Blood First Edition Paper, prepublished online June 28, 2017; DOI 10.1182/blood-2017-02-767400

Article Title

KLF1 drives the expression of fetal hemoglobin in British HPFH

Short Title for the Running Head

KLF1 drives fetal hemoglobin in British HPFH

Authors and Affiliations

Beeke Wienert,[1] Gabriella E. Martyn,[1] Ryo Kurita,[2] Yukio Nakamura,[3] Kate G.R. Quinlan,[1] Merlin Crossley[1]

[1]School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Australia; [2]Research and Development Department, Central Blood Institute, Blood Service Headquarters, Japanese Red Cross Society, Koto-ku, Tokyo, Japan; and [3]Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki, Japan

Corresponding Author

Merlin Crossley
School of Biotechnology and Biomolecular Sciences, University of New South Wales, NSW, 2052, Australia
m.crossley@unsw.edu.au
Tel: +61 2 9385 7916; Fax: +61 2 9385 7920

Word Count for Text: 1394

Word Count for Abstract: 191

Figure/Table Count: 2

Reference Count: 27

Appropriate Scientific Category: RED CELLS, IRON, AND ERYTHROPOIESIS

Key Points

1. Introduction of the British HPFH mutation into the fetal globin promoter in a human cell model causes elevated fetal globin expression.

2. The British HPFH mutation creates a *de novo* binding site both *in vitro* and *in vivo* for the potent erythroid activator KLF1.

Abstract

β-hemoglobinopathies are amongst the most common single locus inherited diseases. In this condition high fetal hemoglobin (HbF) levels have been found to be beneficial and boosting fetal hemoglobin expression is seen as an attractive therapy. Naturally occurring mutations in the fetal globin promoter can result in high HbF persisting into adulthood in a benign condition known as Hereditary Persistence of Fetal Hemoglobin (HPFH). Individuals with one form of HPFH, British HPFH, carry a T to C substitution at position -198 of the fetal *globin* gene promoter. These individuals exhibit HbF levels of up to 20%, enough to ameliorate the symptoms of β-hemoglobinopathies. Here we use CRISPR-mediated genome editing to introduce the -198 substitution into human erythroid HUDEP-2 cells and show that this mutation is sufficient to substantially elevate expression of HbF. We also examined the molecular mechanism underlying the increase in fetal globin expression. Through a combination of *in vitro* and *in vivo* studies we demonstrate that the mutation creates a *de novo* binding site for the important erythroid gene activator Krüppel-like Factor 1 (KLF1/EKLF). Our results indicate that introducing this single naturally occurring mutation leads to significantly boosted HbF levels.

Introduction

Fetal hemoglobin (HbF) levels are typically downregulated postnatally to ~1% of total hemoglobin. However, in some individuals higher levels of HbF persist throughout adulthood. This benign condition, Hereditary Persistence of Fetal Hemoglobin (HPFH), has been associated with mutations in regulatory regions of fetal *γ-globin*. Several HPFH variants have been identified, often in families that presented to the clinic as a result of other disease-causing hemoglobin mutations. Family members who were HPFH carriers tended to experience milder symptoms of disease and thus high levels of HbF are considered beneficial[1].

A -198T>C mutation in the promoter of the *γ-globin* gene was first described in 1974 in a large British family[2,3] and is known as British-type HPFH (Figure 1A and Supplemental Figure 1). Carriers have HbF levels of ~5-20%. Previously it was shown that the mutation creates a sequence resembling an SP1 binding site and that the activator SP1 could bind

this site *in vitro*[4]. Accordingly it was proposed that the ubiquitous SP1 transcription factor is responsible for driving high HbF levels in British HPFH individuals[4,5]. It was therefore unexpected that subsequent transactivation assays showed that elevated activity of the mutant promoter is erythroid-specific[4,5] suggesting that an erythroid-specific transcription factor is involved in the over-expression of HbF. Given these results, we reinvestigated and found that the erythroid transcription factor KLF1[6–8] binds to and drives the expression of the -198T>C promoter.

Study design

In vitro binding assays were performed as previously described[9].

HUDEP-2 cells[10] were used as a model to study the effects of -198T>C HPFH. Cells were transfected with CRISPR/Cas9 plasmid px458[11] encoding a guide RNA targeting the *γ-globin* promoter and a single-stranded oligodeoxynucleotide containing the -198T>C mutation. Transfected cells were enriched by FACS and clonal populations were established.

γ-globin expression was determined by qRT-PCR using primers to amplify the region of interest and expression normalized to 18S rRNA. Flow cytometry was performed on permeabilized cells stained with an APC-conjugated HbF antibody (MHFH05).

Chromatin Immunoprecipitation (ChIP) was performed as previously described[12] using an antibody against KLF1 (TA305808) and control IgG (sc-2028).

Hemoglobin HPLC was performed as previously described[13].

Chromatin-conformation-capture assays were performed as previously described[12].

See Supplemental Methods for more details.

Results and discussion

SP/KLF transcription factor family members have DNA-binding domains composed of three Cys_2His_2-type zinc fingers that recognize CNCCC-motifs where N can be any base but mostly is G or A[14]. The DNA-binding domains of all members of the SP/KLF transcription factor family are virtually identical in the three residues in each zinc finger that make primary contacts to DNA (Figure 1B). Hence, it is not surprising that the binding sites of different SP/KLF family members are highly similar as determined by Chromatin Immunoprecipitation followed by Sequencing (ChIP-Seq) experiments[15–18] (Figure 1C). Accordingly, we reasoned that, as SP1 has been shown to bind *in vitro*[4,5], any of the members of the SP/KLF family could potentially interact with the -198T>C *γ-globin* promoter.

To test this prediction *in vitro* we performed Electrophoretic Mobility Shift Assays (EMSA) with erythroid KLF (KLF1/EKLF, Figure 1D) and 12 other KLF proteins (Supplemental Figure 2), and assessed binding to the WT and -198T>C *γ-globin* promoters. All tested KLFs preferentially bound the mutant *γ-globin* promoter. Next we investigated the abundance of KLF and SP-proteins in erythroid tissues and cell lines, since the -198T>C mutation activates *γ-globin* expression erythroid-specifically[4,5]. We found that KLF1 ~~and KLF3 are~~ is the most abundantly expressed in differentiated human $CD34^+$ bone marrow cells[19] and erythroblast HUDEP-2 cells[10] making it a likely candidate to be accountable for elevated HbF levels in British HPFH (Supplemental Figure 3).

We then moved to engineer HUDEP-2 cells to stably carry the mutation. HUDEP-2 cells are an attractive recently-described model for primary adult erythroblasts, given their expression patterns of globins[10] and of the appropriate KLFs/SP proteins (Supplemental Figure 3). We introduced targeted substitution mutations into the *γ-globin* locus using CRISPR/Cas9 genome editing and homology-directed repair to establish clonal cell populations carrying the -198T>C mutation (Supplemental Figure 4).

First we examined mRNA expression levels of *β-like globins* by RT-PCR (Figure 2A). Clonal WT HUDEP-2 cells express ~0.5-1% *γ-globin* mRNA [γ/γ+β]. Introducing the -198T>C mutation into a single fetal *globin* promoter elevated the percentage of *γ-globin* mRNA to 4-

6%. Accordingly we also observed that the percentage of HbF-immunostaining cells ("F-cells") increased in cells carrying the -198T>C mutation (Figure 2B).

After confirming that the -198T>C mutation reproduces an HPFH phenotype in HUDEP-2 cells, we investigated the binding of transcription factors to the -198T>C *γ-globin* promoter. We first sought to generate homozygous HUDEP-2 -198T>C cells. As it proved difficult to reproducibly target all four copies of the *γ-globin* genes, we introduced an intermediate step and first engineered a parental cell line carrying only one *γ-globin* gene and promoter on each allele (HUDEP-2($\Delta^G\gamma$)WT, Supplemental Figure 5). We then introduced the -198T>C mutation into these cells homozygously, applying the same strategy as outlined previously (HUDEP-2($\Delta^G\gamma$)-198T>C). We found that *γ-globin* mRNA levels were strongly elevated compared to the parental cells and that the amount of F-cells increased substantially in clonal HUDEP-2($\Delta^G\gamma$)-198T>C cells (Figure 2C and D). We also performed High-performance liquid chromatography (HPLC) for hemoglobin. In HUDEP-2($\Delta^G\gamma$)WT cells only adult hemoglobin was detected whereas HbF made up 52-77% in cells with -198T>C (Figure 2E). Western Blots confirmed this finding and HbF repressor levels (BCL11A/ZBTB7A) were unchanged (Supplemental Figure 6).

We then performed KLF1 ChIP and saw consistently higher enrichment of the *γ-globin* promoter in HUDEP-2($\Delta^G\gamma$)-198T>C cells compared to the WT (Figure 2F). This effect was augmented when we performed the KLF1 ChIP in five day differentiated HUDEP-2($\Delta^G\gamma$)WT and -198T>C cells (Figure 2G). To ensure that the antibody is specific for KLF1 we also carried out ChIP experiments across a range of loci using KLF1 knockout and WT HUDEP2 clones. In KLF1 knockout cells we observed only background binding, equivalent to that obtained with control IgG, thereby validating the specificity of the antibody (Supplemental Figure 7). Interestingly we also observed that binding of KLF1 to the *β-globin* promoter was significantly reduced in HUDEP-2($\Delta^G\gamma$)-198T>C cells suggesting that KLF1 may switch from activating *β-globin* in WT[7,20] to driving *γ-globin* expression in -198T>C cells. Hence, we then looked at interaction frequencies of the locus-control-region (LCR) with the globin promoters by chromosome-conformation-capture (3C). We observed that the LCR indeed interacts more frequently with the *γ-globin* promoter and less often with the *β-globin* promoter in HUDEP-2($\Delta^G\gamma$)-198T>C compared to WT cells (Figure 2H).

These findings support our hypothesis that the -198T>C mutation creates a functional binding site for the erythroid activator KLF1 and that *de novo* binding of KLF1 contributes to the upregulation of *γ-globin* expression in British HPFH by shifting the LCR towards the fetal globin promoters (Figure 2I). This is consistent with the commonly accepted model of the LCR interacting competitively with the globin promoters. Although present in appropriate levels, BCL11A and ZBTB7A clearly do not silence the HbF expression at the -198T>C promoter suggesting that *de novo* binding of KLF1 overrides the developmentally determined gene expression program of globins. This is analogous to what we[12] and others[21] have seen when other transcription factors that induce chromatin-looping are targeted to the *γ-globin* promoter.

Several approaches have been proposed to reactivate the expression of HbF in adult life including targeting the HbF repressor BCL11A[22] and its regulatory elements[23,24] and forced chromatin looping mediated by an artificial transcription factor[21]. If the efficiency of targeting is sufficient[18] then introducing naturally occurring mutations associated with HPFH could represent an alternative approach.

Beyond the potential therapeutic aspects, our work is of interest as it shows that the -198T>C mutation generates a *de novo* binding site. An important insight gained from genome-wide association studies is that cis-regulatory mutations are commonly associated with genetic disease, and are as abundant as coding variants[25]. Mutations will typically either disrupt a transcription factor binding site or create a new site. Since a range of deletions, insertions or substitutions can disrupt elements, but only precise changes can create new elements one might expect disruptions are more common than *de novo* sites. However, the first two HPFH mutations we studied, -198T>C and the previously reported -175T>C[12], create *de novo* sites for erythroid-specific transactivators. Similarly, a *de novo* GATA1 site can profoundly affect the expression of genes within the *α-globin* locus[26]. It may be that disruptions are less prevalent than expected because of redundancy observed between multiple *cis*-acting elements[27] but *de novo* sites are in fact more common than initially anticipated.

Acknowledgements

The CRISPR/Cas9 genome editing plasmid px458 was a gift from Feng Zhang, Broad Institute, Cambridge MA (Addgene plasmid # 48138). We would like to thank Laura Norton, Alister Funnell and Richard Pearson, of UNSW Sydney, Australia for valuable discussions and reagents. The authors would also like to acknowledge Helene Lebhar and the UNSW Recombinant Products Facility for assistance in the analysis of hemoglobin variants. This work was supported by funding from the Australian National Health and Medical Research Council to MC (APP1098391). BW was supported by a University International Postgraduate Award. GEM was supported by an Australian Postgraduate Award.

Authorship

Contribution: B.W. performed most of the experiments. G.E.M. performed EMSA assay with KLF1. R.K. and Y.N. provided HUDEP-2 cells. K.G.Q. and M.C. supervised the study. B.W., K.G.Q. and M.C. designed experiments and wrote the manuscript. All authors have read and approved the contents of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

References

1. Powars DR, Weiss JN, Chan LS, Schroeder WA. Is there a threshold level of fetal hemoglobin that ameliorates morbidity in sickle cell anemia? *Blood*. 1984;63(4):921–926.

2. Weatherall DJ, Cartner R, Clegg JB, et al. A form of hereditary persistence of fetal haemoglobin characterized by uneven cellular distribution of haemoglobin F and the production of haemoglobins A and A2 in homozygotes. *Br J Haematol*. 1975;29(2):205–220.

3. Wood WG, MacRae IA, Darbre PD, Clegg JB, Weatherall DJ. The British type of non-deletion HPFH: characterization of developmental changes in vivo and erythroid growth in vitro. *Br J Haematol*. 1982;50(3):401–414.

4. Fischer KD, Nowock J. The T----C substitution at -198 of the A gamma-globin gene associated with the British form of HPFH generates overlapping recognition sites for two DNA-binding proteins. *Nucleic Acids Res*. 1990;18(19):5685–5693.

5. Ronchi A, Nicolis S, Santoro C, Ottolenghi S. Increased Sp1 binding mediates erythroid-specific overexpression of a mutated (HPFH) gamma-globulin promoter. *Nucleic Acids Res*. 1989;17(24):10231–10241.

6. Miller IJ, Bieker JJ. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol Cell Biol*. 1993;13(5):2776–2786.

7. Perkins AC, Sharpe AH, Orkin SH. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*. 1995;375(6529):318–322.

8. Perkins A, Xu X, Higgs DR, et al. Krüppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood*. 2016;127(15):1856–1862.

9. Crossley M, Whitelaw E, Perkins A, et al. Isolation and characterization of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. *Mol Cell Biol*. 1996;16(4):1695–1705.

10. Kurita R, Suda N, Sudo K, et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS ONE*. 2013;8(3):e59890.

11. Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281–2308.

12. Wienert B, Funnell AP, Norton LJ, et al. Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nat Commun*. 2015;6:7085.

13. DeWitt MA, Magis W, Bray NL, et al. Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med*. 2016;8(360):360ra134.

14. Wierstra I. Sp1: emerging roles--beyond constitutive activation of TATA-less housekeeping genes. *Biochem Biophys Res Commun*. 2008;372(1):1–13.

15. Tallack MR, Whitington T, Yuen WS, et al. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res*. 2010;20(8):1052–1063.

16. Burdach J, Funnell AP, Mak KS, et al. Regions outside the DNA-binding domain are critical for proper in vivo specificity of an archetypal zinc finger transcription factor. *Nucleic Acids Res*. 2014;42(1):276–289.

17. Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008;133(6):1106–1117.

18. Chong HK, Biesinger J, Seo YK, Xie X, Osborne TF. Genome-wide analysis of hepatic LRH-1 reveals a promoter binding preference and suggests a role in regulating genes of lipid metabolism in concert with FXR. *BMC Genomics*. 2012;13:51.

19. Norton LJ, Funnell APW, Burdach J, et al. KLF1 directly activates expression of the novel fetal globin repressor, ZBTB7A/LRF, in erythroid cells. *Blood Advances*. 2017; 1:685-692.

20. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*. 1995;375(6529):316–318.

21. Deng W, Rupon JW, Krivega I, et al. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*. 2014;158(4):849–860.

22. Xu J, Peng C, Sankaran VG, et al. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science*. 2011;334(6058):993–996.

23. Bauer DE, Kamran SC, Lessard S, et al. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science*. 2013;342(6155):253–257.

24. Canver MC, Smith EC, Sher F, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015;527(7577):192–197.

25. Deplancke B, Alpern D, Gardeux V. The genetics of transcription factor DNA binding variation. *Cell*. 2016;166(3):538–554.

26. De Gobbi M, Viprakasit V, Hughes JR, et al. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science*. 2006;312(5777):1215–1217.

27. Hay D, Hughes JR, Babbs C, et al. Genetic dissection of the α-globin super-enhancer in vivo. *Nat Genet*. 2016;48(8):895–903.

Figure legends

Figure 1: KLF1 binds to and activates the -198T>C *γ-globin* promoter *in vitro*. (A) The organization of the human *β-globin* locus (chromosome 11). The *β-like globin* genes are indicated by black boxes with their conventional gene symbols indicated above. The locus control region (LCR) is represented as a grey rectangle, with the DNase I hypersensitive sites (HS) within the LCR being represented by black boxes (HS1-4). The British HPFH (-198T>C) mutation in the promoter of the *γ-globin* gene is also depicted. (B) Aligned protein sequences of the three zinc finger DNA-binding domains of all members of the SP/KLF family of transcription factors. Highlighted are residues -1, +3 and +6 of each zinc finger which are responsible for making contact to DNA. (C) *In vivo* binding motifs of transcription factors SP1[18], KLF1[15], KLF3[16] and KLF4[17] as determined by ChIP-Seq. (D) EMSA showing that KLF1 can bind *in vitro* to a -198T>C probe but fails to bind to a WT probe containing the -198 region of the *γ-globin* promoter. Lanes 1-3 contain the WT probe for the -198 site (-209 to -187 bp) and lanes 4-6 contain the HPFH -198T>C mutant probe. Lanes 1 and 4 contain nuclear extracts from COS cells transfected with a pcDNA3 empty vector. Lanes 2-3 and 5-6 contain nuclear extracts from COS cells overexpressing KLF1. Binding of KLF1 to the -198T>C HPFH mutant probe can be observed in lane 5, with a supershift of KLF1 with an anti-KLF1 antibody in lane 6.

Figure 2: KLF1 binds to and activates the -198T>C *γ-globin* promoter in HUDEP-2 cells. (A) mRNA expression levels of *γ–globin* displayed as a percentage of *γ–* and *β-globin* [γ/γ+β]. mRNA levels were determined by qPCR and Ct values were normalized to rRNA levels of 18S. Primer efficiencies were confirmed to be equivalent. Shown are clonal HUDEP-2 WT cells (n=4) and two clones (n=2 for #15 and n=3 for #17) each carrying one allele with the -198T>C mutation. *γ-globin* mRNA levels are significantly higher in cells carrying the -198T>C mutation (**p<0.01). (B) Flow cytometry analysis of fetal hemoglobin levels in clonal WT HUDEP-2 cells (n=3) and a clonal HUDEP-2 cell population that is heterozygous for the -198T>C mutation (n=3). Representatively shown is the median of three experiments (±SD). Cells were permeabilized and then stained with APC-conjugated fetal hemoglobin antibody. The amount of cells expressing high HbF ("F-cells") was determined by flow cytometry. The same gating strategy was applied for WT HUDEP-2 cells and -198T>C HUDEP-2 cells using FlowJo software. (C) mRNA expression levels of *γ–globin* displayed as a percentage of *γ–* and *β-globin* [γ/γ+β]. mRNA levels were determined by qPCR and Ct values were normalized to rRNA levels of 18S. Shown are parental HUDEP-2($\Delta^G\gamma$)WT cells and HUDEP-2($\Delta^G\gamma$) cells homozygous for the -198T>C mutation. *γ-globin* mRNA levels are significantly higher in cells carrying the -198T>C mutation (***p<0.0001) (D) Flow cytometry analysis of fetal hemoglobin levels in HUDEP-2($\Delta^G\gamma$)WT and HUDEP-2($\Delta^G\gamma$)-198T>C cells. Representatively shown is the median of three experiments (±SD). Cells were permeabilized and then stained with APC-conjugated fetal

hemoglobin antibody. The amount of cells expressing high HbF ("F-cells") was determined by flow cytometry. (E) HPLC traces depicting hemoglobin production in HUDEP-2($\Delta^G\gamma$)WT and three clonal ($\Delta^G\gamma$)-198T>C cell populations. Highlighted in red is the peak for fetal hemoglobin. Percentages are HbF over total hemoglobin (HbF and HbA0). (F) ChIP-qPCR analysis of the relative enrichment of KLF1 at various genomic loci in HUDEP-2($\Delta^G\gamma$)WT (n=2) and -198T>C cells (n=4). The tested genomic loci were the *γ–globin* promoter, the *β–globin* promoter and the *VEGFA* promoter (-ctrl). The *SP1* promoter served as a positive control (+ctrl) for successful pulldown with the respective antibody and all values were normalized to enrichment at the *SP1* promoter. Enrichment of the *γ–globin* promoter after KLF1 pulldown is significantly higher in HUDEP-2($\Delta^G\gamma$)-198T>C cells whereas enrichment of the *β–globin* promoter is significantly lower (**p<0.01). (G) ChIP-qPCR analysis of the relative enrichment of KLF1 at genomic loci in five day differentiated HUDEP-2($\Delta^G\gamma$)WT (n=3) and -198T>C cells (n=3). Again, KLF1 enrichment is significantly higher (*p<0.05) at the *γ–globin* promoter but lower (**p<0.01) at the *β–globin* promoter. (H) 3C assay measuring locus-wide crosslinking frequencies in HUDEP-2($\Delta^G\gamma$)WT (n=3) and three clonal -198T>C cell populations. A schematic of the human β-globin locus is shown on top of the graph. The x-axis indicates distances in kb from the *ε-globin* gene. Vertical lines represent HindIII restriction sites. The dark grey bar denotes the anchor HindIII fragment containing hypersensitive site 3. Beige bars denote analyzed HindIII fragments. Shown is mean±s.e.m. (I) Proposed model of molecular mechanism in British HPFH. In WT adult erythroid cells KLF1 drives the expression of adult *β-globin* via the promoter (top panel). The presence of the -198T>C mutation (bottom panel) allows binding of KLF1 to the fetal *γ-globin* promoter and leads to upregulation of *γ-globin* expression through recruitment of the LCR. In these cells binding of KLF1 to the *β-globin* promoter and *β-globin* transcript levels are reduced.

Figures

Figure 1

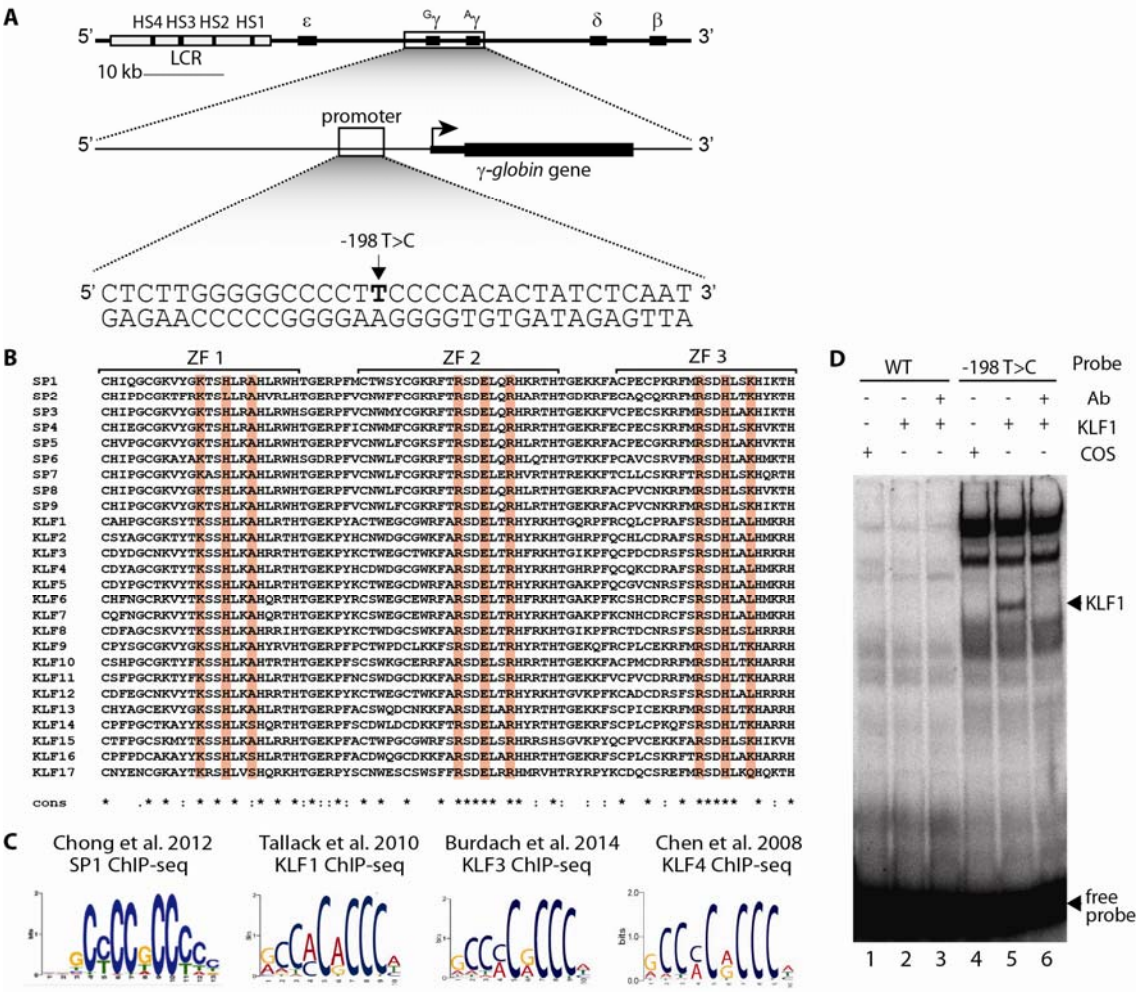
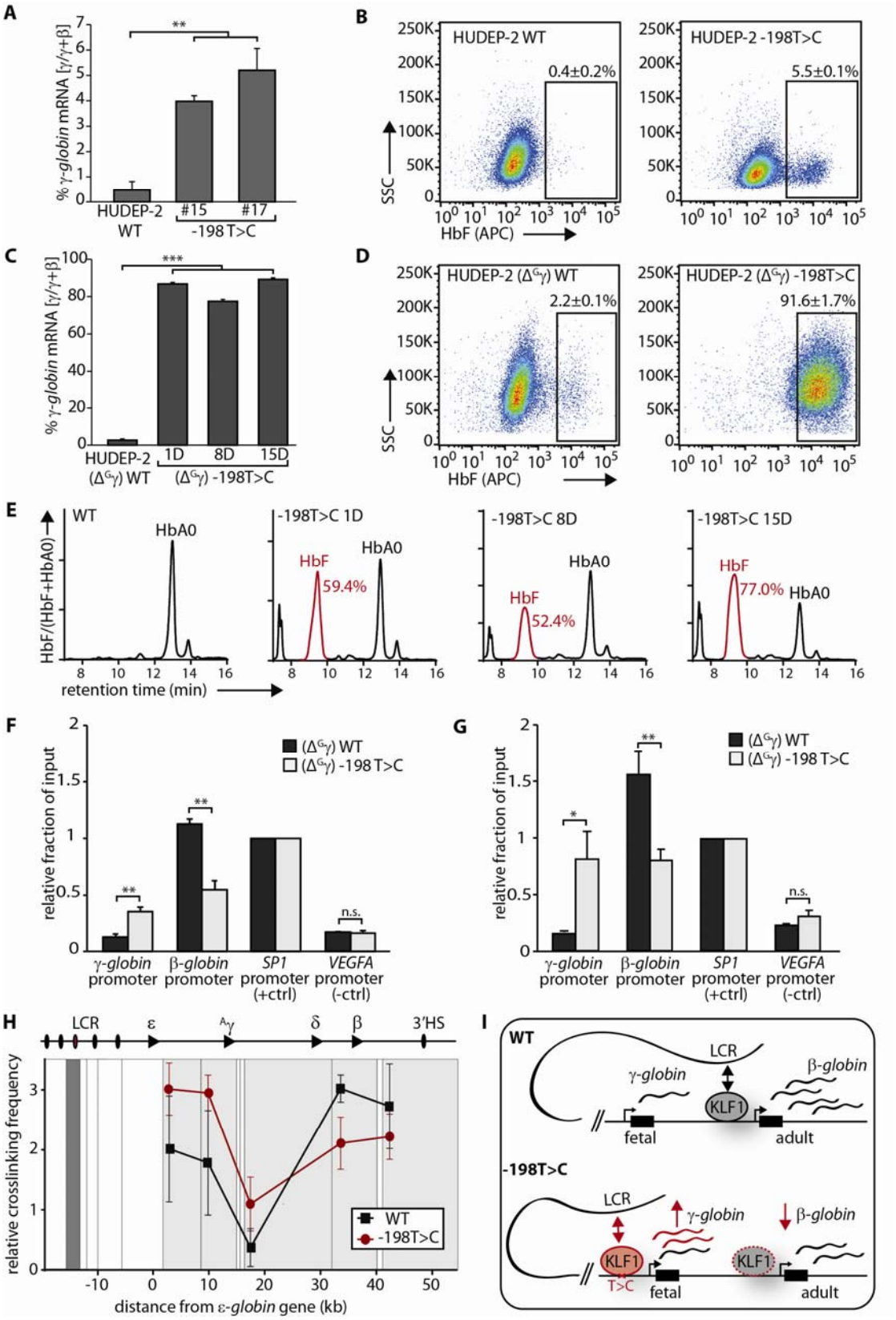


Figure 2



Prepublished online June 28, 2017;
doi:10.1182/blood-2017-02-767400

KLF1 drives the expression of fetal hemoglobin in British HPFH

Beeke Wienert, Gabriella E. Martyn, Ryo Kurita, Yukio Nakamura, Kate G.R. Quinlan and Merlin Crossley

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.