Imbalances of any of three globin chains cause their aggregation, hemolysis, and failure in erythropoiesis.

As can be seen from Table 1, the TATA box of the *HBB* gene in a healthy individual (ODN 1) appears as CATAAAA [Muncie et al., 2009] instead of canonical TATAAAA. It has been demonstrated [Stewart et al., 2001; Stewart et al., 2006; Muncie et al., 2009] that a T-to-C substitution at position 1 in the TATA box variously affects transcription and, therefore, TBP/TATA complex formation. The values that we obtained for this noncanonical TATAAAA *HBB* were as follows: k_a, which characterizes the rate of association of complexes with TBP and the TATA box of *HBB* in healthy individuals, was $(1.4 \pm 0.1) \times 10^4$ M^{-1} sec^{-1}, the dissociation rate constant, k_d, was $(7.1 \pm 0.7) \times 10^{-4}$ sec^{-1}, the equilibrium dissociation constant, K_D, was 50 ± 7 nM, the half-life of the complexes was 16 ± 2 min, and the free energy of association of TBP/TATA complexes (ΔG^0) was 10 ± 1 kcal/mol.

As can be seen from Table 1, the values of the TBP/TATA association constants for *HBB genes* with different SNPs in the TATA box are reduced from eightfold to 34-fold. A 2.6-fold increase in the rate of association of TBP/TATA complexes and a threefold increase in the rate of their dissociation as compared with the norm was observed only once, when the SNP was NG_0000007.3:g.70518A>T (–27A>T, ODN N8), which was found members of a Corsican family [Badens et al., 1999] with signs of β^+-thalassemia. That was accompanied by a threefold decrease in the half-life of the complexes as compared with the norm (healthy individuals): the half-life was reduced to 5 min. The affinity of TBP for the TATA box with that SNP (60 ± 10 nM) is decreased by ~20% as compared with the norm (50 ± 7 nM), and the amount of β-globin mRNA in *HeLa* is decreased about fivefold [Badens et al., 1999]. The free energy of association of TBP/TATA complexes (ΔG^0) is not changed.

The strongest, 34-fold decrease in the rate of association of TBP/TATA complexes ($k_a = (4.1 \pm 0.5) \times 10^2$ M^{-1} sec^{-1}) is

Table 1. Kinetic and Thermodynamic Characteristics of TBP/TATA Interactions

ODN Number	Gene; OMIM	Allele[a]	Disease (OMIM)	Sequences, 5′-3′ Strands	k_a (M^{-1} c^{-1})	k_d (c^{-1})	$t_{1/2}$ (min)	K_D (nM)	ΔG–(kcal/mol)
1	*HBB*; 141900	WT	Normal	cagggctgggCATAAAAgtcagggca	$(1.4 \pm 0.1) \times 10^4$	$(7.1 \pm 0.7) \times 10^{-4}$	16 ± 2	50 ± 7	10 ± 1
2		rs33981098:A>G (–31A>G)	β-Thalassemia (613985)	cagggctgggC**G**TAAAAgtcagggca	ND	ND	ND	99 ± 9[a]	
3		rs33980857:T>A(–30T>A)	β-Thalassemia (613985)	cagggctgggCA**A**AAAAgtcagggca	$(1.7 \pm 0.2) \times 10^3$	$(1.9 \pm 0.3) \times 10^{-4}$	60 ± 10	110 ± 20	9.5 ± 0.9
4		rs33980857:T>C (–30T>C)	β-Thalassemia (613985)	cagggctgggCA**C**AAAAgtcagggca	$(8 \pm 2) \times 10^2$	$(1.1 \pm 0.3) \times 10^{-4}$	100 ± 30	140 ± 60	10.7 ± 1.5
5		rs34598529:A>G (–29A>G)	β-Thalassemia (613985)	cagggctgggCAT**G**AAAgtcagggca	$(1.1 \pm 0.3) \times 10^3$	$(5.3 \pm 0.4) \times 10^{-4}$	22 ± 2	500 ± 200	8.6 ± 1.2
6		rs33931746:A>G (–28A>G)	β-Thalassemia (613985)	cagggctgggCATA**G**AAgtcagggca	$(1.2 \pm 0.1) \times 10^3$	$(6.6 \pm 0.9) \times 10^{-4}$	18 ± 2	560 ± 80	8.5 ± 0.9
7		rs33931746:A>C (–28A>C)	β-Thalassemia (613985)	cagggctgggCATA**C**AAgtcagggca	$(4.1 \pm 0.5) \times 10^2$	$(2.0 \pm 0.3) \times 10^{-4}$	58 ± 9	500 ± 100	8.6 ± 1.4
8		NG_000007.3:g.70518A>T (–27A>T)	β-Thalassemia (613985)	cagggctgggCATAA**T**Agtcagggca	$(3.6 \pm 0.4) \times 10^4$	$(2.2 \pm 0.4) \times 10^{-3}$	5 ± 1	60 ± 10	10 ± 0.9
9	*HBD*; 142000	WT	Normal	acaggaccagCATAAAAggcagggca	$(2.3 \pm 0.3) \times 10^4$	$(8 \pm 2) \times 10^{-4}$	14 ± 4	36 ± 4	10 ± 1
10		rs35518301:A>G (–31A>G)	δ-Thalassemia (613985)	acaggaccagC**G**TAAAAggcagggca	$(2.0 \pm 0.1) \times 10^3$	$(2.0 \pm 0.4) \times 10^{-4}$	60 ± 10	100 ± 20	9.5 ± 0.9
11	*HBZP1*; HGNC: 4836	WT	Normal	ctgccacacccacattattagaaaat	$(2.3 \pm 0.3) \times 10^3$	$(4.2 \pm 0.6) \times 10^{-4}$	28 ± 4	160 ± 30	9.3 ± 1
12		NG_000006. 1:g..20573T>C (–70T>C)	α-Thalassemia (604131)	ctgccacaccCACATTAT**C**agaaaat	$(2.4 \pm 0.2) \times 10^3$	$(2.5 \pm 0.3) \times 10^{-4}$	46 ± 6	100 ± 10	9.5 ± 1.3
13	*MBL2*; 154545	WT	Normal	catctatttcTATATAGcctgcaccc	$(3.0 \pm 0.3) \times 10^4$	$(1.0 \pm 0.1) \times 10^{-3}$	12 ± 1	34 ± 5	10.2 ± 1
14		rs72661131:T>C (–35T >C)	Immunodeficiency (614372)	catctatttcTA**C**ATAGcctgcaccc	$(1.2 \pm 0.1) \times 10^4$	$(6.4 \pm 0.7) \times 10^{-4}$	18 ± 2	54 ± 7	9.9 ± 0.9
15	*TP1I*; 190450	WT	Normal	cgcggcgctcTATATAAgtgggcagt	$(1.8 \pm 0.1) \times 10^5$	$(7.2 \pm 0.8) \times 10^{-4}$	16 ± 2	4.1 ± 0.5	11.4 ± 0.9
16		rs1800202:T>G (–24T>G)	TPI-deficiency (615512)	cgcggcgctcTATA**G**AAgtgggcagt	$(5.0 \pm 0.8) \times 10^3$	$(5.1 \pm 0.7) \times 10^{-4}$	23 ± 3	100 ± 20	9.5 ± 1.4
17	*F9*; 300746	WT	Normal	acagctcagcTTGTACTttggtacaa	$(2.4 \pm 0.2) \times 10^3$	$(1.4 \pm 0.1) \times 10^{-3}$	8.3 ± 0.6	580 ± 20	8.5 ± 0.6
18		NG_007994.1:g.4975G>C (–48G>C)	Hemophilia B Leyden (306900)	acagctcagcTT**C**TACTttggtacaa	$(1.0 \pm 0.1) \times 10^3$	$(5.6 \pm 0.9) \times 10^{-4}$	21 ± 3	540 ± 30	8.6 ± 1.2
19	*IL1B;* 147720	WT	Normal	ttttgaaagcCATAAAAacagcgagg	$(1.0 \pm 0.1) \times 10^4$	$(2.1 \pm 0.5) \times 10^{-4}$	60 ± 10	20 ± 4	10.5 ± 1.6
20		rs1143627:C>T (–31C>T)	Lung cancer (211980)	ttttgaaagc**T**ATAAAAacagcgagg	$(1.6 \pm 0.2) \times 10^5$	$(8 \pm 1) \times 10^{-4}$	14 ± 2	4.8 ± 0.6	11.3 ± 1.5
21	*F3*; 134390	WT	Normal	gccggcccTTTATAgcgcgcggggca	$(2.3 \pm 0.2) \times 10^4$	$(1.7 \pm 0.2) \times 10^{-3}$	6.8 ± 0.8	77 ± 9	9.7 ± 1
22		NG_029366.1:g.5078C>T (–21C>T)	Myocardial infarction (608446)	gccggcccTTTATAg**T**gcgcggggca	$(1.0 \pm 0.1) \times 10^5$	$(1.9 \pm 0.2) \times 10^{-3}$	6.1 ± 0.6	19 ± 1	10.5 ± 1
23	*NOS2*; 163730	WT	Normal	atggggtgagTATAAATActtcttgg	$(4.2 \pm 0.4) \times 10^5$	$(8 \pm 1) \times 10^{-4}$	14 ± 2	1.9 ± 0.3	12 ± 4
24		NG_011470.1:g.4980T>C (–60T>C)	Susceptibility to malaria (611162)	atggggtgagTATAAATAc**C**tcttgg	$(4.0 \pm 0.3) \times 10^5$	$(7.3 \pm 0.9) \times 10^{-4}$	16 ± 2	1.8 ± 0.2	11.9 ± 1.1

ODN, oligodeoxyribonucleotide number; N.D., not determined; $k\pm$, standard deviation.

[a]dbSNP or HGVS as per journal recommendations. ODN N18 and N24 SNPs numbers correspond to the current version of the reference human genome. Variant names given in parentheses, for example, (–35T>C), refer to polymorphism name in original articles (5′-3′ DNA strand, 35 nucleotides to the left of +1 transcription start site).

observed when the SNP is rs33931746 (–28A>C, ODN N7). The rate of their dissociation is decreased about 3.6-fold and their half-life is increased to 58 ± 9 min, whereas the affinity (K_D) is reduced 10-fold and becomes nearly nonspecific. The carrier of this polymorphism has β–thalassemia major [Poncz et al., 1982], depends on red blood cell transfusions for the rest of his life and is a compound heterozygote [Surrey et al., 1985]. The free energy of TBP/TATA binding is decreased by 1.4 kcal/mol, which is the cost for the departure from the better sequence (CATAAAA → CATACAA). It has also been demonstrated that transcription in *HeLa* is decreased twofold to threefold, and transcription in vitro is decreased fourfold [Surrey et al., 1985]. Opinion has it [Bucher, 1990; Wang et al., 2003] that the probability of this substitution in the TATA box is less than 1%.

A rs33980857 (–30T>A, ODN N3) found in a Yugoslavian and a Turkish individual with β^+-thalassemia intermedia [Muncie et al., 2009] leads to higher than an eightfold decrease in the rate of association of TBP/TATA complexes and a 3.7-fold decrease in the rate of their dissociation. Their half-life is increased to 60 ± 10 min (16 ± 2 min in healthy individuals), which enables the assembly of transcription complexes with the production rate of normal β-globin RNA between 8% and 13%. The free energy of TBP/TATA binding is decreased slightly. It should be noted that T at position 3 in the TATA box reportedly occurs with the highest probability and displays the highest transcriptional activity in diverse systems, whereas T to A substitutions bring transcription down to undetectable levels [Wang et al., 2003].

A rs33980857T>C leads to higher than an 18-fold decrease in the rate of association of TBP/TATA complexes (ODN N4), the strongest decrease (6.5-fold) in the rate of dissociation, and the longest (100 ± 30 min) half-life as compared with that in healthy individuals, and a moderate, 2.8-fold decrease in the affinity of TBP for ODN N4 with this SNP. The free energy of TBP/TATA binding is increased by 0.7 kcal/mol. It had previously been demonstrated [Wobbe et al., 1990] that an T to C substitution at position 3 in the TATA box of the AdML promoter (TATAAAA) is preferable to an T to A substitution, with a ~16% residual transcriptional activity.

A rs34598529 (–29A>G, ODN N5) resulted in a 25% production of normal globin mRNA in the erythroid cells of affected individuals and in *HeLa* [Antonarakis et al., 1984]. The rate of association of complexes containing TBP and the TATA boxes with this SNP is reduced 13-fold, the rate of their dissociation changes little and so does their half-life ($t_{1/2}$ = 22 ± 2 min) as compared with those in healthy individuals. The K_D value, which characterizes TBP/TATA affinity, is decreased 10-fold. ΔG^0 is decreased by 1.4 kcal/mol, and the energy of binding is spent to probably compensate for a low affinity. Nevertheless, the homozygotes for this allele are affected by β-thalassemia intermedia and do not depend on red blood cell transfusions [Antonarakis et al., 1984].

A rs33931746A>G (–28A>G, ODN N6) found in a transfusion-dependent patient with β-thalassemia intermedia in a Chinese population [Orkin et al., 1983] leads to a 12-fold decrease in the rate of association of TBP/TATA complexes. The rate of their dissociation and half-life change little (7.5%). The K_D value, which characterizes TBP/TATA affinity, is decreased 11-fold. This mutation is accompanied by the highest decrease in ΔG^0 (1.5 kcal/mol) in β-globin genes. It is possible that this energy is used for strengthening low-affinity TBP/TATA complexes. The amount of β-globin mRNA was decreased 10-fold in the total RNA isolated from the erythroid cells of patients who were carriers of that allele and threefold to fivefold in *HeLa* [Orkin et al., 1983] as compared with the norm. The same A to G substitution at position 5 of the TATA box in in vitro experiments with AdMLP resulted in a reduction in transcription to undetectable levels [Wobbe et al., 1990; Wang et al., 2003].

Regrettably, we could not determine the association or dissociation rate constant of TBP/TATA complexes for a rs33981098A>G (–31A>G, ODN N2) [Takihara et al., 1986]. The K_D value determined under equilibrium conditions was 99 ± 9 nM. Half the normal amount of mRNA is synthesized from the promoter with this mutation in the TATA box in *cos*-cells [Takihara et al., 1986].

The TATA-box sequence in the δ-globin (*HBD*) gene is identical to that in the β-globin (*HBB*) gene, CATAAAA; however, its flanking sequences are different [Frischknecht et al., 2005]. As a result, the rate of association of TBP/TATA complexes for the ODN identical to the WT δ-globin promoter is 1.6 times higher than that for the ODN of the β-globin gene promoter in healthy individuals (k_a = (2.3 ± 0.3) × 10^4 and (1.4 ± 0.1) × 10^4 M^{-1} c^{-1}, respectively), the affinity is increased 1.4-fold, and the free energy of TBP/TATA binding remains unchanged. Whenever there is rs35518301A>G (–31A>G, ODN10) in the TATA box, the rate of association of TBP/TATA complexes is decreased 12-fold as compared with healthy individuals, the dissociation rate is decreased fourfold, and the half-life is increased fourfold (this is 14 ± 4 min in the norm and 60 ± 10 min in δ-thalassemia). The TBP/TATA affinity is decreased 2.8-fold as compared with the norm, ΔG^0 is decreased by 1.4 kcal/mol and all these characteristics are associated with δ-thalassemia intermedia. Carriers of homozygous alleles with SNPs have a very low HbA2 production rate [Frischknecht et al., 2005].

A –70 T>C SNP (NG:000006.1:g.20573T>C) found in the *HBZP1*(HGNC:4836) pseudogene creates a new promoter element, GATA/TATA, which inhibits transcription of the downstream α-globin genes and so accounts for a deficiency of α-globin chains and α-thalassemia intermedia. When this mutation is the case, the TBP/TATA association rate remains unaffected, whereas the dissociation rate is decreased 1.7-fold and the half-life is increased 1.7-fold as compared with the norm. The constant of dissociation of the TBP/TATA complexes, which characterizes TBP/TATA affinity, is increased 1.6-fold; it is also expected that the erythroid-specific activator GATA-1 should bind well to the mutant region. Collectively, all these factors are the reason why the transcription initiation complex is assembled on the pseudo promoter and why TBP and TFs are lured from the promoters of WT α-globins [De Gobbi et al., 2006].

An SNPs Influence on the Association and Dissociation Rates of Complexes of TBP with the TATA Boxes of the *MBL2, TPI,* Factor *IX or F9, IL1B, TF,* or *F3,* and *NOS2A* Genes

The rs72661131T>C in the TATA box of the mannose-binding lectin gene promoter (*MBL2*; MIM #154545) [Boldt et al., 2006] leads to a 2.5-fold decrease in the rates of association of the complexes and a 1.6-fold decrease in dissociation rates. The half-life of the complexes is increased 1.6-fold and the affinity of TBP for the TATA box is decreased 1.6-fold. A rs72661131C>T substitution at position 3 of the TATA box leads to a slight decrease in the free energy of association of TBP/TATA complexes (ΔG^0), by 0.3 kcal/mol (from 10.2 to 9.9 kcal/mol). MBL is a serum protein, which binds to mannose and N-acetyl glucosamine on the surface of microorganisms, opsonizes, and eventually kills them. MBL plays an important role in tissue homeostasis and removal of endogenous wastes. This is why it is believed that genetically determined variation in MBL concentrations in human blood serum accounts for varying sensitivity to infections and predisposition to autoimmune, inflammatory, metabolic, and cardiovascular diseases [Madsen et al., 1998]. Low MBL levels are associated with increased risk of recurrent infections [Eisen et al., 2005].

The rs1800202T>G SNP in the TATA box of the triosephosphate isomerase gene (*TPI*; MIM #190450) [Watanabe et al., 1996] leads to a 35.8-fold decrease in TBP/TATA association rates and a slight decrease (by 30%) in dissociation rates. The half-life of the complexes is increased 1.4-fold. When the TATA box is unaffected, the half-life is 16 min, and when the TATA box contains this SNP, the half-life is 23 min. This mutation leads to a decrease in ΔG^0 by 1.9 kcal/mol. TPI-deficient (MIM #615512) people were observed to have insufficient mRNA, the enzyme activity in erythrocytes and other cells was 2%–10% of the norm; patients had neurological and muscular disorders, and hemolytic anemia. Other carriers of the heterozygous alleles were observed to have a moderate decrease in TPI activity in vivo, by 26%–50% [Watanabe et al., 1996]. We have demonstrated that the –24T>G SNP (very rare, with its prevalence being less than 1%) [Patikoglou et al., 1999] associated with these disorders causes the largest, 35.8-fold decrease in TBP/TATA association rate and the largest decrease in affinity among this group of genes. With such decreases, the probability of complex formation is lowered. The dissociation rates are decreased 1.4-fold, and the half-life of the complexes is increased 1.5-fold. As can be seen from Table 1, the association of TBP with one of the very good TATA-box sequences, TATATA, in which the fifth "T" is replaced by "G," is accompanied by the largest decrease in free energy of binding for this group of genes, by 2.26 kcal/mol. It is possible that this energy is used for introducing conformational changes to and strengthening low-affinity TBP/TATA complexes.

The SNP in the TATA-like element of the coagulation Factor *IX* or *F9* gene (MIM #300746) leads to a 2.3-fold decrease in the rates of association of TBP/TATA complexes, a 2.5-fold decrease in dissociation rates and a 2.5-fold decrease in the half-life of these complexes. The SNP has little effect on the affinity of TBP for this TATA-like element (the affinity remains to be very low, K_D = 580 and 540 nM) or on the free energy of association. Hemophilia B Leyden (MIM #306900) is characterized by very low levels of factor 9 (blood coagulation factor) in the blood plasma during infancy (60% of the norm as a maximum). The promoter of the WT *Factor IX* gene contains a region for binding to the hepatocyte nuclear factor HNF4 controlling factor IX expression in healthy individuals [Burgner et al., 2003]. The binding site for HNF-4 is located in the TATA box. The NG_007994.1:g.4975G>C (human reference genome), –48G>C SNP destroys it, and so HNF-4 becomes unable to effectively bind to the target anymore, which leads to hemophilia B Leyden. For TBP, the nucleotide sequence of that region is where the interaction is practically nonspecific.

The rs1143627C>T, –31C>T SNP in the TATA box of the interleukin-1 beta gene (*IL1B*; MIM #147720) [Zienolddiny et al., 2004] (ODN 20), which improves the sequence of the TATA box, leads to a 15.8-fold increase in the rates of association of TBP/TATA complexes and a 3.7-fold decrease in their dissociation rates. The half-life of the complexes is increased nearly fourfold. The rs1143627C>T substitution leads to an increase in ΔG^0 by 0.8 kcal/mol, no cost is required of departure from the consensus sequence and improved longevity. The rs1143627C>T polymorphism is associated with increased risk of a large number of inflammatory diseases, stomach cancer and lung cancer (MIM #211980).

The NG_029366.1:g.5078C>T, –21C>T SNP in the TATA box of the tissue factor, *TF*, or coagulation factor *F3* gene (MIM #134390) [Arnaud et al., 2000] leads to a 4.6 increase in the rate of association of TBP/TATA complexes and has little effect on their dissociation rates or half-lives. The increased association rates due to the SNP are associated with a 4.2-fold increase in TBP/TATA affinity. The free energy of association for these complexes is increased by 0.8 kcal/mol. The tissue factor is a membrane protein expressed in many tissues, including the outermost layer of the vessel walls, where it rapidly activates coagulation whenever integrity is compromised. The SNP in the TATA element of the tissue factor gene is consistent with the known enhancement in gene expression [Arnaud et al., 2000], which leads to increased risk of thrombophlebitis and myocardial infarction (MIM #608446).

The SNP in the sequence that flanks the TATA box of the *NOS2* gene (MIM #163730). The NG_011470.1:g.4980T>C, SNP-60T>C reference human genome or 21T>C [Burgner et al. 2003] in the sequence that flanks the TATA box of the NO synthase gene promoter has little effect (changes are within 10%) on the kinetic characteristics of TBP/TATA complexes and none on ΔG^0. The gene product is inducible nitric oxide synthase (NO synthase), which is involved in many responses produced by the organism, including the immune response. NO synthase also contributes to the pathogenesis of infections and inflammatory conditions and has a role in the regulation of blood pressure [Burgner et al., 2003]. It should be noted that the observed differences in k_a and k_d are not only because of the SNPs within the TATA box, but also because of SNPs in its flanking sequences and because of the structural differences between these sequences.

Based on the results obtained, we cannot say that a low affinity of a TF for its target DNA site should accelerate the dissociation of the complexes (the TF should not linger in a nonspecific DNA region), nor can we say that association rates (k_a) should be the same for specific and nonspecific sites [Dinan et al., 2009]. As can be seen from the values obtained for k_a and k_d (Table 1), a reduction in affinity can lead to a change in both association and dissociation rates; particularly, if the interaction is less specific, the dissociation rates may decrease, which implies that the complexes become longer living (for example, the TBP/TATA complexes in the *MBL2* and *TPI* genes). We are strongly tempted to say that these conditions are specifically set up for a functional multiprotein complex (in our case, the preinitiation complex) to have enough time to be assembled on these "weak" TBP/TATA complexes.

Also, it should be noted that in vivo TBP can be part of TFIID complexes of different composition (from 8 to 14 TBP-associated factors, TAFs), which play critical roles in tissue- and cell-specific transcription [Thomas et al., 2006; Deato et al., 2008; Papai et al., 2011]. Some authors have demonstrated that TBP alone is enough to enable TATA-dependent basal transcription to run; however, activated transcription or transcription from TATA-less promoters absolutely requires TAFs [Klages et al., 1995; Martinez et al., 1995]. The latter is also true of the in vivo initiation of transcription in human and yeast TATA-less promoters [Kim et al., 2005].

Various TAFs within the dynamic TFIID structure perform various functions that have importance for transcriptional regulation in eukaryotic cells. Depending on the core promoter sequence, TAFs can stabilize or destabilize TFIID/DNA binding and influence the formation of the preinitiation complex and promoter strength [Tjian et al., 1996]. TAFs can compensate for a low specificity for TBP binding [Hahn et al., 1989] by changing the rate of TBP dissociation from promoters [De Graaf et al., 2010] and hinder the initiation of transcription from very weak and cryptic promoters [Verrijzer et al., 1996]. It is likely that the interactions between TBP and TATA boxes in the β-globin gene promoter with the –29A>G, –28A>G, and –28A>C SNPs as well as between TBP and TATA boxes in the Factor *IX* gene promoter, which have, according to our results, a very low specificity, are in vivo stabilized and strengthened by TAFs. This enables a large preinitiation complex to form and regulated transcription to run.

Conclusions

Quantitative changes in gene expression caused by polymorphisms in regulatory regions play an important role in human susceptibility to various diseases, environmental conditions, and drug therapy. We have for the first time demonstrated experimentally that the implications of SNPs, associated with increased risk of various human hereditary pathologies, for the interaction between TBP and TATA boxes at a molecular level come through changes in TBP/TATA association/dissociation rates. SNPs lead to a 2.5–36-fold decrease in TBP/TATA association rates. The Gibbs free energy values were all negative, indicating that TBP/TATA binding was spontaneous. Because TBP/TATA binding occurs under the most standardized reaction conditions, at the same temperature, decreased δG^0 values indicate that the TBP/TATA complex is destabilized with SNPs associated with diseases.

A good agreement between the K_D values characterizing TBP/TATA affinity, obtained under equilibrium conditions [Savinkova et al., 2013] and inferred from kinetic constants, suggests that the kinetic parameters of the interactions between TBP and the TATA-containing ODNs were determined correctly. All the characteristics of TBP/TATA association that we have obtained fall within the range of values typical of TBP/TATA interactions.

Obviously, TBP binds to the consensus and polymorphic TATA-box variants in vivo and in vitro under different conditions, with different TAFs and TFs. Nevertheless, we observe a correlation between the extent, to which the rates of association and dissociation of the complexes of TBP and SNP-containing TATA boxes change in vitro, and the degree of severity of the associated diseases. This provides evidence that the results obtained adequately portray the molecular interactions occurring in vivo.

Acknowledgments

We are grateful to Vladimir Filonenko for translating this article from Russian into English, to Vladimir Koval for the MALDI-TOF/MS analysis of oligodeoxynucleotides, and to Olga Arkova for technical assistance in preparing the manuscript.

Author Contributions: I.D. and T.A. performed the research. L.S., M.P., N.K., and S.P. designed the study, and L.S. and I.D. wrote the article.

Disclosure statement: The authors declare no conflict of interest.

References

Antonarakis SE, Irkin SH, Cheng TC, Scott AF, Sexton JP, Trusko SP, Charache S, Kazazian HH Jr. 1984. Beta-thalassemia in American Blacks: novel mutations in the "TATA" box and an acceptor splice site. Proc Natl Acad Sci USA 81:1154–1158.

Arnaud E, Barbalat V, Nicaud V, Cambien F, Evans A, Morrison C, Arveiler D, Luc G, Ruidavets JB, Emmerich J, Fiessinger JN, Aiach M. 2000. Polymorphisms in the 5′ regulatory region of the tissue factor gene and the risk of myocardial infarction and venous thromboembolism: the ECTIM and PATHROS studies. Etude Cas-Te.moins de l'Infarctus du Myocarde. Paris Thrombosis case-control study. Arterioscl Thromb Vasc Biol 20:892–898.

Badens C, Jassim N, Martini N, Mattei JF, Elion J, Lena-Russo D. 1999. Characterization of a new polymorphism, IVS-I–108 (TC), and a new beta-thalassemia mutation, –27 (AT), discovered in the course of a prenatal diagnosis. Hemoglobin 23:339–344.

Boldt AB., Culpi L, Tsuneto LT, de Souza IR, Kun JF, Petzl-Erler ML. 2006. Diversity of the MBL2 gene in various brasilian populations and the case of selection at the mannose-binding lectin locus. Hum Immunol 67:722–734.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254.

Bucher P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol 212:563–578.

Burgner D, Rockett K, Ackerman H, Hull J, Usen S, Pinder M, Kwiatkowski DP. 2003. Haplotypic relationship between SNP and microsarellite marcers at the NOS2A locus in two populations. Gen Immun 4:506–514.

Deato MD, Marr MT, Sottero T, Inouye C, Hu P, Tjian R. 2008. MyoD targets TAF3/TRF3 to activate myogenin transcription. Mol Cell 32:96–105.

De Graaf P, Mousson F, Geverts B, Scheer E, Tora L, Houtsmuller AB, Timmers HT. 2010. Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. J Cell Sci 123:2663–2671.

De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ, Vernimmen D, Yoshinaga Y, de Jong P, Cheng JF, Rubin EM, et al. 2006. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science 312:1215–1217.

Dinan CT, Luijsterburg MS, Hofer T, Gesa von Bornstaedt, Vermeulen W, Houtsmuller AB, Roel van Driel. 2009. Assembly of multiprotein complexes that control genome function. J Cell Biol 185:21–26.

Drachkova IA, Repkova MN, Lysova MV, Arshinova TV, Savinkova LK. 2005. Interaction of proteins from general transcription complex RNA polymerase II with oligoribonucleotides. Mol Biol (Russia) 39:139–146.

Eisen DP, Minchinton RM. 2005. Impact of mannose-binding lectin on susceptibility to infectious diseases. Clin Infect Dis 37:1496–1505.

Flint J, Harding RM, Boyce AJ, Clegg JB. 1998. The population genetics of the hemoglobinopathies. Bailliere's Clin Hematol 11:1–50.

Frischknecht H, Dutly F. 2005. Two new delta-globin mutations: Hb A2-Ninive delta133 (H11) Val-Ala] and a delta(+)-thalassemia mutation [-31 (A –> G)] in the TATA box of the delta-globin gene. Hemoglobin 29:151–154.

Galanello R, Origa R. Beta thalassemia. Orphanet J Rare Diseases 2010. 5:11–27.

Gelinas R, Endlich B, Pfeiffer C, Yagi M, Stamatoyannopoulos G. 1985. G to A substitution in the distal CCAAT box of the A gamma-globin gene in Greek hereditary persistence of fetal haemoglobin. Nature 313:323–325.

Hahn S, Buratowski S, Sharp F, and Guarente L. 1989. Yeast TATA_binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc Natl Acad Sci USA 86:5718–5722.

Jain S, Edison ES, Mathews V, Shaji RV. 2012. A novel δ-globin gene mutation (HBD: c.323 G>A) masking the diagnosis of β-thalassemia: a first report from India. Int J Hematol 95:570–572.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. Nature 436:876–880.

Klages N, Strubin M. 1995. Stimulation of RNA polymerase II transcription initiation by recruitment of TBP *in vivo*. Nature 374:822–823.

Lacerra G, Musollino G, Scarano C, Lagona LF, Caruso DG, Testa R, Prezioso R, Di Noce F, Medulla E, Friscia MG, Mastrullo L, Caldora M, et al. 2008. Molecular evidences of single mutational events followed by recurrent crossing-overs in the common delta-globin alleles in the Mediterranean area. Gene 410:129–138.

Levings PP, Bungert J. 2002. The human beta-globin locus control region. Eur J Biochem 269:1589–1599.

Madsen HO, Satz ML, Hogh B, Svejgaard A, Garred P. 1998. Different molecular events result in low protein levels of mannan-binding lectin in populations. J Immunol 161:3169–3175.

Martinez E, Zhou Q, L'etoile ND, Oelgeschlager T, Berk AJ, and Roeder RG. 1995. Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding. Proc Natl Acad Sci USA 92:11864–11868.

Muncie HL, Campbell JC. 2009. Alpha and beta thalassemia. A. Fam Phys 80:339–344.

Orkin SH, Sexton JP, Cheng TC, Goff SC, Giardina PJ, Lee JI, Kazazian HH Jr. 1983. ATA box transcription mutation in beta-thalassemia. Nucl Acids Res 11:4727–4734.

Papai G, Weil AP, and Schultz P. 2011. New insights into the function of transcription factor TFIID from recent structural studies. Curr Opin Genet Dev 21:219–224.

Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev 13:3217–3230.

Poncz M, Ballantine M, Solowiejczyk D, Barak I, Schwartz E, Surrey S. 1982. Beta-thalassemia in a Kurdish Jew. Single base changes in the T-A-T-A box. J Biol Chem 257:5994–5996.

Ponomarenko PM, Savinkova LK, Drachkova IA, Lysova MV, Ponomarenko MP, Kolchanov NA. 2008. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. Dokl Biochem Biophys 419:88–92.

Pugh BF. 1995. Purification of the human TATA-binding protein, TBP. In: Tymms MJ, editor. *Methods in molecular biology: In vitro transcription and translation protocols*. Totowa, NJ: Humana Press Inc. p 37.

Rockman MV, Wray GA. 2002. Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol 19:1991–2004.

Savinkova L, Drachkova I, Arshinova T, Ponomarenko P, Ponomarenko M, Kolchanov N. 2013. 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 8(2):e54626.

Schechter A. 2008. Hemoglobin research and the origins of molecular medicine. Blood 112:3927–3938.
Steinberg MH, Adams JG. 1991. Hemoglobin A2: origin, evolution, and aftermath. Blood 78:2165–2177.
Stewart JJ, Stargell LA. 2001. The stability of the TFIIA–TBP–DNA complex dependent on the sequence of the TATAAA element. J Biol Chem 276:30078–30084.
Stewart JJ, Fischback JA, Chen X, Stargell LA. 2006. Non optimal TATA element exhibits diverse mechanistic consequences. J Biol Chem 281:22665–22673.
Surrey S, Delgrosso K, Malladi P, Schwartz E. 1985. Functional analysis of a beta-globin gene containing a TATA box mutation from a Kurdish Jew with beta thalassemia. J Biol Chem 260:6507–6510.
Takihara Y, Nakamura T, Yamada H, Takagi Y, Fukumaki Y. 1986. A novel mutation in the TATA box in a Japanese patient with beta$^+$-thalassemia. Blood 67:547–550.
Thein SL. 2005. Pathophysiology of β-thalassemia—A guide to molecular therapies. Am Soc Hematol Educ Program 2005:31–37.
Thomas MC and Chiang CM. 2006. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41:105–178.
Tjian R. 1996. The biochemistry of transcription in eukaryotes: a paradigm for multi-subunit regulatory complexes. Philos Trans R Soc Lond B Biol Sci 351:491–499.
Verrijzer CP, Tjian R. 1996. TAFs mediate transcriptional activation and promoter selectivity. Trends Biochem Sci 21:338–42.
Wang Y, Kato N, Hoshida Y, Yoshida H, Taniguchi H, Goto T, Moriyama M, Otsuka M, Shiratori Y, Ito Y, Omata M. 2003 Interleukin-1beta gene polymorphisms associated with hepatocellular carcinoma in hepatitis C virus infection. Hepatology 37:66–71.
Watanabe M, Zingg BC, Mohrenweiser HW, 1996. Molecular analysis of a series of alleles in humans with reduced activity at the triosephosphate isomerase locus. Am J Hum Genet 58:308–316.
Wobbe CR, Struhl K. 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Mol Cell Biol 10:3859–3867.
Zienolddiny S, Ryberg D, Maggini V, Skaug V, Canzian F, Haugen A. 2004. Polymorphisms of the interleukin-1 beta gene are associated with increased risk of non-small cell lung cancer. Cancer 109:353–356.