JOURNAL OF UNDERGRADUATE RESEARCH

Mechanisms of Pol II Recruitment to the Human Beta-Globin Locus Control Region Super-Enhancer

Breanna Johnson, Matthew Gibbons, Fabia Bushra, Yu Fang, Jorg Bungert

College of Liberal Arts and Sciences, University of Florida

Jorg Bungert, Department of Biochemistry and Molecular Biology

Abstract

The human β-globin gene locus consists of five genes that encode subunits of hemoglobin. The locus control region (LCR) regulates these genes, acting as a super-enhancer to promote high-levelβ-globin expression in erythroid cells. The LCR, like other super-enhancers, is known to recruit transcription complexes and generate non-coding enhancer RNA (eRNA). Past research has suggested that super-enhancers form phase-separated transcription initiation domains that concentrate RNA polymerase II (Pol II) and transfer Pol II to target gene promoters during transient looping interactions. The aim of this project is to examine the role of involved transcription initiation (TFIIH) in eRNA production and Pol II transfer in the globin gene locus. The LCR was immobilized on magnetic beads and incubated with erythroid nuclear protein extracts. Proteins not bound were washed away and bound proteins were subjected to western blotting experiments. CUT & RUN experiments were performed to examine the localization of TFIIH in the globin gene locus in intact cells. The data obtained suggests that TFIIH is recruited to the LCR to aide in the generation of eRNA. In the future, the laboratory plans to deplete TFIIH and examine production of eRNA and Pol II recruitment at the LCR.

Keywords: TFIIH, LCR, eRNA, Transcription, Pol II, Super-Enhancer

Introduction

The human β-globin gene locus is a sequence of five genes linearly arranged on a chromosome. The gene is expressed in erythroid cells, and helps to define the structure of the polypeptide chains found in adult hemoglobin (U.S. National Library of Medicine, n.d.). Mutations in the β-globin gene have been known to cause genetic disorders known as β-haemoglobinopathies, which include sickle cell disease and β-thalassemia (Reinisch *et al*., 2016). These conditions can lead to abnormally formed hemoglobin proteins or a decreased production of hemoglobin within the body, which in turn can lead to a lessened perfusion of oxygen throughout the body, leading to conditions such as life-threatening anemia, failure to thrive in children, and blockages in the flow of blood causing pain crises (Medline Plus, n.d., n.d.). Additionally, the β-globin gene locus has been shown to serve as a very effective model for the regulatory systems that are active during cell differentiation and development (Zhou *et al.*,

2010). Thus, investigation of the β-globin gene allows scientists a better understanding of β-haemoglobinopathic conditions and eventually leads to novel treatment development. Additionally, the study of the gene provides an insightful look into regulatory mechanisms during the processes of cell development. For these reasons, the regulation of the β-globin gene has become a subject of intense study.

The human β-globin gene is regulated by a super-enhancer known as the locus control region, or LCR (Gurumurthy *et al.*, 2021). The LCR has been generally considered necessary for high-level expression of globin genes to occur during the development of erythroid cells, and its deletion has been shown to significantly reduce the expression of β-globin genes in both mice and humans (Bender *et al.*, 2000; Reik *et al.*, 1998). Additionally, the LCR has been found to contain many sites that are very susceptible to digestion to the enzyme DNase I, known as DNase I-hypersensitive sites (HS Sites) (Leach et al., 2001).

In order to commence the transcription of RNA, RNA polymerase II (Pol II) is loaded onto DNA by the pre-initiation complex (PIC) (Patel *et al.*, 2018). The transcription factor that commences involved transcription initiation, Transcription factor II H (TFIIH), is a component of the PIC, and is what opens the DNA template and thus allows for transcription to occur (Rimel & Taatjes, 2018). Thus, the association of TFIIH to the PIC allows for transcription to be functional, allowing for the potential production of eRNA if the PIC forms at an enhancer region (Gibbons et al., 2022). In addition, it has been observed that TFIIH has been recruited at the basal promoter but not in upstream activating sequences in yeast (Gibbons et al., 2022). This potentially implies that TFIIH association to the basal promoter is linked to the recruitment of Pol II at basal promoter elements via TFIID/TFIIB-mediated recruitment. Pol II has been known to be recruited to HS sites on the LCR both *in vitro* and *in vivo*, which implies that transcription complexes first form at the LCR before they are eventually transferred to globin gene promoters (Vieria *et al.*, 2004; Levings *et al.*, 2002). Thus, a model has been proposed suggesting that the LCR serves as a reservoir for transcription complexes, which are transferred to the globin genes during transient contacts.

To determine the role of TFIIH in the production of eRNA and the transfer of Pol II in the globin gene locus, the LCR was immobilized on magnetic beads before being incubated with erythroid nuclear protein extracts. Any unbound proteins and proteins bound to the beads were

subjected to western blotting experiments and were subsequently visualized in order to determine whether or not TFIIH binds to the LCR and β-globin gene in both undifferentiated and differentiated cells. Finally, in order to examine the genome-wide association of TFIIH in intact cells, CUT & RUN experiments were performed (see below).

Methods

Formation of DNA Constructs

pLCR and pβAX plasmids were first linearized using XhoI. Afterwards, a phenol extraction was performed in order to purify the DNA and remove unwanted proteins. For precipitation, a tenth of a volume of 3M sodium acetate (pH 5.8) was added to the pβAX and LCR digestion products, and afterwards 2.5 volumes of 100% ethanol was added. A fill-in reaction was then performed in order to incorporate biotinylated dGTP, biodUTP, biodCTP, biodATP and Klenow polymerase on the 5' overhangs of the linearized pLCR and pβAX DNA, and thus allow the DNA fragments to attach to streptavidin coated magnetic beads. Afterwards, the post fill-in reaction samples were digested with NotI enzyme in order to isolate the LCR and β-globin fragments. An agarose gel was run of the NotI digested products, and a gel extraction was then performed on the resulting DNA bands using the QIAquick Gel Extraction Kit in order to isolate the LCR and β-globin genes that were labeled at one end with biotin.

Immobilization of DNA Constructs

Using a kilobase binder kit (Dynal), the samples of isolated LCR and β-globin genes obtained from the gel extraction step were then attached to magnetic beads coated with streptavidin at room temperature for 2 hours using a binding buffer provided by the kit. An aliquot of the beads was then taken, digested with EcoRI, and analyzed under a 1% agarose gel in order to determine the binding efficiency of the LCR/β-globin (βAX) constructs onto the beads. The DNA concentrations were then measured using an Implen Nanophotometer.

Binding reaction

Protein extracts were prepared from MEL (murine erythroleukemia) cells as described by Leach et al. Two separate cultures of 10^8 undifferentiated and differentiated MEL cells were pelleted via centrifuge and washed in phosphate buffered saline (13.7 mM NaCl, 0.27 mM KCl, 0.43 mM Na2HPO4, 7H2), 0.14 mM KH2PO4), respectively. The washed cells were then

resuspended in 1 mL lysis buffer (20 mM HEPES [pH 7.6], 20% glycerol, 10 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 0.1% Triton X-100, 1 mM DTT, 1 mM PMSF), before being incubated on ice for 3 minutes and subsequently centrifuged at 2,000 rpm for 10 minutes at a temperature of 4 C. The resultant pellets were then rocked in 1 mL nuclear extraction buffer (20 mM HEPES [pH 7.6], 20% glycerol, 400 mM NaCl, 1.5 ml MgCl2, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF) for 1 hour at 4 C, before being centrifuged at 10,000 rpm for 10 minutes at 4 C. The supernatant was then dialyzed in dialysis buffer (20 mM HEPES [pH 7.8], 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF) for 16 hours in order to produce MEL protein extract. Final protein concentration was then measured from the resultant extracts, with the expected final protein concentration being between 4 and 6 mg/ml

After being attached to the magnetic beads, the DNA was then incubated with 50 µl MEL nuclear protein extract, either prepared from undifferentiated MEL cells (2µg/µl) or prepared from differentiated MEL cells (grown in the presence of 2% DMSO, 3µg/µl), 25µl 2X binding buffer (336 mM HEPES, pH 7.9, 160 mM KCl, 40 mM MgCl2, 4 mM dithiothreitol, 10 mM phenylmethylsulphonyl fluoride, and 20% glycerol), and water (to a total volume of 100 µl) for 30 minutes at 30^0C on a roller. All unbound material (supernatant) was then removed using a Dynal magnet and remaining beads were resuspended in buffer A (10 mM HEPES [pH 7.6], 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM EGTA [pH 8.0], 10% glycerol, 10 mM glycerolphosphate, 1 mM DTT, 0.2 mM PMSF). In some reactions, 500ng of non-biotinylated LCR (pLCR) or β-globin plasmids (βA/X) were included and served as competitor DNA.

Western Blotting

10 µl of the beads and supernatant samples were combined with 10 µl of SDS-PAGE Sample Loading Buffer (Biosciences), heated for 5 min at 90^0C, placed on ice for 5 min, and loaded onto a 4-20% Mini-PROTEAN gel (BIO-RAD). The pre-stained protein standard (Precision Plus Protein "Kaleidoscope") was run in one of the lanes. The gel was run in 1x Running buffer (3g Tris base, 14.3 g Glycine, 1 g SDS, pH 8.3) for about 45 min at room temperature at 110V. After electrophoresis, the proteins in the gel were transferred to a nylon membrane using the Transblot Turbo system (BIORAD) for 40 min. The membranes were then washed with water and then TBST Buffer (3.63 g Tris base, 8.77 g NaCl, 1 L ddH_20, adjust to pH 7.6 with HCl). The

membrane was then rocked at room temperature in 5% milk in TBST buffer for 1 hour, and the membrane was incubated with two 1:5000 dilutions of RNA Pol II (ab5408 mouse mAb Abcam 1 mg/mL) and ERCC (10818-1-AP Rabbit PolyAb Proteintech 30 µg/mL) antibodies. ECL solution A and B were then spread on the membranes and the membranes were then subjected to fluorescence imaging using the Amersham Typhoon system.

CUT & RUN

To determine where Pol II and TFIIH interacts with DNA in intact erythroid cells, the Cleavage Under Targets & Release Using Nuclease (CUT&RUN) technique was utilized. 4,000,000 cells were propagated for use in the Cut and Run by growing K562 cells in DMEM, 10% FFBS and 1% Pen/Strep. These cells were then harvested and washed, before being loaded onto nickel beads at a rate of 500,000 cells per reaction and permeabilized with 0.01% Digitonin solution. A positive and negative control reaction were performed in duplicate by using 0.5 µg of H3K4me3 and IgG antibodies for the positive and negative control reactions respectively. Both the positive and negative controls were spiked-in with SNAP-CUTANA K-MetStat Panel.

The same TFIIH antibodies described for the transfer experiment was used for the CUT & RUN experiment. For the Pol II antibody, a rabbit Ab from Bethyl laboratories was used (A300-653A-T). The remaining reactions were performed in duplicate with 0.5 µg ERCC2 (TFIIH) antibody and 0.5 µg Pol II antibody. All reactions were incubated overnight at 4^0C, and afterwards pAG-MNase was added to each reaction. 100 mM CaCl was then added in order to initiate the DNA cleavage activity of pAG-MNase, and all reactions were incubated at 4^0C for 2 hours. The reactions were then stopped, and afterwards spike-in E. Coli DNA was added to each sample. DNA was then purified and quantified using the QuBit fluorometer. A subsequent next-generation sequencing (NGS) of the cleaved fragments was then used to identify the binding sites of TFIIH and RNA pol II.

Results

***In vitro* association of TFIIH with the LCR and the β-globin gene**

The concentrations of βAX and LCR following the gel extraction step on the bands from the NotI digested products were 6.48 mg/µL and 1.363 mg/µL, respectively. The data from the

western blot show that TFIIH is present mostly in the supernatant (Figure 1A) but can be detected at the LCR and β-globin beads (Figure 1B). RNA Pol II was detectable in all of the supernatant samples analyzed (Figure 1C). Somewhat unexpectedly, the interaction of TFIIH with the β-globin gene on the beads was weaker and disappeared in the absence of the LCR competitor. After the RNA Pol II and ERCC2 antibodies were run on a western blot and analyzed using an SDS-PAGE, a band of 250 kDa and 87 kDa respectively was detected for the Pol II and ERCC2 antibodies, respectively (Figure 1D).

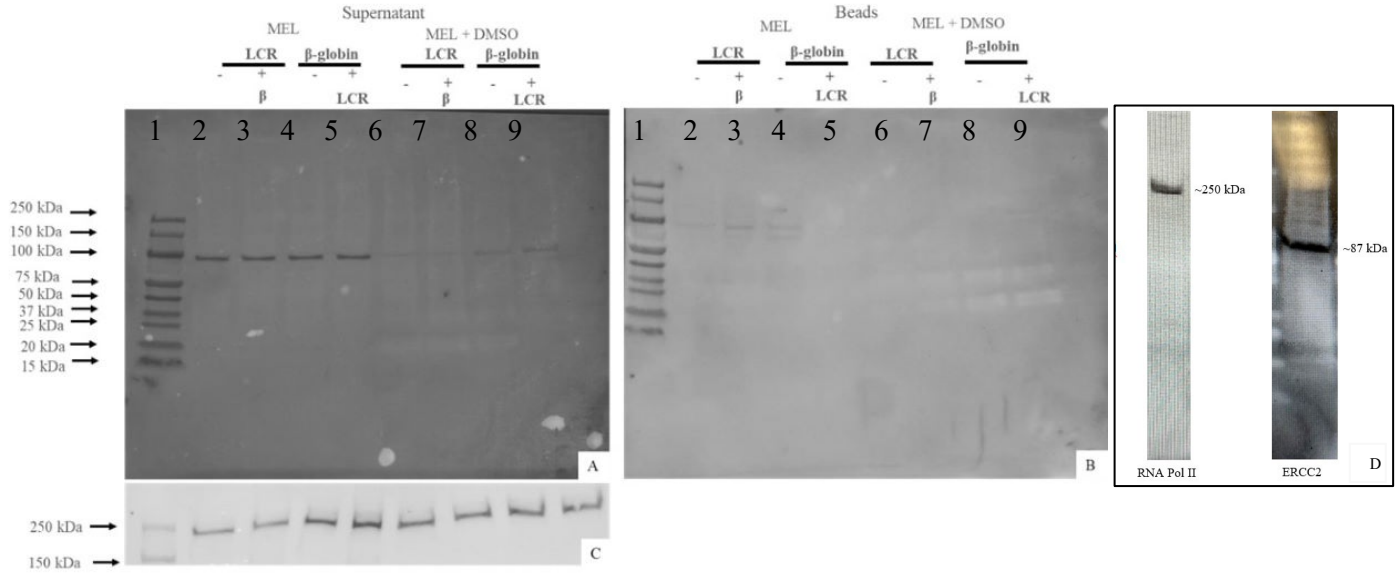


Figure 1A: Presence of TFIIH in the supernatant. The 1st lane contains a protein standard (Kaleidoscope ladder.) Lanes 2 and 3 contain LCR incubated with MEL extracts in the absence (lane 2) or presence (lane 3) of beta-globin competitor. Lanes 4 and 5 contain beta-globin incubated with MEL extracts in the absence (lane 4) or presence (lane 5) of LCR competitor. Lanes 6-7 contain LCR incubated with MEL + DMSO in the absence (lane 6) and presence (lane 7) of beta-globin competitor. Lanes 8-9 contain beta-globin incubated with MEL + DMSO in the absence (lane 8) and presence (lane 9) of LCR competitor. Figure 1B: Presence of TFIIH in the beads. The lanes are loaded in the same manner as in Figure 1B. Figure 1C depicts the levels of RNA pol II found in the supernatant. Figure 1D depicts RNA Pol II (Left) and TFIIH antibodies (right) blotted onto whole protein lysate from K562 cells.

CUT & RUN

The CUT & RUN experiment was performed as described in the Methods section. Transcriptionally active promoters contain nucleosomes that are tri-methylated at histone H3 lysine (K) 4 (H3K4me3). Antibodies specific for H3K4me3 served as a positive control in the experiments. Only one of the H3K4me3 reactions provided enough DNA for the sequencing reactions. The table in figure 2A summarizes the data. It also shows the combination of bar-codes that were spiked-in in the different samples. Figure 3 depicts the results of the subsequent H3K4me3 CUT & RUN reaction.

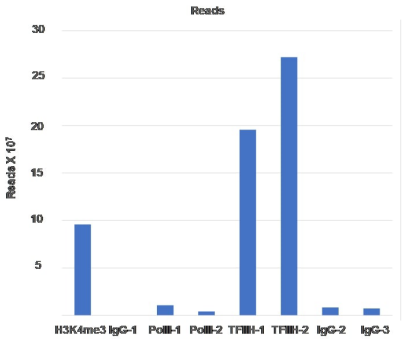
Figure 2: A. Table summarizing the data from the Illumina sequencing reaction. B. Graph depicting the number of sequencing reads for the different antibodies used in the experiment.

The graph in figure 2B depicts the number of reads pers sample. As can be seen, the H3K4me3, TFIIH-1 and TFIIH-2 samples yielded relatively high sequencing reads. Low sequencing reads for the negative control IgG antibodies was expected. Low sequencing reads for the Pol II antibody was unexpected and may suggest that the Pol II epitope is inefficiently recognized by the antibody in the context of the cellular transcription complexes. The data are currently being analyzed for visualizing the binding peaks for H3K4me3, TFHH-H and Pol II.

A

Sample	Reads	Pct of lane	Pct of proj	Index
H3K4Me3	95816338	3.19%	16.14%	CGCTCATT+AGGCTATA
IgG-3131	743	0.00%	0.00%	CGCTCATT+GCCTCTAT
Pol-II-1	10782167	0.36%	1.82%	CGCTCATT+AGGATAGG
Pol-II-2	4136819	0.14%	0.70%	CGCTCATT+TCAGAGCC
TFIIH-1	195504072	6.50%	32.92%	CGCTCATT+CTTCGCCT
TFIIH-2	271939475	9.04%	45.79%	CGCTCATT+TAAGATTA
IgG-1-2938	8327740	0.28%	1.40%	ATTACTCG+AGGATAGG
IgG-2-2938	7318151	0.24%	1.23%	ATTACTCG+TCAGAGCC

B



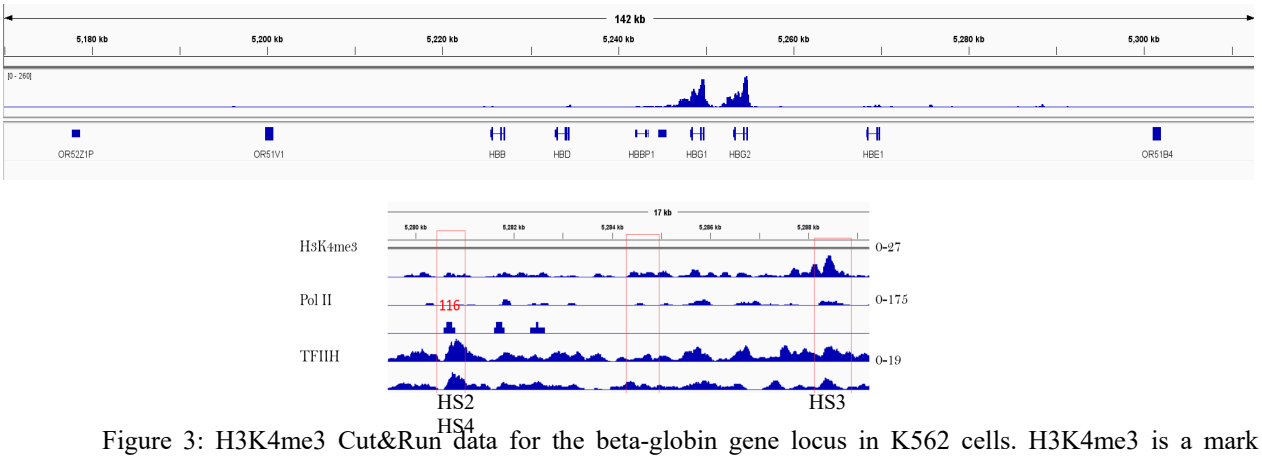


Figure 3: H3K4me3 Cut&Run data for the beta-globin gene locus in K562 cells. H3K4me3 is a mark for actively transcribed regions in the genome. The two gamma globin genes HBG1 and HBG2 are highly expressed in K562 cells. There is little H3K4me3 at the LCR (HS2 HS3 and HS4), with more in HS4. There is also TFIIH in HS2. TFIIH CUT & RUN shows background and Pol II CUT & RUN did not work, likely because the antibody used is not suitable for CUT & RUN.

Discussion

Among the LCR and βAX samples containing and lacking competitors for both the supernatant and bead agarose gels, all undifferentiated cells displayed a band of <100 kDa. This indicates that TFIIH can bind to both the β-globin gene and the LCR. Additionally, due to the fact that a less intense band was displayed in the β-globin lane containing the LCR competitor, it can be surmised the presence of the LCR gene can lead to less TFIIH binding to the βAX gene. Since previous studies have shown that the LCR hypersensitive sites are transcribed, the finding of TFIIH at both the globin gene and the LCR is not surprising (Vieira *et al.*, 2004). However, it should be noted that the binding reactions were carried out under conditions not optimized for transcription.

Additionally, although the Pol II levels remain constant in both differentiated (DMSO + MEL) and undifferentiated (MEL) cells, almost no TFIIH was shown to be present on the beads in the differentiated cells. This could be due to the fact that experiments were carried out under conditions not optimized for transcription. The β-globin gene is expressed more efficiently in differentiated cells. Experiments will have to be repeated under conditions allowing transcription initiation. Furthermore, it is possible that eRNA formation is more efficient in undifferentiated MEL cells and could be promoted by an activity that is lost when the cells differentiate into taking on erythroid-specific functions, hence the loss of TFIIH binding in the samples with added DMSO. It is also possible that there is simply not a high enough concentration of TFIIH found in the differentiated cells to allow for sufficient binding to the LCR and β-globin gene.
With respect to the CUT & RUN experiment, it was found that the gamma globin genes HBG1 and HBG2 were expressed at high levels in K562 cells. Additionally, H3K4me3 was found at the LCR as expected, although in low amounts, which indicates a small amount of transcription of the LCR. Meanwhile, the Pol II CUT & RUN did not work properly, implying that a different antibody will need to be used in order to later obtain results for Pol II.

Acknowledgements

I wish to thank Jorg Bungert for his mentorship and guidance throughout the project. In addition, I am grateful to Fabia Bushra for the technical assistance that she provided, and for the

assistance that Matt Gibbons and Yu Fang provided in running the CUT & RUN reaction. I would also like to thank Randall Johnson for his assistance in proofreading the final manuscript.

References

U.S. National Library of Medicine. (n.d.). HBB Hemoglobin Subunit Beta [homo sapiens (human)] - gene - NCBI. National Center for Biotechnology Information. Retrieved April 1, 2023, from https://www.ncbi.nlm.nih.gov/gene/3043?report=full_report

Beta thalassemia: MedlinePlus Genetics. (n.d.). https://medlineplus.gov/genetics/condition/beta-thalassemia/

National Library of Medicine. (n.d.). Sickle Cell Disease. Sickle Cell Anemia | MedlinePlus. https://medlineplus.gov/sicklecelldisease.html

Dever, D. P., Bak, R. O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C. E., Pavel-Dinu, M., Saxena, N., Wilkens, A. B., Mantri, S., Uchida, N., Hendel, A., Narla, A., Majeti, R., Weinberg, K. I., & Porteus, M. H. (2016). CRISPR/Cas9 β-globin Gene Targeting in Human Haematopoietic stem cells. *Nature, 539*(7629), 384–389. https://doi.org/10.1038/nature20134

Patel, A. B., Louder, R. K., Greber, B. J., Grünberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., & Nogales, E. (2018). Structure of Human TFIID and Mechanism of TBP Loading Onto Promoter DNA. *Science* (New York, N.Y.), *362*(6421), eaau8872. https://doi.org/10.1126/science.aau8872

Dean A. (2011). In the Loop: Long Range Chromatin Interactions and Gene Regulation. *Briefings in Functional Genomics, 10*(1), 3–10. https://doi.org/10.1093/bfgp/elq033

Gibbons, M. D., Fang, Y., Spicola, A. P., Linzer, N., Jones, S. M., Johnson, B. R., Li, L., Xie, M., & Bungert, J. (2022). Enhancer-Mediated Formation of Nuclear Transcription Initiation Domains. *International Journal of Molecular Sciences, 23*(16), 9290. https://doi.org/10.3390/ijms23169290

Gurumurthy, A., Yu, D. T., Stees, J. R., Chamales, P., Gavrilova, E., Wassel, P., Li, L., Stribling, D., Chen, J., Brackett, M., Ishov, A. M., Xie, M., & Bungert, J. (2021). Super-enhancer Mediated Regulation of Adult β-globin Gene Expression: the Role of eRNA and Integrator. *Nucleic Acids Research, 49*(3), 1383–1396. https://doi.org/10.1093/nar/gkab002

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., … & Young, R. A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. *Cell, 155*(4), 934-947.

Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E., & Groudine, M. (1998). The Locus Control Region is Necessary for Gene Expression in the Human Beta-globin Locus but not the Maintenance of an Open Chromatin Structure in Erythroid Cells. *Molecular and Cellular Biology, 18*(10), 5992–6000. https://doi.org/10.1128/MCB.18.10.5992

Bender, M. A., Bulger, M., Close, J., & Groudine, M. (2000). Beta-globin Gene Switching and DNase I Sensitivity of the Endogenous Beta-globin Locus in Mice do not Require the Locus Control Region. *Molecular Cell, 5*(2), 387–393. https://doi.org/10.1016/s1097-2765(00)80433-5

Leach, K. M., Nightingale, K., Igarashi, K., Levings, P. P., Engel, J. D., Becker, P. B., & Bungert, J. (2001). Reconstitution of Human β-globin Locus Control Region Hypersensitive Sites in the

Absence of Chromatin Assembly. *Molecular and Cellular Biology, 21*(8), 2629-2640. https://doi.org/10.1128/MCB.21.8.2629-2640.2001

Levings, P. P., & Bungert, J. (2002). The Human Beta-Globin Locus Control Region. *European Journal of Biochemistry, 269*(6), 1589–1599. https://doi.org/10.1046/j.1432-1327.2002.02797.x

Vieria, K. F., Levings, P. P., Hill, M. A., Crusselle, V. J., Kang, S. L., Engel, J. D., Bungert, J. (2004). Recruitment of Transcription Complexes to the β-Globin Gene Locus in Vivo and in Vitro. *Journal of Biological Chemistry, 279*(48). https://doi.org/10.1074/jbc.M408883200

Rimel, J. K., & Taatjes, D. J. (2018). The Essential and Multifunctional TFIIH complex. *Protein Science: a Publication of the Protein Society, 27*(6), 1018–1037. https://doi.org/10.1002/pro.3424

Zhou, Z., Li, X., Deng, C., Ney, P. A., Huang, S., & Bungert, J. (2010). USF and NF-E2 Cooperate to Regulate the Recruitment and Activity of RNA polymerase II in the Beta-Globin Gene Locus. *The Journal of Biological Chemistry, 285*(21), 15894–15905. https://doi.org/10.1074/jbc.M109.098376