

## Characterization of a Transcript Found Within the *HBS1L-MYB* Intergenic Region

Tasha A. Morrison, Ibi Wilcox, Hong-Yuan Luo, Martin H. Steinberg and David H.K. Chui.  
Department of Medicine, Boston University School of Medicine, Boston, MA

Fetal hemoglobin (HbF,  $\alpha_2\gamma_2$ ) is the main hemoglobin in the fetus, but its levels decline after birth to <1% in adults. However, there are individuals who continue to have high levels of HbF well into adulthood. High HbF in sickle cell anemia is often associated with milder disease hence the effort to understand more about its regulation for possible therapeutic gain. Some of the variation in HbF among patients with sickle cell anemia can be accounted for by differences in three major HbF quantitative trait loci (QTL) that were found in genome-wide association studies to be associated with high HbF. However, depending on the population studied, these QTLs only account for up to 10-50% of HbF variance. Therefore, determining other factors for high HbF levels is critical in understanding the regulation of HbF expression and finding other possible therapeutic targets to induce HbF production in patients with hemoglobinopathies.

Previously, we found that HbF variability is associated with a 3-bp deletion within the *HBS1L-MYB* intergenic region on chr6q23 (Farrell et al, *Blood* 2011). Functional analyses of the region around the 3-bp deletion using ChIP assay in K562 cells, which are heterozygous for the 3-bp deletion, found that four transcription factors that are involved in erythroid cell differentiation—GATA-2, TAL1, E47 and RUNX1—all bind in the region of the 3-bp deletion. ENCODE datasets show that RNA polymerase II and a 50 bp RNA transcript are found in this same area. This work has been confirmed and extended by Stadhouders et al (*JCI*, 2014), who found that within the *HBS1L-MYB* intergenic region is an enhancer region, which includes the 3-bp deleted region and erythroid-specific transcription binding sites that functions to increase expression of *MYB*, a gene that encodes the transcription factor c-Myb, which is known to regulate erythroid differentiation and hemoglobin expression. Based on these observations, we set out to determine the extent of the 50-bp transcript found by ENCODE, and whether this transcript could possibly be a long noncoding RNA (lncRNA) that functions to regulate *MYB* expression and its downstream targets.

Using cDNA from K562 cells and the 50-bp transcript as a guide, we were able to amplify by PCR a fragment greater than 1 kb, which includes the 3-bp region and the binding sites for the erythroid-specific transcription factors mentioned above. Sequencing of this PCR product confirmed that the transcript is transcribed from the *HBS1L-MYB* intergenic region. Next, 5'- and 3'-rapid amplification of cDNA ends (RACE) reactions were done to determine the full length of this transcript. Sequencing results of the 3' RACE product show that the 3'-end of the transcript is more than 100 bp downstream from what was originally found by PCR. Results for the 5'-end are pending. These results were also confirmed in primary erythroid cells derived from cord blood CD34+ cells. It is known that lncRNAs display differential expression during development and tend to be tissue-specific. Therefore, we looked at expression of this transcript during differentiation of primary erythroid progenitors cells, and found that its expression modulates during this process. We also found that this transcript is tissue-specific—expressed mostly in tissue of hematopoietic lineage. Further analyses are underway to determine its function.

This novel transcript within the *HBS1L-MYB* intergenic region is a candidate for being a lncRNA. Since this fragment is transcribed from a region known to be associated with HbF variability and regulate *MYB* expression, it will be of interest to determine its role as it relates to erythroid differentiation and hemoglobin expression.

