Blood Cells, Molecules and Diseases 69 (2018) 1–9

Contents lists available at ScienceDirect

Blood Cells, Molecules and Diseases

journal homepage: www.elsevier.com/locate/bcmd

A long noncoding RNA from the *HBS1L-MYB* intergenic region on chr6q23 regulates human fetal hemoglobin expression

Tasha A. Morrison[a], Ibifiri Wilcox[a], Hong-Yuan Luo[a], John J. Farrell[b], Ryo Kurita[c], Yukio Nakamura[d], George J. Murphy[a], Shuaiying Cui[a], Martin H. Steinberg[a,e], David H.K. Chui[a,e,*]

[a] Department of Medicine, Section of Hematology-Oncology, Boston University School of Medicine, Boston, MA 02118, USA
[b] Department of Medicine, Section of Biomedical Genetics, Boston University School of Medicine, Boston, MA 02118, USA
[c] Research and Development Department, Central Blood Institute, Blood Service Headquarters, Japanese Red Cross Society, Tokyo, Japan
[d] Cell Engineering Division, RIKEN BioResource Center, Ibaraki, Japan
[e] Department of Pathology and Laboratory Medicine, Boston University School of Medicine, Boston, MA 02118, USA

ARTICLE INFO

Keywords:
Long noncoding RNA
HbF quantitative trait loci
Regulation of HbF expression

ABSTRACT

The *HBS1L-MYB* intergenic region (chr6q23) regulates erythroid cell proliferation, maturation, and fetal hemoglobin (HbF) expression. An enhancer element within this locus, highlighted by a 3-bp deletion polymorphism (rs66650371), is known to interact with the promoter of the neighboring gene, *MYB*, to increase its expression, thereby regulating HbF production. RNA polymerase II binding and a 50-bp transcript from this enhancer region reported in ENCODE datasets suggested the presence of a long noncoding RNA (lncRNA). We characterized a novel 1283 bp transcript (*HMI-LNCRNA;* chr6:135,096,362–135,097,644; hg38) that was transcribed from the enhancer region of *MYB*. Within erythroid cells, *HMI-LNCRNA* was almost exclusively present in nucleus, and was much less abundant than the mRNA for *MYB*. *HMI-LNCRNA* expression was significantly higher in erythroblasts derived from cultured adult peripheral blood $CD34^+$ cells which expressed more *HBB*, compared to erythroblasts from cultured cord blood $CD34^+$ cells which expressed much more *HBG*. Down-regulation of *HMI-LNCRNA* in HUDEP-2 cells, which expressed mostly *HBB*, significantly upregulated *HBG* expression both at the mRNA (200-fold) and protein levels, and promoted erythroid maturation. No change was found in the expression of *BCL11A* and other key transcription factors known to modulate *HBG* expression. *HMI-LNCRNA* plays an important role in regulating *HBG* expression, and its downregulation can result in a significant increase in HbF. *HMI-LNCRNA* might be a potential therapeutic target for HbF induction treatment in sickle cell disease and β-thalassemia.

1. Introduction

Sickle cell disease (SCD) and β-thalassemia major are the most prevalent hemoglobinopathies worldwide. Once found primarily in regions of the world where malaria was and may still be endemic, these diseases are now widespread due to human migration and are increasingly important to global health. Fetal hemoglobin (HbF; $\alpha_2\gamma_2$) can inhibit polymerization of deoxy-sickle hemoglobin, and can also compensate for the lack of adult hemoglobin (HbA; $\alpha_2\beta_2$) in β-thalassemia major. Therefore, HbF is the major modifier of disease severity for both SCD and β-thalassemia major [1,2].

Genome-wide association studies (GWAS) have found multiple single nucleotide polymorphisms (SNPs) marking three major quantitative trait loci (QTL) associated with HbF levels — chr11p15 (*HBB* gene cluster), chr2p16 (*BCL11A*) and chr6q23 (*HBS1L-MYB* intergenic polymorphisms or HMIP) [3]. Together they account for 20–45% of HbF variance in different populations. In addition, other cis-acting elements such as the *HBD-HBBP1* intergenic region and transcription factors including MYB, KLF1, BCL11A, ZBTB7A, CHD4, NR2C1/NR2C2 and KDM1α, play important roles in regulating *HBG* expression [4–7]. Nevertheless, significant gaps of knowledge on the regulation of *HBG* still remain.

The 126 kb *HBS1L-MYB* intergenic region on chr6q23 is between the genes *HBS1L* which is a member of the GTP-binding elongation factor family with no known association with erythroid-specific traits, and *MYB* which encodes for the transcription factor c-MYB. c-MYB regulates proliferation and maturation of erythroid cells, and modulates gene expression within the *HBB* gene cluster [8,9]. A distal enhancer

* Corresponding author at: Department of Medicine, Boston University School of Medicine, 72 East Concord Street, Evans 248, Boston, MA 02118, USA.
E-mail address: david.chui@bmc.org (D.H.K. Chui).

https://doi.org/10.1016/j.bcmd.2017.11.003
Received 12 October 2017; Received in revised form 27 November 2017; Accepted 28 November 2017
Available online 29 November 2017
1079-9796/

located at ~84 kb upstream of *MYB* has been shown by GWAS, insertional mutagenesis, long-range interaction demonstrable by chromosome conformation capture (3C) analysis, and gene editing with Cas9 nucleases [10–13]. This enhancer encompasses a 3-bp deletion polymorphism (rs66650371), which is surrounded by binding sites for erythroid-specific transcription factors TAL1/E47, GATA, RUNX1, LDB1 and KLF1, and is likely the functional motif to account for most of the effect upon HbF level by this QTL [10,12,13]. Alteration of this enhancer by polymorphisms such as rs66650371 reduced its interaction with the *MYB* promoter, which led to downregulation of *MYB* and upregulation of *HBG* expression. Furthermore, ENCODE datasets annotated RNA polymerase II occupancy and a 50-bp RNA transcript adjacent to rs66650371. This led us to hypothesize that this transcript is part of a long noncoding RNA (lncRNA) [10]. LncRNAs are usually > 200 nucleotides long, are transcribed throughout the genome, and have broad functionality.

We now report the characterization of a novel 1283 bp lncRNA, herein named the *HBS1L-MYB* intergenic long noncoding RNA (*HMI-LNCRNA*). *HMI-LNCRNA* is transcribed from the enhancer for *MYB*, and its downregulation significantly increased *HBG* expression at both the mRNA and protein levels in human adult-like erythroid cells. These observations suggest that *HMI-LNCRNA* has an important role in silencing *HBG* expression in adults, and could become a therapeutic target for increasing HbF in patients with SCD and β-thalassemia major.

2. Materials and methods

2.1. K562 cells

K562 cells were cultured at 37 °C in RPMI medium containing 10% FBS and 2% penicillin/streptomycin.

2.2. RNA extraction

Total RNA was extracted using RNeasy Mini Kit (Qiagen), treated with DNase (RNase-Free DNase Set, Qiagen), followed by RNA cleanup using RNeasy Mini Kit. For tissue-specificity experiment, multiple human organ RNA panels (Invitrogen and Clontech) were also treated with DNase, followed by RNA cleanup.

2.3. Reverse transcription polymerase chain reaction RT-PCR

cDNA was synthesized from total RNA using SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). PCR reactions were done using the Multiplex PCR kit (Qiagen). The following primers were used to amplify the 1180 bp product: 5′-ATCGCTCATGAGAAATG TGG-3′ (forward) and 5′-GGAACCGCCCTGATAACATT-3′ (reverse).

2.4. Rapid amplification of cDNA ends (RACE)

5′- and 3′-RACE were done using the FirstChoice RLM-RACE Kit (Ambion), following the manufacturer's instructions, using SuperTaq Plus Polymerase (Life Technologies) for PCR reactions. The following gene-specific primers were used: 5′-GTCTAATGGTGTGGCTCACAAA-3′ (5′-outer), 5′-CCCCAGCTTCCTTATCTGTAAA-3′ (5′-inner), 5′-TTCACT CTGGACAGCAGATGTT-3′ (3′-outer) and 5′-CGGTTCCCTCAGAAGACA CTTA-3′ (3′-inner). RACE PCR products were ligated to pCRII vector using TA Cloning Dual Promoter Kit (Invitrogen), transformed into One Shot INVαF chemically competent *E. coli* (Invitrogen), and grown on LB plates containing ampicillin and X-Gal. Insert-positive white colonies were picked and grown for DNA extraction. PCR to amplify insert (Forward: 5′-TGTGGAATTGTGAGCGGATA-3′ and Reverse: 5′-GTTTTC CCAGTCACGACGTT-3′), and DNA sequencing were done to determine the 5′- and 3′-ends.

2.5. DNA sequencing

PCR products were purified using AccuPrep PCR Purification Kit, and prepared for sequencing using ABI Big Dye Terminator v3.1 Cycle Sequencing Kit. Sequence data was analyzed on FinchTV version 1.5.0. NCBI BLAST was used to determine length and location of sequence.

2.6. Human umbilical cord blood-derived erythroid progenitor (HUDEP) cells

HUDEP cells are immortalized erythroid cell lines derived from cord blood $CD34^+$ mononuclear cells [14]. HUDEP-1 and HUDEP-2 cells were maintained in expansion medium—StemSpan SFEM medium (StemCell Technologies) supplemented with SCF (50 ng/ml, Invitrogen), EPO (3 U/ml, Invitrogen), dexamethasone (1 μM, Sigma), doxycycline (1 μg/ml, Clontech), L-glutamine (1%, Life Technologies) and penicillin/streptomycin (2%, Life Technologies). For erythroid maturation, cells were cultured in differentiation medium—IMDM medium (Invitrogen) supplemented with heat inactivated human serum from human male AB plasma (5%, Sigma), EPO (3 U/ml, Invitrogen), insulin (10 μg/ml, Sigma), doxycycline (1 μg/ml, Clontech), holotransferrin (500 μg/ml, Sigma), heparin (3 U/ml, Sigma), SCF (100 ng/ml, Invitrogen), L-glutamine (1%, Life Technologies) and penicillin/streptomycin (2%, Life Technologies)—for 5 days. For further erythroid maturation, doxycycline was removed and cells were cultured for 2 more days.

2.7. Primary $CD34^+$ mononuclear cells

Primary $CD34^+$ mononuclear cells derived from cord blood and peripheral blood (StemCell Technologies) were expanded for up to 7 days in StemSpan SFEM II medium (StemCell Technologies) supplemented with StemSpan CC100 (1 ×, StemCell Technologies) and penicillin/streptomycin (2%, Life Technologies). To induce erythroid differentiation, cells were cultured for up to 10 days in StemSpan SFEM II medium (StemCell Technologies) supplemented with SCF (10 ng/ml, Invitrogen), EPO (5 U/ml, Invitrogen), IL-6 (10 ng/ml, Sigma) and penicillin/streptomycin (2%, Life Technologies).

2.8. Nuclear and cytoplasmic fractionation

Nuclear/Cytosol Fractionation Kit (BioVision) was used following manufacturer's instructions.

2.9. Quantitative PCR (qPCR)

RNA was used for qPCR using TaqMan RNA-to-C_T 1-Step Kit (Applied Biosystems) and the following TaqMan gene expression assays (Applied Biosystems): *HBG1/2* (Hs00361131_g1), *HBB* (Hs00758889_s1), *MYB* (Hs00920556_ m1), *HBS1L* (Hs04188641_g1), *HMI-LNCRNA* (custom TaqMan assay designed by Applied Biosystems to target genome position chr6: 135,096,354–135,097,644, hg38; assay ID number AJI1MTQ), *BCL11A* (Hs01093197_m1), *CHD4* (Hs00172349_m1), *KLF1* (Hs0061 0592_m1), *ZBTB7A* (Hs00252415_s1), *NR2C1* (Hs00915957_m1), *NR2C2* (Hs00991824_m1), *KDM1a* (Hs01002741_m1) and *ACTB* (Hs01060665_g1). QPCR reactions were run on a StepOne Plus qPCR machine (Applied Biosystems). *ACTB* was used as the endogenous control.

2.10. Western blot analysis

Cell pellets were suspended in Roche lysis buffer (protease inhibitor, 0.3% NP40, 10% glycerine, 2 mM EDTA, 246 mM NaCl, 10% phosphatase inhibitor, PBS and water), placed on ice for 1 h and centrifuged at 14,500 rpm for 15 min at 4 °C to extract protein. Standard methodology was used for Western blot analysis. The following antibodies were used: c-MYB (ab109127, Abcam), HBB (16216-1-AP, Proteintech),