lncRNA CRNDE204 regulates erythropoiesis and myelopoiesis by binding to PUS1

lncRNA *CRNDE204* regulates erythropoiesis and myelopoiesis by binding to PUS1

Jingwen Hu

A Thesis Submitted to the Faculty of

The Harvard Medical School

in Partial Fulfillment of the Requirements

for the Degree of Master of Medical Sciences in Immunology

Harvard University

Boston, Massachusetts.

May, 2021

Thesis Advisor: Dr. Carl Novina Jingwen Hu

lncRNA *CRNDE204* regulates erythropoiesis and myelopoiesis by binding to PUS1

Abstract

Hematopoiesis is a crucial process that produces all cells within the hematopoietic system. Mutations in genes that disrupt normal hematopoiesis can cause bone marrow failure disorders ranging from myelodysplastic syndromes (MDS) to leukemia. Recent studies have pointed out the association between ribosomopathy and hematopoietic failure, and hematopoietic deficiency is a common clinical manifestation of multiple ribosomopathies, including Shwachman Diamond Syndrome and Diamond-Blackfan anemia (DBA). Shwachman Diamond Syndrome (SDS) is a rare and inherited form of ribosomopathy that is caused by mutations in SBDS gene, and patients with SDS lack myeloid progenitor cells, which causes severe anemia and neutropenia, and have increased risk of developing MDS or acute myeloid leukemia in later stages of life. Although standard care for SDS and other forms of MDS can relieve the symptoms, patients usually require lifelong treatment and multiple hospital visits to monitor their current status. Until now, the only cure for SDS is hematopoietic stem cell transplantation, yet due to its high risk of side effects, this is not widely used for treating SDS patients and is only considered if the patient has severe symptoms or starting to develop malignancies. The absence of curative treatment of SDS suggests there is a need to understand the molecular mechanism of SDS pathogenesis.

The Novina Lab published the first map of SDS hematopoiesis at single cell resolution. Analysis of this SDS patient data identified a lncRNA first described in colorectal cancers, called *colorectal neoplasia differentially expressed* (*CRNDE*), as the most downregulated genes during SDS hematopoiesis and the expression of it is increased in acute myeloid leukemia. Our data points out the role of *CRNDE* in promoting normal erythropoiesis. *In vitro* experiments show that one

splice variant isoform of *CRNDE*, *CRNDE204*, can promote differentiation of erythrocytes and megakaryocytes in K562 cells. However, another isoform, *CRNDE201*, represses differentiation. Based on these observations, we performed RNA-seq and proteomics analysis and found that *CRNDE204* regulates key erythropoiesis and myelopoiesis genes and pathways. In addition, a yeast three-hybrid screen identified 24 proteins that could bind to *CRNDE204*, including PUS1 which is the key regulator of erythrocyte development and its mutation causes sideroblastic anemia. This is further validated by computational analysis and knock-down assays, and knock-down of PUS1 result in a decrease in both erythroid and megakaryocyte differentiation in K562 cells. Collectively, our findings pinpoint that *CRNDE204* regulates erythropoiesis and myelopoiesis by interacting with PUS1, which makes *CRNDE204* reintroduction as promising genetic therapy target for bone marrow failure disorders.

Table of Contents

Table of Figures

Acknowledgement

I would like to express my sincere gratitude to my Principal Investigator, Dr. Carl Novina for providing me with this invaluable opportunity to work in his lab and for the advices on conducting research as well as for my education and career. I would like to thank Dr. Leon Wert-Lamas and Dr. Kushani Shah for their work and support in RNA-seq computational analysis, yeast-three hybrid experiments, experimental design and troubleshooting guidance throughout my work towards my master's thesis. I would like to thank Dr. Huan Yang from Lodish Lab for providing us with the shRNA constructs for knockdown experiments. Lastly, I would like to thank all members of Novina Lab for providing technical support and troubleshooting advice throughout my experiments. This accomplishment would not be possible without the support of my lab mates.

I also wanted to thank the Master's Immunology program at Harvard Medical School. I would like to thank Dr. Shiv Pillai for mentorship and advice on my education and career. Also, I would like to thank Dr. Gavin Porter for helping me to understand the techniques used in immunology, and Selina Sarmiento for helping me with the administrative work throughout the program.

Lastly, I would like to thank my family members for their continuous support and encouragement during my master's studies. I would also like to thank everyone in Class of 2021 cohort for their support in the past two years.

"This work was conducted with support from Students in the Master of Medical Sciences in Immunology program of Harvard Medical School. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University and its affiliated academic health care centers."

Chapter 1: Background

1.1 Hematopoiesis

Overview of hematopoiesis:

Hematopoiesis is a crucial process of constant production of blood cells that starts in early embryonic development and continues to throughout life. In early human embryonic development, hematopoiesis can be divided into the primitive wave and the definitive wave. The primitive wave happens in blood islands of the yolk sac. Primitive hematopoiesis starts from a primitive progenitor cell called hemangioblast, which is the progenitor for endothelial cells and common myeloid progenitor cells[1]. During the first transient wave of primitive hematopoiesis, large nucleated erythrocytes are produced. This process is followed by a second transient wave of primitive hematopoiesis where cells from erythron-myeloid lineage and lymphoid lineages start to develop[2]. Previous studies had identified the presence of IL-7Rα expressing lympho-myeloid restricted progenitor cells happens before the definitive hematopoiesis stages, suggesting the development of adaptive immune system initiates at early stages of embryonic development[3]. Definitive hematopoiesis originates in the aorta-gonad-mesonephros (AGM) region and is also the place for the production and renewal of definitive HSC and progenitor cells before the whole hematopoietic system moved to liver and bone marrow[4]. With the help of lineage tracing, recent evidence shows megakaryocyte progenitors can emerge within few cell divisions after HSC labeling, followed by the presence of erythroid and myeloid progenitors which appears around 2 weeks after initial HSC labeling. Cells from the lymphoid lineages appears after 3 weeks from initial HSC labeling. This difference can depend on the evolutionary origin of the blood cells, life span, the function of cell in normal homeostasis and whether the cell is easily depleted under stress condition[5].

Regulation of erythropoiesis:

Adult erythropoiesis is a complex process that produces mature red blood cell from HSC. This process involves the engagement phase, differentiation phase and maturation phase[6]. During the engagement phase, pluripotent HSC differentiates into erythroid lineage-committed burst-forming unit erythroid cell (BFU-E), and then to colony-forming unit erythroid cell (CFU-E)[7]. Terminal erythroid differentiation is characterized by the differentiation from CFU-E to reticulocytes, where the cell synthesizes hemoglobin, condense its size, expulse organelles and become enucleated reticulocytes[6]. This process happens in erythroblastic islands, which consist of a specialized stromal macrophage surrounded by erythroid cells at different developing stages[8]. Several studies have shown that the central macrophage functions to provide cellular support in erythroid differentiation and proliferation, iron transfer in heme synthesis, and phagocytose the extruded nucleus from erythroblasts[6]. The final step of erythropoiesis involves the maturation of erythrocytes from reticulocytes, where the cell gains its biconcave shape and enters the bloodstream. Mature erythrocytes circulate in the bloodstream, but they gradually lose hemoglobin and shrink in size during their lifespan, and will be phagocytosed by liver/spleen macrophages in the end[9].

Erythropoiesis is regulated by a network of interactions between multiple transcription factors, growth factors, non-coding RNAs and chromatin modifications[10]. GATA family transcription factors, GATA1, GATA2 and GATA3 are crucial in regulating hematopoietic development[11]. GATA1 is considered as the master transcription factor in erythropoiesis and is highly expressed in erythroid and megakaryocytic lineage cells[12]. The expression of GATA1 is silenced in HSC and early progenitor stages epigenetically through methylation at its upstream

element, and its expression start to accumulate in common myeloid progenitors and reaches peak at proerythroblast stage[13]. Knock-out of GATA1 during embryonic development is lethal due to severe anemia caused by apoptosis of erythroid progenitor cells, and mutations in GATA1 can cause a range of hematological disorders including Diamond-Blackfan anemia and myeloproliferative neoplasms[14]. GATA2 is highly expressed in HSCs, hematopoietic progenitor cells and mast cells, and is important in regulating the fate and proliferation of progenitor cells[15]. However, the expression of GATA2 is downregulated in terminal differentiated erythrocytes and macrophages, and is dispensable for the terminal differentiation of these cells[16]. Evidence shows that GATA1 represses the expression of GATA2 by binding to GATA-switch sites and promotes the expression of GATA1 in erythroid lineage cells[17]. This GATA-switch mechanism is important in maintaining the commitment to erythroid lineages.

Erythroid lineage surface markers:

Expression of different surface markers can be used to identify the development stage of erythroid cells in flow cytometry studies. In particular, the expression of transferrin receptor (CD71) and glycophorin A (CD235a) is used to determine the maturation stage of erythroid cells: From the proerythroblast stage, the expression of CD71 gradually declines while the expression of CD235a gradually increases, and mature erythrocytes are low in CD71 and high in CD235a expression[18]. The expression of integrins CD41 and CD61 complex is used in flow cytometry studies to determine the maturation of megakaryocytes. CD41 and CD61 forms heterodimer complexes and is highly expressed on platelets as fibrinogen receptors in platelet aggregation[19].

1.2 Ribosomopathy

Overview of Ribosomopathy:

Ribosome is the central location for protein production. In human, the ribosome consists of a large subunit (60s) and small (40s) subunit, where each subunit is made up by ribosomal RNA (rRNA) and ribosomal proteins[20]. Ribosomal proteins function in rRNA processing and maintain ribosome assembly. Depletion of a ribosomal protein in the ribosome subunit can affect the production of other ribosomal proteins of the same subunit, disrupt the rRNA processing pathway, and cause an overall decrease in production of mature ribosomes and polysomes[21]. Mutations in any of the ribosome components – either in ribosomal proteins, ribosomal RNAs or ribosomal assembly proteins – can result in a heterogeneous group of diseases called ribosomopathy[22] (Figure 1).

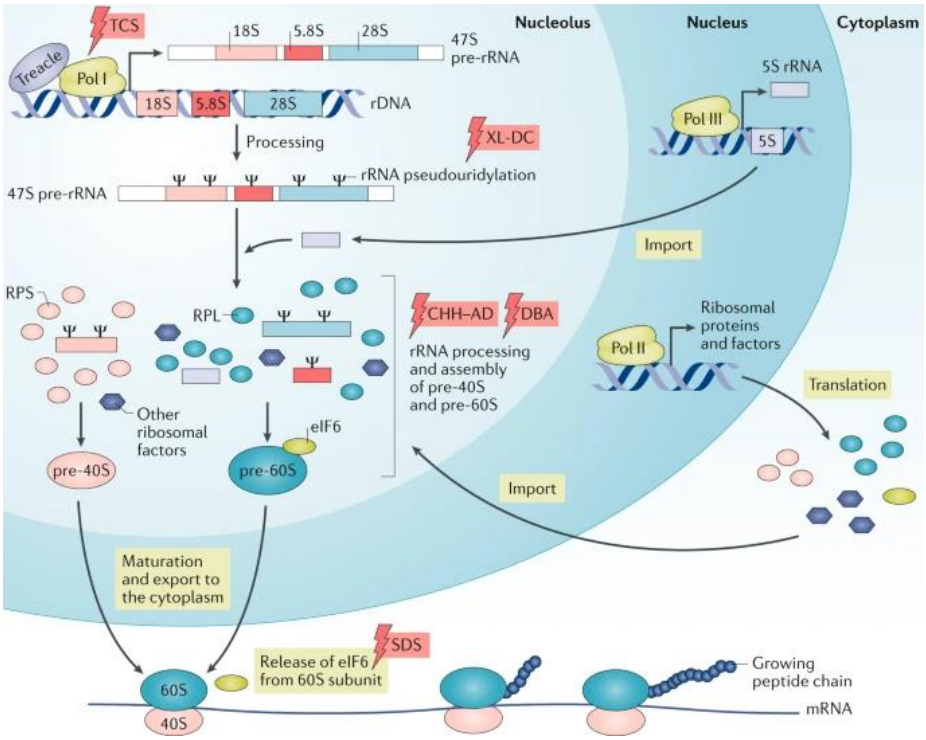


Figure 1: Ribosomopathies affect multiple steps in ribosome biosynthesis pathway. Adapted from Aspesi, A. & Ellis, S.R. Rare ribosomopathies: insights into mechanisms of cancer. Nat Rev Cancer **19**, 228-238 (2019)[23]

The most typical clinical symptoms of ribosomopathies include bone marrow failure, organ and developmental abnormalities and increased risk of developing cancers[24]. However, the reason why ribosomopathies tend to have defects in the blood lineages and why there is an increased incidence of cancers in patients with ribosomopathy is still unresolved. It seems paradoxical that a decrease in protein synthesis pathway could lead to cancers, where protein synthesis is activated. Recent studies have revealed the function of p53 in ribosomopathies. Ribosomopathies including Diamond-Blackfan anemia, Shwachman Diamond Syndrome (SDS) and Treacher Collins syndrome (TCS) have mutations in ribosomal proteins that can cause accumulation of p53 and activate the apoptosis pathway[23]. These observations suggest cell may lose the expression of p53 under selective pressure, and this will also cause p53 to lose its tumor suppressor function, thus increase the risk of cancer[25]. Another possible explanation to this paradox is the defects in ribosome components could produce a onco-ribosome that may alters the translation of specific proteins or create an unstable cellular environment that promotes cancer[23].

Shwachman Diamond Syndrome (SDS):

Shwachman Diamond Syndrome (SDS) is a rare inherited ribosomopathy characterized by bone marrow failure, exocrine pancreas dysfunction and skeletal abnormalities. Mutations associated with the *SBDS* gene, located at chromosome 7q11, are found in a majority of SDS patients, among which 183-184TA>CT and 258+2T>C mutations most common among all sites identified. The 183-184TA>CT mutation creates a premature stop codon and results in a truncated protein, while the 258+2T>C mutation and another 8bp deletion mutation disrupts the splice sites[26]. Research on *SBDS* gene has revealed its function in stabilizing mitotic spindle[27],

inducing FAS ligand during apoptosis[28], responding to cellular stress[29], and most importantly, regulating ribosome assembly process[30]. Other genes associated with ribosomal assembly, including DNAJC21, EFL1 and SRP54, are shown to be associated with SDS phenotype[31]. Based on the heterogenous nature of the symptoms and the complex molecular mechanisms affecting ribosome biosynthesis, finding a cure for SDS has been challenging. Current treatments for SDS aim to alleviate the symptoms: Blood transfusions and G-CSF are used to treat patients with anemia and neutropenia with recurrent infections. Pancreatic enzyme supplements are given to replenish the loss of pancreatic function. If patients show severe pancytopenia or signs of developing acute leukemia, hematopoietic stem cell transplantation is used and is the only curative treatment option[32].

1.3 Long non-coding RNA (lncRNA)

Overview of lncRNA:

Long non-coding RNA (lncRNA) are identified from RNA mapping as a novel class of RNA that exceed 200 nucleotides and do not code for functional proteins[33]. Although lncRNAs do not code for protein, they are heavily involved in many essential cellular processes, including transcriptional control of gene expression, post-transcriptional processing of nascent mRNAs, and epigenetic silencing of genes (Figure 2).

lncRNAs can recruit chromatin remodeling complexes and induce epigenetic chromosome silencing[34]. For example, the lncRNA *HOTAIR* recruits the Polycomb complex PRC2 by binding to the EZH2 subunit. This interaction is responsible for genome-wide trimethylation of H3K27[35]. Another well-studied lncRNA *Xist* is known to regulate X-

chromosome inactivation by recruiting SHARP-SMRT complex that activates the transcriptional silencer HDAC3 to induce the silenced state, and *Xist* maintains the silenced state by recruiting PRC2 through a HDAC3-dependent manner[36].

lncRNA can bind to RNA-binding proteins and recruit protein factors that regulates transcription[37]. Signal-induced lncRNAs can bind to the RNA-binding protein TLS, which binds to and inhibits the activity of CREB-binding proteins and the histone acetyltransferase p300 on cyclin D1 promoter and silence the expression of cyclin D1[38]. Moreover, lncRNAs are involved in post-transcriptional modification of proteins. Antisense lncRNA can bind to the 5'UTR containing the 5' splice site of *Zeb2* gene. This can prevent the binding of splice factors to the 5' splice site and allows ribosome to bind to the IRE element to regulates the translation of ZEB2 protein[39].

Furthermore, lncRNAs are also shown to modulate protein phosphorylation and activity[34]. For instance, the lncRNA *NKILA* binds to and forms a stable complex with the phosphorylation motifs of NF-κB/IκB and inhibits IKK-mediated phosphorylation of NF-κB[40]. The lncRNA *lnc-DC* is expressed exclusively in human dendritic cells and regulates the differentiation of dendritic cells. *lnc-DC* binds to STAT3 to promote its phosphorylation, and this binding prevents STAT3 from binding to SHP1 that results in dephosphorylation of STAT3[41].

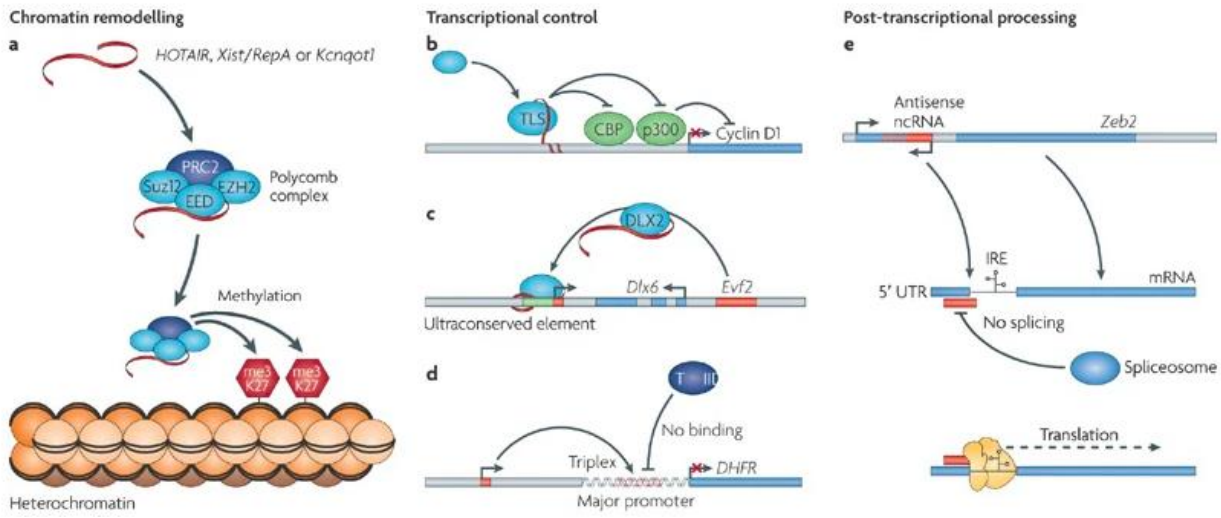


Figure 2: Function of lnRNAs in regulating gene expression. Adapted from Mercer, T.R., Dinger, M.E. & Mattick, J.S. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**, 155-159 (2009).[37]

lncRNA in cancers:

Growing evidence has shown that lncRNAs play important roles in human diseases, especially in cancers. In particular, the functions of lncRNAs *HOTAIR* and *MALAT1* are well-studied and have been used as prognostic markers in multiple tumors. The lncRNA *HOTAIR* is found to be upregulated in solid tumors including lung, breast and colorectal cancers, and also in hematological tumors. Cancers with upregulated expression of *HOTAIR* is associated with poor prognosis, metastasis and tumor cell proliferation, and there has been some therapies targeting *HOTAIR* in these cancers[42]. Like *HOTAIR*, the lncRNA *MALAT1* also interacts with the Polycomb repressive complex by binding to its subunit EZH2. In addition, *MALAT1* interacts with SR proteins to regulate pre-mRNA splicing[43]. The oncogenic role of *MALAT1* was first identified in early-stage non-small cell lung cancer patients and its overexpression predicts poor prognosis and metastasis[44]. *MALAT1* overexpression is also seen cancers including breast cancer, melanoma, renal cell carcinoma, and mantle cell lymphoma, in which the upregulation of *MALAT1* induces cancer cell proliferation and metastasis[45]. However, some recent studies have

shown *MALAT1* have tumor suppressive functions. In breast cancer and colorectal cancers, *MALAT1* expression is positively correlated with the expression of the tumor suppressor PTEN, and *MALAT1* regulate the expression of genes including EPCAM and ITGB4 that are involved in cell migration[46]. Due to the heterogenous nature of tumors, a lncRNA may have opposite function within the same type of cancer. A thorough analysis on the molecular pathways modulated by lncRNA would be necessary to understand the exact function of lncRNA in different cancers.

lncRNA and hematopoiesis:

lncRNAs play important roles in regulating normal hematopoiesis at different stages. Over 150 HSC-specific lncRNAs (lncHSC) were identified from mouse HSC and regulate differentiation both in vitro and in vivo by binding to hematopoietic transcription factor E2A[47]. The process of erythron-megakaryocytic development is dependent on lncRNA regulation. A study by Alvarez-Dominguez et al identified a total of 655 lncRNAs from mouse fetal liver during development, which are tissue specific and are dynamically regulated by transcription factors including GATA1, TAL1 and KLF1[48]. Knock-down of these lncRNAs affects erythrocyte maturation as a reduction of the percentage of enucleated cells after the knockdown. RNA-FISH shows distinct localization patterns of the lncRNAs and indicates differential functions of the lncRNAs in regulating erythropoiesis[49].

lncRNAs are also shown to regulate the development of lymphoid lineage cells. Brazao et al. have identified over 4,500 lncRNAs loci expressed in different stages in B cell development, and the expression of a group of the identified lncRNAs are regulated by the B cell lineage-specific transcription factor PAX5[50]. lncRNA *Morrbid* regulates cytotoxic T cell response to

viral infections and negatively regulate the AKT signaling in CD8+ T cells. The *Morrbid* locus also interacts with the pro-apoptotic factor BCL2L11 to regulate the effector cell fate[51]. lncRNAs including NeST, Th2 LCR and Flicr have roles in regulating the function of CD4+ T cells through epigenetic regulation[52]. Recent studies have shown that lncRNAs can contribute to resistance to immunotherapies by affecting the antigen presentation pathways, upregulating the expression of PD-1, modulating T cell activity and recruiting immunosuppressive cells to tumor sites[53].

On the other hand, lncRNAs contribute to the pathogenesis of bone marrow failure disorders and hematological malignancies. In vivo shRNA screening has identified lncRNAs that are required for the progress of acute myeloid leukemia by maintaining the stemness of the leukemic cells. In particular, lncRNAs *Pvt1* and *Lilam* are highly upregulated in AML cells and their expression is positively correlated with the protein level of MYC, the key transcription factor in AML pathogenesis[54].

lncRNA *CRNDE*:

The lncRNA *CRNDE* (colorectal neoplasia differentially expressed) is located on chromosome 16 and has 24 different splice variants annotated in *Ensembl*. As its name suggests, *CRNDE* was first identified to be upregulated in colorectal adenomas and adenocarcinomas[55]. Elevated expression of *CRNDE* is associated with cell proliferation, tumor progression, metastasis and poor prognosis in colorectal cancer patients[56]. Up to now, the isoform landscape for *CRNDE* is still unclear, and no standard nomenclatures of different *CRNDE* splice variants has been identified. However, different *CRNDE* isoforms has been identified in different forms of leukemias, and a specific splice variant of *CRNDE* has been found to be upregulated in

multiple hematological malignancies, including both acute and chronic myeloid leukemia, acute lymphoid leukemia and myelodysplastic syndromes (MDS)[57].

Apart from being an oncogene, *CRNDE* also functions in maintaining the multipotency of progenitor cells and regulating cell differentiation. The promoter of mouse *CRNDE* locus is bound to Myc family transcription factors, and knockdown of *CRNDE* causes a decrease in the expression of pluripotency transcription factors, including Klf4 and SOX2[58]. Lin et. al. identified most abundant expression of *CRNDE* in both human iPSCs and differentiating neurons, which also suggested the potential role of *CRNDE* in regulating neurogenesis and brain development[59]. Moreover, *CRNDE* has been found to elevated in glioma stem cells and inhibits the function of miR-186 to maintain the self-renewal and proliferation of glioma stem cells[60].

Finally, *CRNDE* has role in regulating metabolism. Ellis et.al. observed that the nuclear expression of *CRNDE* was downregulated by insulin and IGF signaling, whereas inhibiting PI3K/mTOR and Raf/MAPK signaling pathways could rescue the effect. Moreover, they identified that *CRNDE* could regulate genes involved in glucose and lipid metabolism, which overlaps the downstream genes regulated by insulin/IGF signaling[61]. This suggests additional role of *CRNDE* in regulating cellular process and could be applied to cancer therapies.

In summary, current research has identified *CRNDE* as a targetable oncogene in that it can promote cell differentiation and proliferation, alter cell metabolism and promote the stemness of cancer stem cells.

Chapter 2: Data and Methods

2.1 Brief introduction

From analysis of bulk RNA-seq data of SDS patients and human bone marrow CD34+ cells, we identified the expression of lncRNA *CRNDE* and its isoforms are silenced in SDS patients. We showed an overexpression of specific isoform of *CRNDE*, *CRNDE204*, which was silenced in several myelodysplastic syndromes, could promote proliferation and differentiation in both erythropoietic and megakaryocytic pathways in K562 cells, yet the same effect was not seen in other isoforms derived from *CRNDE* splicing axis. RNA-seq and transcriptional network analysis further showed *CRNDE204* is involved in the key regulatory pathway of driving regulates erythropoiesis and myelopoiesis. Yeast 3-hybrid identified a total of 24 proteins as binding partners of *CRNDE204*. Bioinformatics analysis and knock-down assays revealed PUS1 and NR1H3 as the key factors in regulating erythropoiesis and myelopoiesis. Taken together, we revealed a unique role of the *CRNDE* isoform expression axis in regulating hematopoiesis, whose modulation could constitute a novel therapy for bone marrow failure and cancer.

2.2 Material and Methods

2.2.1 Cell Culture and generation of stable cell lines

K562 cells were obtained from ATCC and cultured in RPMI 1640 medium (ThermoFisher, Waltham, MA) with 10% fetal bovine serum. Stable cell lines were generated by transfecting pcDNA3.1(-)-*CRNDE201* and pcDNA3.1(-)-*CRNDE204* plasmids into K562 cells using Lipofectamine LTX (Invitrogen, 15338100) following manufacturers protocol. 48 hours after

transfection, cells were selected using 400μg/mL Geneticin (Gibco, 10131035), for two weeks and until un-transfected cells died off.

For maintaining the stable cell lines, cells were cultured at 37C, 5% CO_2 in RPMI 1640 medium supplemented with 10% FBS and 500ug/mL Geneticin.

2.2.2 Assays for cell differentiation

For erythroid differentiation assay, each cell line (K562 control, K562 *CRNDE201*, K562 *CRNDE204*) was plated in triplicates in a 24-well plate at 100,000 cells/well. 40uM Hemin in NaOH was added to induce differentiation. Control cells received the same amount of 50uM NaOH. Cells were incubated for 72 hours before evaluation. Cell differentiation was evaluated by staining of erythroid lineage markers CD71 and CD235a (Biolegend 334107, 349104) conjugated with APC and FITC respectively. Fluorescence signals were detected and analyzed using flow cytometry on BD LSR Fortessa.

For megakaryocyte differentiation assay, each cell line was plated in triplicates in a 12 well plates at 500,000 cells/well. PMA in DMSO was added to induce differentiation. Control cells received the same amount of DMSO. Cells were incubated for 72 hours before evaluation. Cell differentiation was evaluated by staining of megakaryocyte lineage markers CD41/CD61 (Biolegend 359805) conjugated with PE. Fluorescence signals were detected and analyzed using flow cytometry on BD LSR Fortessa.

2.2.3 RNA sequencing and bioinformatics analysis

RNA was isolated from K562 stable cell lines using RNeasy Plus Mini Kit (Qiagen, 74136) following manufacturers protocol. Whole RNA was sent to Genewiz for library

preparation using Illumina kit with rRNA depletion (Illumina Ribo-Zero rRNA removal) and sequencing on Illumina HiSeq 150 reads with paired-end sequencing and 35 million read depth. Processing of raw data was done by Dr. Leon Wert-Lamas.

2.2.4 Protein Sequencing and Bioinformatic Analysis

CRNDE201 and 204 overexpressing stable K562 cells and CRNDE204 knock-down cells were used for XRNAX, a method for generic purification of protein crosslinked RNA. 100μM 4-thouridine (4sU) (Sigma, T4509) was added to media consisting of 10 million cells and incubated for 16 hours before UV crosslinking. Downstream steps for XRNAX experiments were followed as previously described by Trendel et al[62]. Processing of raw data was done by Dr. Leon Wert-Lamas.

2.2.5 Yeast-three hybrid screening

This method was to Dr. Leon Wert-Lamas and Sophie Kim. For Y3H high throughput screening, two centopools containing all proteins in human ORF in PdestAD vector were prepared in SC-W media, and *CRNDE201* or *CRNDE204* was cloned into pIIIa vector and was kept in SC-U media at 30°C overnight. 10uL of vectors containing centopool and 10uL of vectors containing *CRNDE201* or *CRNDE204* were aliquoted into a plate with YPD and incubated overnight at 30°C. The yeast colonies were selected in SC-WU media and grown overnight at 30°C. The cells were plated on SC-WUH media and incubated at 30°C for 3 to 7 days until colonies form. The colonies were picked and transformed to SC-WU media and incubated at 30°C overnight. The yeast cells were lysated and DNA were extracted to perform PCR and sequencing. For PCR, AD/Term primers and pIIIa F1/R1 primers were used. DNA

extracts were sent to Genewiz for sequencing. For pairwise screening, the proteins that were positive in high throughput screening were cloned into yeast and 10uL of yeast containing target proteins and 10uL of yeast transfected with *CRNDE201* or *CRNDE204* were aliquoted to 96-well plate containing YPD and incubated overnight at 30°C, selected in SC-WV media, plated on SC-WUH media and incubated for 3-7 days until colonies form.

2.2.6 *CRNDE* binding partner knockdown and differentiation assay

To knockdown PUS1, NR1H3, RPL22 and NPM2 in K562 cells, three shRNA constructs in plasmids for each protein targets were designed and kindly provided to us by Dr. Huan Yang. Lipofectamine 2000 (Invitrogen, 11668500) were used for shRNA transfection following manufacturer's protocol.

For erythroid and megakaryocyte differentiation assays, K562 cells overexpressing *CRNDE204* were used in this process. Cells were plated in triplicates in 24-well plates at 170,000 cells/well and were transfected by shRNA constructs. The cells were incubated for 24 hours before induction of differentiation. For erythroid differentiation, 40uM Hemin in NaOH was added to induce differentiation. Control cells received the same amount of 50uM NaOH. Cells were incubated for 48 hours before evaluation. Cell differentiation was evaluated by staining of erythroid lineage markers CD71 and CD235a (Biolegend 334107, 349104) conjugated with APC and FITC respectively. Fluorescence signals were detected and analyzed using flow cytometry on BD LSR Fortessa. For megakaryocytic differentiation, 10nM PMA in DMSO was added to induce differentiation. Control cells received the same amount of DMSO. Cells were incubated for 48 hours before evaluation. Cell differentiation was evaluated by

staining of megakaryocyte lineage markers CD41/CD61 (Biolegend 359805) conjugated with PE. Fluorescence signals were detected and analyzed using flow cytometry on BD LSR Fortessa.

Table 1: List of shRNA constructs used in this thesis.

shRNA	Sequence
PUS1 shRNA1	5'- GCA CGC ACA ACT TCC ACA ATT -3'
PUS1 shRNA2	5'- GAG CTT CAT GAT GCA TCA GAT -3'
PUS1 shRNA3	5'- GCA CTT CGA GAA GTA CAA CCA -3'
NR1H3 shRNA1	5'- GTG CAG GAG ATA GTT GAC TTT -3'
NR1H3 shRNA2	5'- GCA ACT CAA TGA TGC CGA GTT -3'
NR1H3 shRNA3	5'- GAT CTG GGA TGT GCA CGA ATG -3'
RPL22 shRNA1	5'- GTT CTG AAG TTC ACT CTT GAT -3'
RPL22 shRNA2	5'- CGT GAC TGG TTG CGC GTA GTT -3'
RPL22 shRNA3	5'- GCA CAC AAT TAT GTC TGC TAA -3'
NPM2 shRNA1	5'- GTT GCT TCA TAC GAT TTG CTT -3'
NPM2 shRNA2	5'- GTT ATG AAG CAT CAG ACC TAA -3'
NPM2 shRNA3	5'- GAG GAA ATA AGA GCC AGC GTT -3'

2.2.7 Validation of knockdown efficiency

To validate the knockdown efficiency of the shRNAs used above, K562 cells overexpressing *CRNDE204* were plated into 6-well plates and transfected with the scramble control and three different shRNA plasmids against each protein in duplicates using Lipofectamine 2000 (Invitrogen, 11668500) following manufacturer's protocol. After 24 hours, the cells were sorted by BD FACSAria II cell sorters to obtain cells that took up the plasmid

(appeared as BFP positive). Total mRNA was extracted from sorted cells using RNeasy Mini Kit (Qiagen, 74106) and cDNA was generated using SuperScript III First-Strand Synthesis System (Invitrogen, 18080051). The expression of mRNA of PUS1, NR1H3, RPL22 and NPM2 was measured by quantitative PCR (qPCR) with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, 1725271). The relative expression of these mRNAs was normalized by comparing to the expression of housekeeping gene beta-actin.

Table 2: qPCR primers used in this thesis:

Primers	Sequence
Beta-actin_F	5'- AAC TTC TAC AAT GAG CTG CG -3'
Beta-actin_R	5'- CCT GGA TAG CAA CGT ACA TGG -3'
PUS1_F	5'- GTC TGG GAG GAC GGA GAA CAT -3'
PUS1_R	5'- CAG CAC GAT CTT CCG CTT G -3'
NR1H3_F	5'- GTC TCT GTG CAG GAG ATA GTT G -3'
NR1H3_R	5'- GTC TTC AGC AGG GCA ATC T -3'
RPL22_F	5'- AAA GTG AAC GGA AAA GCT GGG - 3'
RPL22_R	5'- TCA CGG TGA TCT TGC TCT TGC - 3'
NPM2_F	5'- GGC TGT TGC TTC ATA CGA TTT G -3'
NPM2_R	5'- CAT CTT CTT GTC CTC CTG GTT T -3'

2.2.8 Statistical analysis

Unpaired t tests, ANOVAs, and correlations were calculated using GraphPad Prism version 9.00 for Mac, GraphPad Software, La Jolla California USA (https://www.graphpad.com/). For enrichment analysis, a Fisher Exact test was done in

GraphPad Prism to show significance. In cell differentiation assays, the expression of surface markers was analyzed using FlowJo software, and significance was calculated using GraphPad Prism software. Volcano plots were generated using RStudio. The results from DATNA were represented by Cytoscape.

2.3 Results

2.3.1 *CRNDE* isoforms are silenced in SDS patients

We analyzed bulk RNA-seq samples from SDS patients and healthy CD34+ control and focused on the differential expression of lncRNAs between patients and healthy control. Our analysis revealed *CRNDE* isoforms *CRNDE201* and *CRNDE204* were downregulated by 5-fold compared with healthy control, which were more significant than any other lncRNA identified from SDS patient samples (Figure 3A). Further analysis of the sequences and structures of *CRNDE* isoforms showed that the different isoforms of *CRNDE* could be categorized into two backbone models. We focused on *CRNDE201* and *CRNDE204* because they represented paradigmatic structure examples and appeared in literature as archetypical silenced or overexpressed isoforms. By comparing the exon maps of the two isoforms, we found out that the structure of *CRNDE201* and *CRNDE204* was different, and alternative splicing event happened at exon 4 of *CRNDE201*(Figure 3B). To predict the folding and 3D structure of *CRNDE201* and *CRNDE204*, we used Venna fold algorithms to generate structure predictions of the two isoforms (Figure 3C and D). The two *CRNDE* isoforms only have 54% structural homology, and *CRNDE204* forms large loop-like structure. We hypothesized that the difference in splicing and folding of these two isoforms may have an impact on their functions in hematopoiesis, and *CRNDE204* may have a unique protein-binding motif or a folding structure that enables it to bind to hematopoietic transcription factors. This difference in their structure may contribute to the difference in their function in regulating hematopoietic cell development.

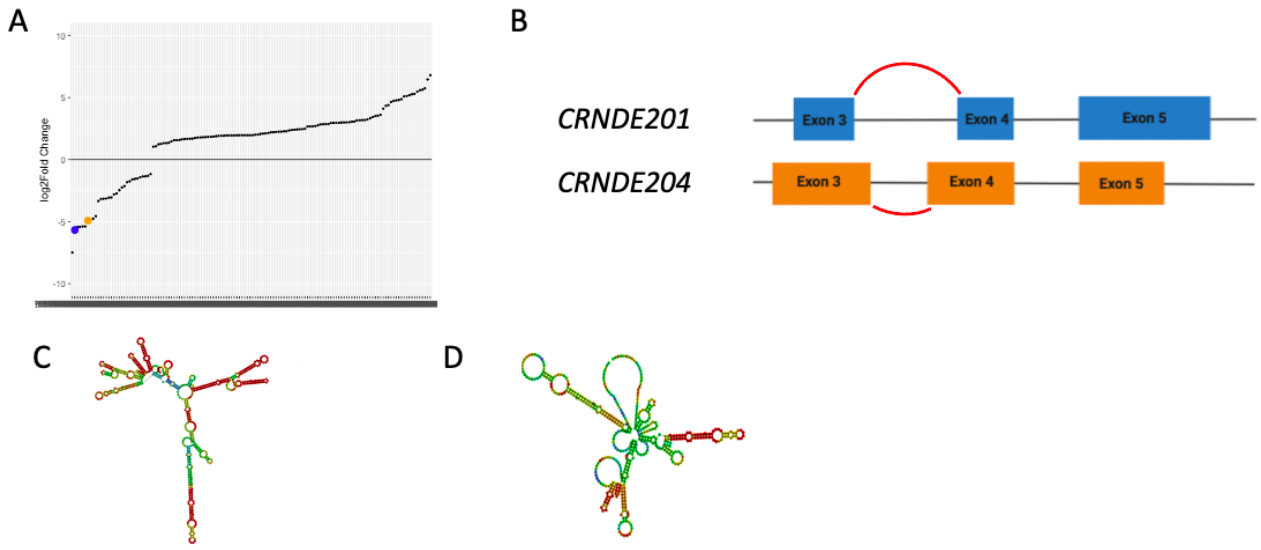


Figure 3: *CRNDE* isoforms are silenced during SDS hematopoiesis. (A) Waterfall plot of log2fold change of differentially expressed genes in SDS hematopoiesis. *CRNDE201* (highlighted in blue) and *CRNDE204* (highlighted in orange) are the most downregulated genes. Figure credits to Dr. Leon Wert-Lamas. (B) *CRNDE* splicing map of isoform *CRNDE201* (top) and *CRNDE204* (bottom). Alternative splicing event is seen in exon 4. (C) RNA structure diagram of *CRNDE201* predicted by Vienna fold program using centroid algorithm. (D) RNA structure diagram of *CRNDE204* predicted by Vienna fold program using centroid algorithm. Figure credits to Dr. Leon Wert-Lamas.

2.3.2 *CRNDE204* promotes erythroid and megakaryocytic differentiation

Because we observed a downregulation of *CRNDE* isoforms in SDS patients, we hypothesized the absence of *CRNDE* in SDS patients could be associated with and may cause the dysregulation in the development and maturation of myeloid-derived cells. To validate this hypothesis, we overexpressed *CRNDE201* and *CRNDE204* in K562 cells and generated stable cell lines. K562 cells were first isolated from bone marrow of a chronic myeloid leukemia (CML) patient in terminal blast crisis by Lozzio et al in 1975 and this cell line was characterized by its undifferentiated nature[63]. Further studies on this cell line had shown K562 cells had the potential to differentiate into erythrocytes[64] and megakaryocytes[65] after stimulating with pro-differentiation agents. We chose K562 cells as our model system because these cells express

erythroid-specific lineage marker CD71 and CD235a and are identified to be similar to premature colony-forming unit erythroid cells and more differentiated erythroblasts. Additionally, erythroid differentiation transcriptional factors GATA1, GATA2 and FOG1 are expressed in K562 cells, which further suggests the erythroid origin of K562 cells[66].

For erythroid differentiation of K562 cells, we chose hemin as differentiating agent because it was described as a strong inducer of erythropoiesis in regulating hemoglobin synthesis and has been used as a prescription drug to treat porphyria-related symptoms[67]. To assess the ability of the cells to differentiate into the mature erythrocytes, erythroid lineage-specific surface markers CD71 and CD235a were used as the indicators of erythrocyte maturation. Mature erythrocytes express high levels of CD235a but low levels of CD71. K562 cells overexpressing *CRNDE204* promoted differentiation towards erythroid lineages, whereas K562 cells overexpressing *CRNDE201* inhibits differentiation when comparing with the control as shown by mean fluorescence intensity of CD71 and CD235a (Figure 4A), and the percentage of CD235a+ cells post differentiation (Figure 4B). *CRNDE204* overexpression increased the percentage of CD235a+ cells by two folds, whereas *CRNDE201* overexpression decreased the percentage of mature erythrocytes by two folds.

We next investigated whether overexpression of *CRNDE204* also promote K562 cells to differentiate along the megakaryocytic lineage. For megakaryocyte differentiation, we chose phorbol 12-myristate 13-acetate (PMA) as the differentiation agent. PMA can cause K562 cells to change their morphology and adhesion properties, undergo cell growth arrest, and promote expression of megakaryocyte-specific protein Pyk2 kinase through MAPK-dependent and independent pathways[68]. PMA-induced megakaryocyte differentiation is also reported to modulate the mitochondrial function by reducing the activity of respiratory chain complex IV[69].

Megakaryocyte differentiation assay results showed increased differentiation indicated by increased CD41/CD61 mean fluorescence intensity in differentiated K562 cells overexpressing *CRNDE204*. However, cells overexpressing *CRNDE201* showed inhibited differentiation compared with control (Figure 4C). Quantification of the number double positive cells showed significant difference between the percentage of mature megakaryocytes post differentiation between cells overexpressing *CRNDE201* and cells overexpressing *CRNDE204* (Figure 4D).

In conclusion, our results showed *CRNDE204* was the key isoform of *CRNDE* in regulating myelopoiesis. Specifically, *CRNDE204* overexpression promoted erythroid and megakaryocyte maturation. Oppositely, *CRNDE201* overexpression suppressed differentiation towards mature erythrocytes and megakaryocytes.

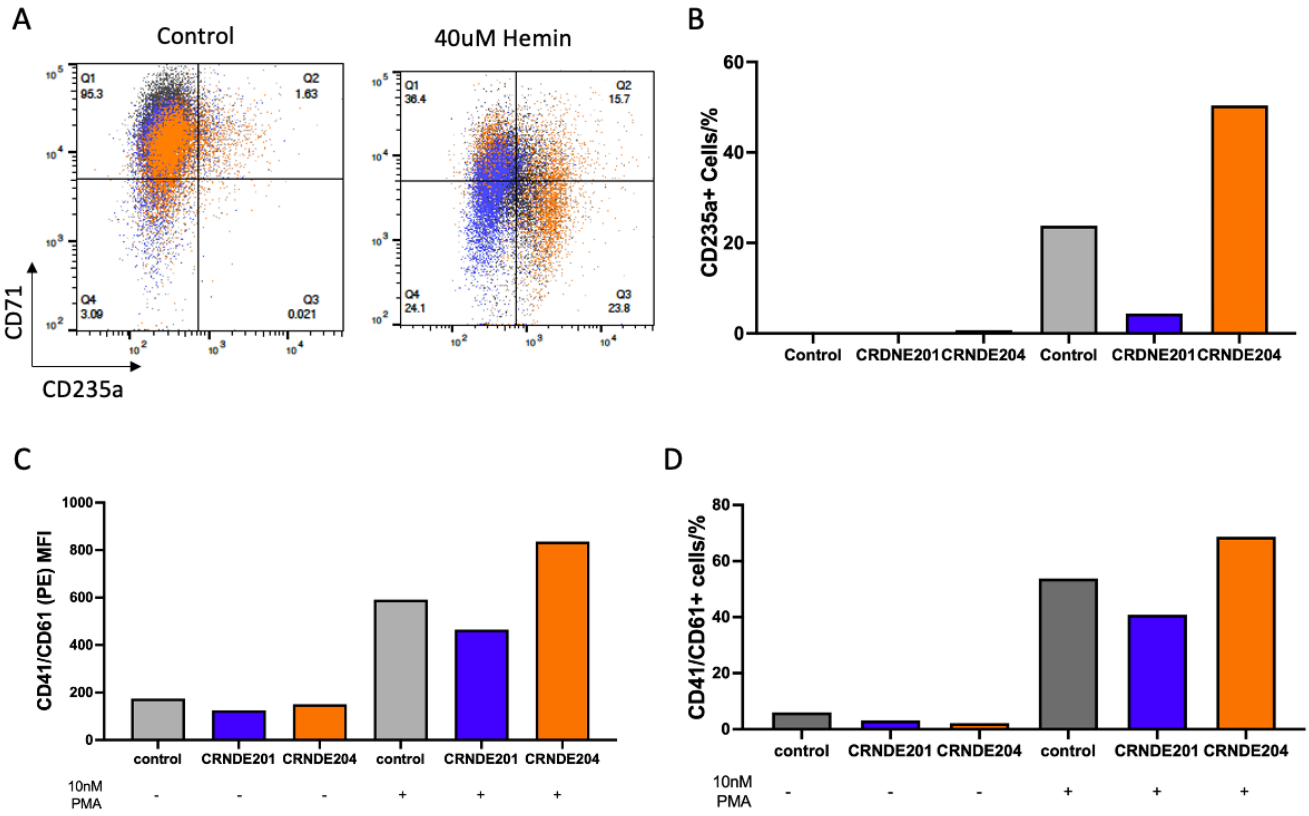


Figure 4: *CRNDE204* promotes erythroid and megakaryocytic differentiation. (A) K562 cells transfected with PCDNA3.1 control, CRNDE201 and CRNDE204 plasmids were induced by NaOH (control) and 40uM hemin for erythroid differentiation assay. The result for differentiation assay was analyzed by flow cytometry:

x-axis represent the surface marker CD235a (FITC) and y-axis represent CD71 (APC). Gray: K562 control; blue: K562 overexpressing *CRNDE201*, orange: K562 overexpressing *CRNDE204*. (B) Bar graph shows quantification of the percentage of CD71- CD235a+ cells. (C) K562 cell lines were induced by DMSO (control) and 10nM PMA for megakaryocyte differentiation. Bar graphs show the average intensity of CD41/CD61+ cells post differentiation (D) Quantification of percentage of CD41/CD61+ cells post differentiation.

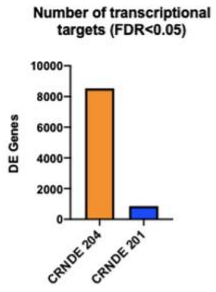
2.3.3 *CRNDE204* regulates erythroid and megakaryocytic development

Our results showed cells overexpressing different *CRNDE* isoforms had distinct patterns of differentiation, we hypothesized there might be a difference in the transcriptional networks regulated by these two different isoforms and wanted to comprehensively investigate the role of *CRNDE* isoforms in transcriptomics and proteomics level. We conducted RNA-seq of K562 cells overexpressing *CRNDE201*, *CRNDE204* and a control plasmid to see whether they have distinct transcriptional program. Transcriptional targets of *CRNDE* were identified by significance (FDR<0.05) and differentially expressed genes between control and *CRNDE* overexpression. Over 8,000 transcriptional targets were identified for *CRNDE204* by this criterion, however, only 866 transcriptional targets were identified for *CRNDE201* (Figure 5A). This difference in transcriptional programs elicited by *CRNDE201* and *CRNDE204* was aligned with the opposite phenotype of the diseases where they accumulate.

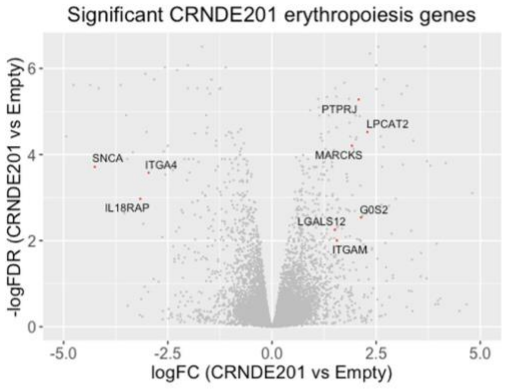
To find whether the genes regulated by *CRNDE204* were involved in erythropoiesis, we overlapped maturation-related erythropoiesis genes in human-derived nucleated red blood cells profiled by Mello et.al to our RNA-seq results conducted and analyzed by Dr. Leon Wert-Lamas and Dr. Kushani Shah. A total of 6569 genes were identified by Mello et.al to be expressed in at least one stage of bone marrow erythropoiesis with differential expression patterns at different stages, which were representative of the gene expression profile throughout erythroid maturation[70]. We chose to use those erythroid-specific genes identified by Mello et.al. as a

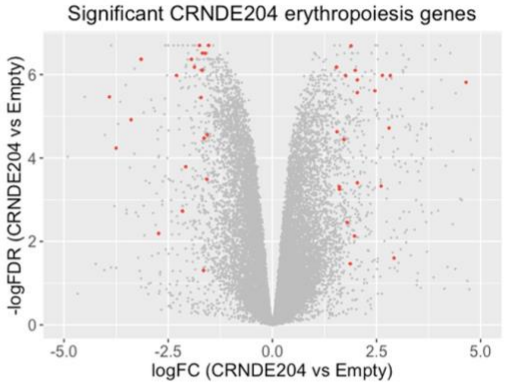
reference for identifying the erythropoiesis transcriptional network of under the control of *CRNDE* isoforms. Similarly, we chose the myelopoiesis genes annotated by Ferrari et.al in 2007 and overlapped with to our RNA-seq results. Ferrari et.al integrated myelopoiesis gene analyses from 24 different experiments of 8 different cell types in myelopoiesis stages and generated a list of 5929 genes that was expressed in every type of cell from myeloid lineage, and further defined clusters of genes that were differentially expressed compared with CD34+ progenitor cells[71]. If the published data overlaps with our RNA-seq and protein-seq data, we could say that the overlapped gene is in either the erythropoiesis or myelopoiesis transcriptional network. We further narrowed down the gene by significance, which were identified as FDR<0.05 and log2-fold change less than -1.5 or greater than 1.5 and were highlighted in red in volcano plots. *CRNDE204* is shown to regulate a larger group of genes that is significant in erythropoiesis (Figure 5B) and myelopoiesis level (Figure 5C) compared with *CRNDE201*. Furthermore, *CRNDE204* overexpression regulated a significant number of proteins that were involved in erythropoiesis and myelopoiesis. Fisher's exact test of these genes or proteins regulated by *CRNDE204* was significantly enriched.

A

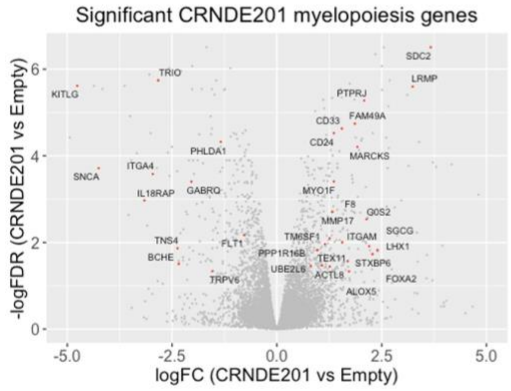

Number of transcriptional targets (FDR<0.05)
DE Genes
10000
8000
6000
4000
2000
0
CRNDE 204
CRNDE 201

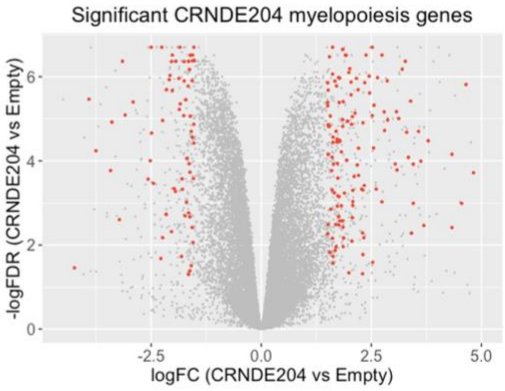
B


Significant CRNDE201 erythropoiesis genes
-logFDR (CRNDE201 vs Empty)
logFC (CRNDE201 vs Empty)
PTPRJ
LPCAT2
MARCKS
SNCA
ITGA4
IL18RAP
G0S2
LGALS12
ITGAM

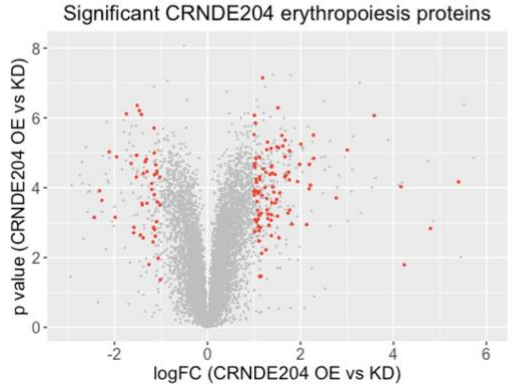

Significant CRNDE204 erythropoiesis genes
-logFDR (CRNDE204 vs Empty)
logFC (CRNDE204 vs Empty)

C


Significant CRNDE201 myelopoiesis genes
-logFDR (CRNDE201 vs Empty)
logFC (CRNDE201 vs Empty)
SDC2
TRIO
LRMP
KITLG
PTPRJ
CD33
FAM49A
CD24
PHLDA1
MARCKS
ITGA4
SNCA
IL18RAP
GABRQ
MYO1F
F8
G0S2
MMP17
SGCG
TM6SF1
ITGAM
LHX1
TNS4
FLT1
PPP1R16B
TEX11
STXBP6
BCHE
UBE2L6
ACTL8
FOXA2
TRPV6
ALOX5


Significant CRNDE204 myelopoiesis genes
-logFDR (CRNDE204 vs Empty)
logFC (CRNDE204 vs Empty)

D


Significant CRNDE204 erythropoiesis proteins
p value (CRNDE204 OE vs KD)
logFC (CRNDE204 OE vs KD)

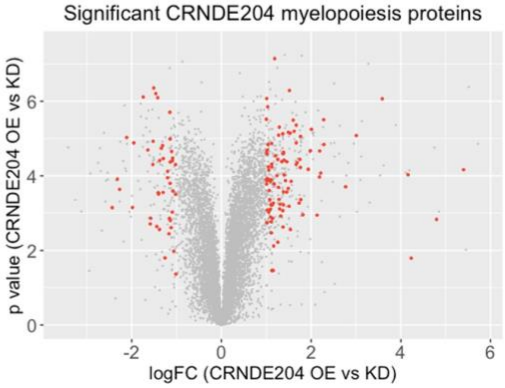

Significant CRNDE204 myelopoiesis proteins
p value (CRNDE204 OE vs KD)
logFC (CRNDE204 OE vs KD)

Figure 5: *CRNDE204* regulate key erythropoiesis and myelopoiesis genes. (A) Quantification of number of significant transcriptional targets (FDR<0.05) between *CRNDE201* and *CRNDE204*. (B) Volcano plots showing the expression of differential expressed genes in *CRNDE201* overexpression vs control (left) and *CRNDE204* overexpression vs control in K562 cells, with significant (FDR<0.05, fold change>1.5 or <-1.5) erythropoiesis genes highlighted. The log fold change in *CRNDE* overexpression vs control is represented on the x-axis and minus log10 FDR is represented on the y-axis. (C) Volcano plots showing the expression of differential expressed genes in *CRNDE201* overexpression vs control (left) and *CRNDE204* overexpression vs control in K562 cells, with significant (FDR<0.05, fold change>1.5 or <-1.5) myelopoiesis genes highlighted. The log fold change in *CRNDE* overexpression vs control is represented on the x-axis and minus log10 FDR is represented on the y-axis. (D) Volcano plot of differential expressed proteins in *CRNDE204* expression vs knockdown, with significant (FDR<0.05, fold change>1.5 or <-1.5). The log fold change in *CRNDE204* overexpression vs knock-down is represented on the x-axis and p value is represented on the y-axis. Left: volcano plot of protein-seq data highlighting significant erythropoiesis protein as red dots. Right: volcano plot of protein-seq data highlighting significant myelopoiesis proteins as red dots.

To determine the gene regulatory pathways involved in *CRNDE204* transcription program, we performed Ingenuity Pathway Analysis (IPA) on protein sequencing data of *CRNDE204* overexpression. We identified important upstream regulators of *CRNDE204* transcriptional network, including the master erythropoiesis regulator GATA1. Interestingly, NR1H3, previously identified as crucial in macrophage activation and innate inflammatory pathways, was identified as upstream regulator of *CRNDE204* transcriptional network. Additionally, IL-6, SMARCA4 and HMGA1 were also identified as upstream regulator of *CRNDE204* (Figure 6). IL-6 is a pro-inflammatory cytokine that is important in regulating inflammatory responses, and it also has function in promoting the emergence of HSCs in embryonic development. IL-6 signaling is also important for HSC-independent neutrophil development and definitive erythroid-myeloid progenitor cells, which is regulated by Notch1 signaling[72]. SMARCA4 is a protein that plays a role in chromatin modification, and it was identified as an important regulator of stemness and differentiation in HSPCs. Knocking down of SMARCA4 in hematopoietic progenitor cells disrupted self-renewal and skewed towards myeloid differentiation[73]. HMGA1 is also reported to be regulate gene expression in embryonic stem cells and is involved in lymphoid tumorigenesis[74]. The collective functions of IL-6, SMARCA4 and HMGA1 suggests a potential role of *CRNDE204* in regulating gene expression

in hematopoietic stem cells and progenitor cells. IPA also identified top regulatory networks and most of them are inflammatory pathways and immunological disorders (Figure 2.4). Consistent with the top regulatory networks identified, inflammatory response, immunological disease and cancer were among the top disease and disorders involved in *CRNDE204* transcription profile.

Taken together, these results showed that *CRNDE204* could promote myeloid lineage development and maturation by regulating a group of genes and proteins that are associated with important in regulatory networks of erythropoiesis and myelopoiesis, and the upstream regulators of *CRNDE204* are associated with hematopoietic stem cell development.

Upstream Regulators	Top Diseases/Disorders	Top Regulator Effect Networks
IL6	Inflammatory Response	DOK8, IFNA4, IFNL1, IRF7, **SAMSN1**, SASH1
LDLR	Organismal Injury and Abnormalities	IL10R1, IL21, IRF1, mir-17
NR1H3	Immunological Disease	PI3K, RAC1, **REL**, **TGM2**
AGT	Cancer	IL28, IRF1
GATA1		DUSP1, IL128, IFNα, IRF1
ZBTB17		BMPR2, CCN1, **STAT1**, VEGF, YAP1
SMARCA4		GATA3, PI3K, **STAT1**
HMGA1		IRF1, LINC01139
		KLF4, **REL**

Figure 6: IPA identified regulatory networks of *CRNDE204*. IPA result of *CRNDE204* overexpression vs knock-down. Genes and proteins that are involved in erythropoiesis or myelopoiesis were highlighted in bold.

2.3.4 Identification of *CRNDE* binding partners through yeast-three hybrid

To identify potential binding partners of *CRNDE204*, we used yeast three-hybrid system (Y3H) to screen whole human ORFeome to find potential interactions between *CRNDE* and

protein partners (Figure 7A). Y3H is a screening system to identify RNA-protein binding relationships. In this system, one of the hybrid proteins is designed to have DNA binding domain, and the other is designed to have the activation domain that can activate the reporter gene. If there is direct binding relationship between the hybrid RNA and the proteins, the reporter gene will be activated and can be detected[75]. Twenty-four binding partners of *CRNDE204* were identified from Y3H (Figure 7B). However, no binding partners of *CRNDE201* was identified in Y3H screening. We overlapped these binding partners to erythropoiesis and myelopoiesis transcriptional programs in previous sections, and highlighted proteins that are important in both erythropoiesis and myelopoiesis. mRNA processing factor CSTF2 also appeared to be involved in regulation of myelopoiesis, which implied different splicing and processing may affect the role of *CRNDE* isoforms in regulating erythropoiesis (Figure 7C).

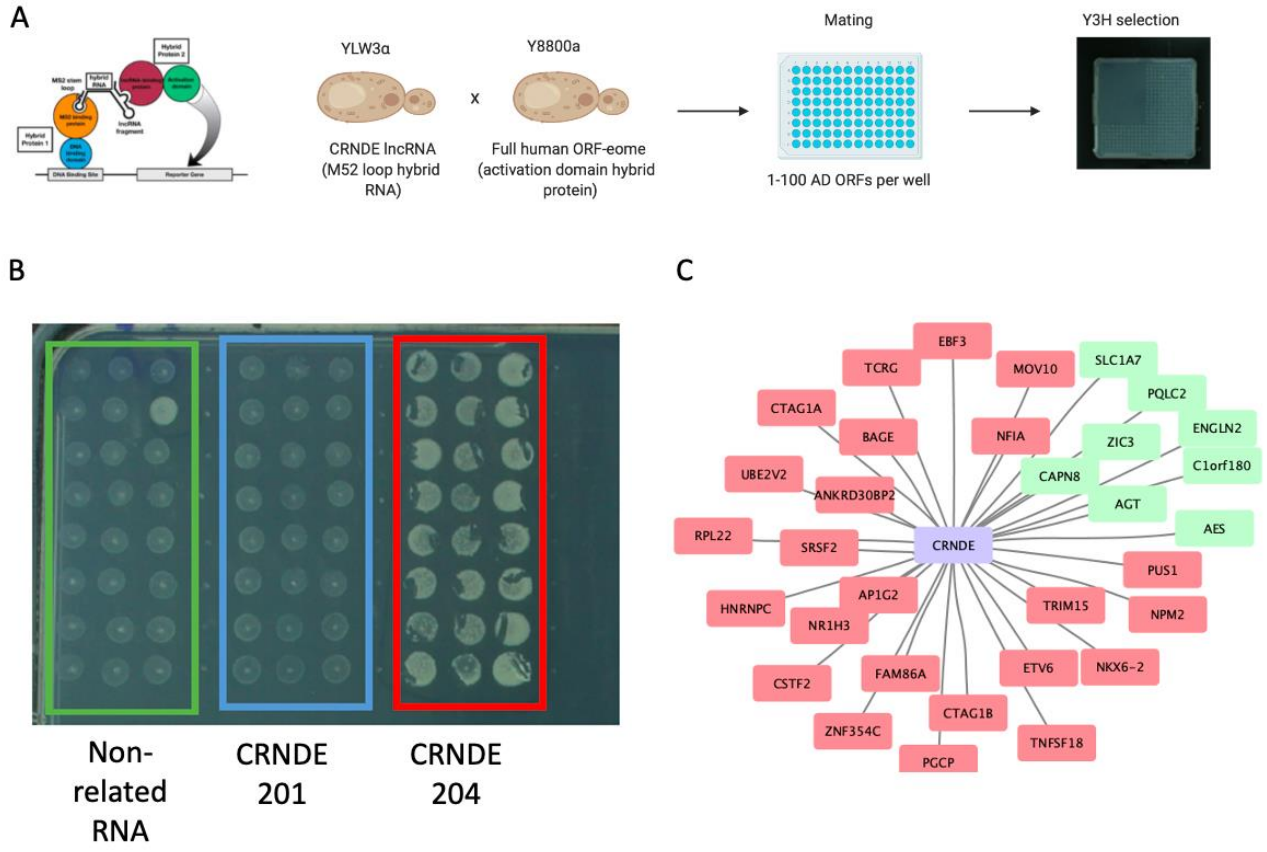


Figure 7: Identification of *CRNDE* binding partners by yeast-three hybrid. (A) schematics of yeast 3 hybrid screening (B) Y3H result of showing the binding partners of *CRNDE* isoforms. (C) Cytoscape represents *CRNDE204* binding partners. Red boxes are factors important in either erythropoiesis or myelopoiesis.

To find lncRNA-transcription factor-gene relationship, we ran DATNA algorithm on RNA-seq data from AML patient data analyzed Bamopoulos et al to predict the number of lncRNA-TF-gene triads formed in erythropoiesis and myelopoiesis[76]. DATNA is an algorithm was developed based on lncMOd, a computational model developed by Li et al that integrates genome-wide lncRNA data and gene expression profile with TF-target regulators to identify lncRNAs that affect TF activity[77]. We then identified erythropoiesis or myelopoiesis-related genes from triads predicted by DATNA (Figure 8A). Among the erythropoiesis triads, PUS1 formed the most triads with *CRNDE204*, which validated its role in regulating erythropoiesis. Similarly, in myelopoiesis, NR1H3 formed the greatest number of triads with *CRNDE204*. We wanted to focus on genes whose expression are correlated with the expression of the protein targets, so we narrowed down the network to show only the concordant triads formed by each transcription factor (Figure 8B). In this view, RPL22 and NPM2 appears more significant than other targets of *CRNDE204* and could also be important in regulating erythropoiesis and myelopoiesis. We chose to focus on PUS1, NR1H3, RPL22 and NPM2 for the following phenotypic experiments because there are evidence showing these proteins as regulators of erythropoiesis and myelopoiesis. PUS1 (pseudouridine synthase 1) encodes for pseudouridine synthase protein and is responsible for synthesis of pseudouridine on tRNAs and maintains their function. Mutations in PUS1 is found in myopathy, lactic acidosis, and sideroblastic anemia (MLASA), a subtype of sideroblastic anemia. Evidence has shown that missense mutation in PUS1 can cause a disruption in normal pseudouridine synthesis, thus reduce translation efficiency, cause mitochondria dysfunction and trigger formation of sideroblast[78]. NR1H3 (or

LXR alpha) belongs to the nuclear receptor family proteins that regulates cholesterol and glucose metabolism. Studies also showed that LXRs were able to repress LPS-induced inflammatory cytokine production and production of other pro-inflammatory molecules[79]. RPL22 encodes for 60S ribosomal protein L22, yet it was identified dispensable for ribosome assembly and protein translation[80]. RPL22 is identified as an RNA-binding protein that can bind to 28S ribosomal RNA, EBV RNA and HCV RNA. RPL22 also directly regulates the translation of its homologous paralog RPL22L1 by interacting with telomerase RNA. Both RPL22 and RPL22L1 bind to SMAD1 mRNA but have opposing functions: RPL22 inhibits the expression of SMAD1, whereas RPL22L1 acts against the inhibition. In regulating hematopoiesis, the balance between RPL22 and RPL22L1 is crucial in maintaining HSC emergence: When RPL22 dominates, SMAD1 expression is repressed and causes anemia; when RPL22L1 dominates, SMAD1 is overexpressed and causes cancer[81]. NPM2 belongs to the nucleoplasmin/nucleophosmin family proteins and can regulate HSC/progenitor cell cycle progression. However, overexpression of NPM2 can inhibit the differentiation of myeloid progenitor cells[82].

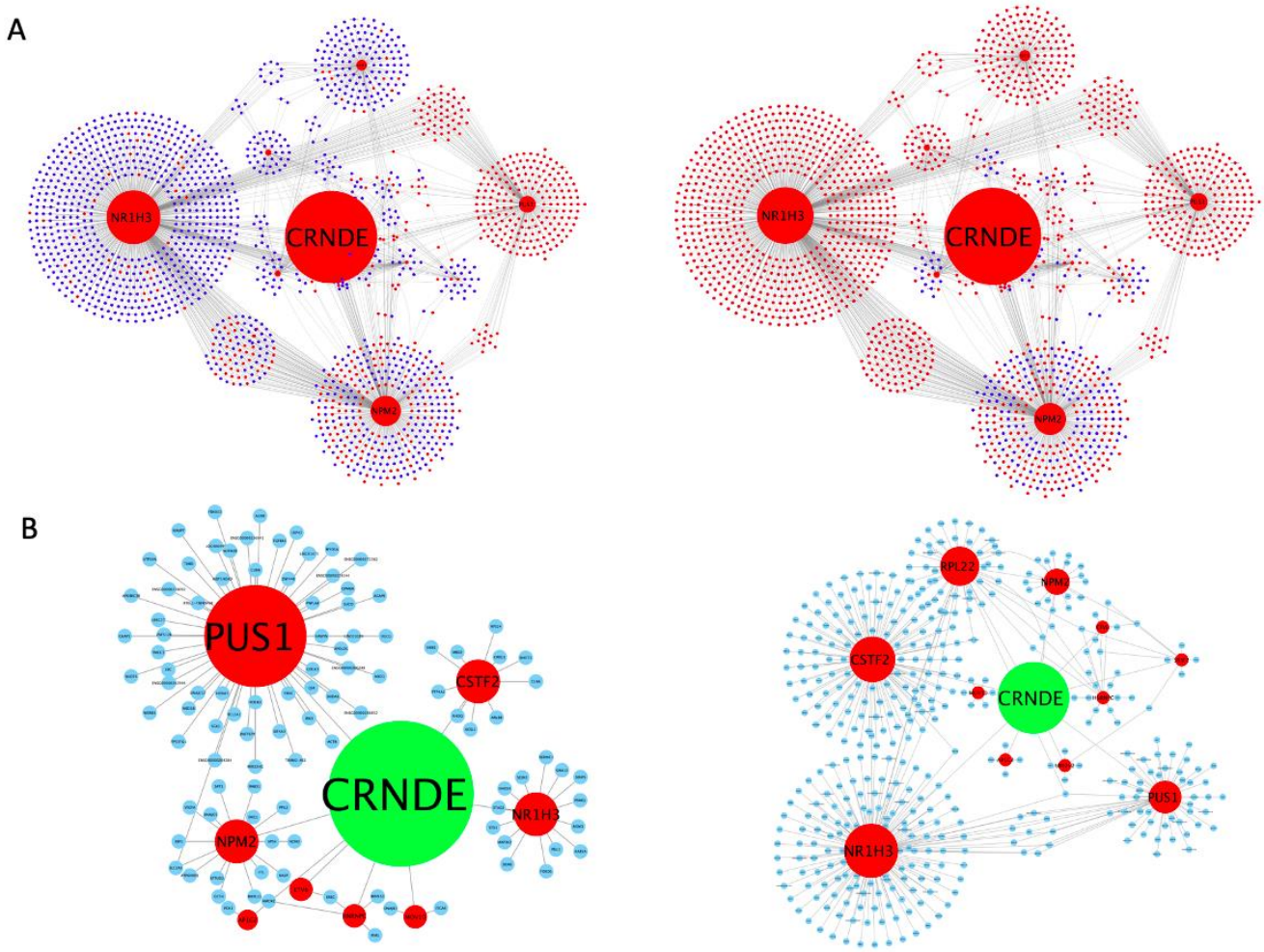


Figure 8: DATNA identified key binding partners of *CRNDE204*. (A) DATNA result on data from Bamopoulos et al showing RNA-transcription factor-gene triads forming by *CRNDE204* and its binding partners. Left: Cytoscape presentation of the binding triads. Red dots are key transcriptional factors in *CRNDE204* regulated erythropoiesis pathways. Size of each node is assigned corresponding to the number of triads each binding partner is forming. Right: Cytoscape presentation of the binding triads. Red dots are key transcriptional factors in *CRNDE204* regulated myelopoiesis pathways. Size of each node is assigned corresponding to the number of triads each binding partner is forming. (B) Cytoscape showing the number concordant triads formed by *CRNDE204*. Left: Cytoscape presentation of the concordant triads in erythropoiesis pathway. Right: Cytoscape presentation of the concordant triads in myelopoiesis pathway.

In summary, our results from yeast-three hybrid and DATNA has identified several binding partners of *CRNDE204* that might be important in regulating key developmental processes in erythropoiesis and myelopoiesis.

2.3.5 *CRNDE204* regulates erythroid and megakaryocytic development by binding to PUS1

After identified PUS1, NR1H3, RPL22 and NPM2 as important erythropoiesis and myelopoiesis regulators in *CRNDE204* transcriptional pathway by Y3H and DATNA, we proceeded to knockdown them to further validate the role of these binding partners of *CRNDE204* in regulating erythropoiesis and myelopoiesis. For each of the protein targets, we designed 3 shRNA constructs and tested the knockdown efficiency in K562 cells overexpressing *CRNDE204* by using RT-qPCR and also to validate if knocking down of these binding partners affects the differentiation towards mature erythrocytes and megakaryocytes.

To test the knockdown efficiency of shRNA knockdown in K562 cells with *CRNDE204* overexpression, we transfected the cells with the shRNA plasmids, sorted the cells that took up the plasmid, and extracted RNA to perform a RT-qPCR to quantify the expression of mRNA after knockdown. The shRNAs against PUS1 showed an average of 80% knockdown efficiency (Figure 9A). All shRNAs against NR1H3 showed an average of 70% knockdown efficiency (Figure 9B). RPL22 shRNA1 showed around 75% knockdown efficiency, but shRNA2 and shRNA3 showed only around 30% knockdown efficiency (Figure 9C). NPM2 shRNA1 and shRNA2 showed an average of 50% knockdown efficiency, whereas shRNA3 had an average of 20% knockdown efficiency (Figure 9D). Taken together, all three shRNA constructs against the protein targets can be considered effective at the mRNA level.

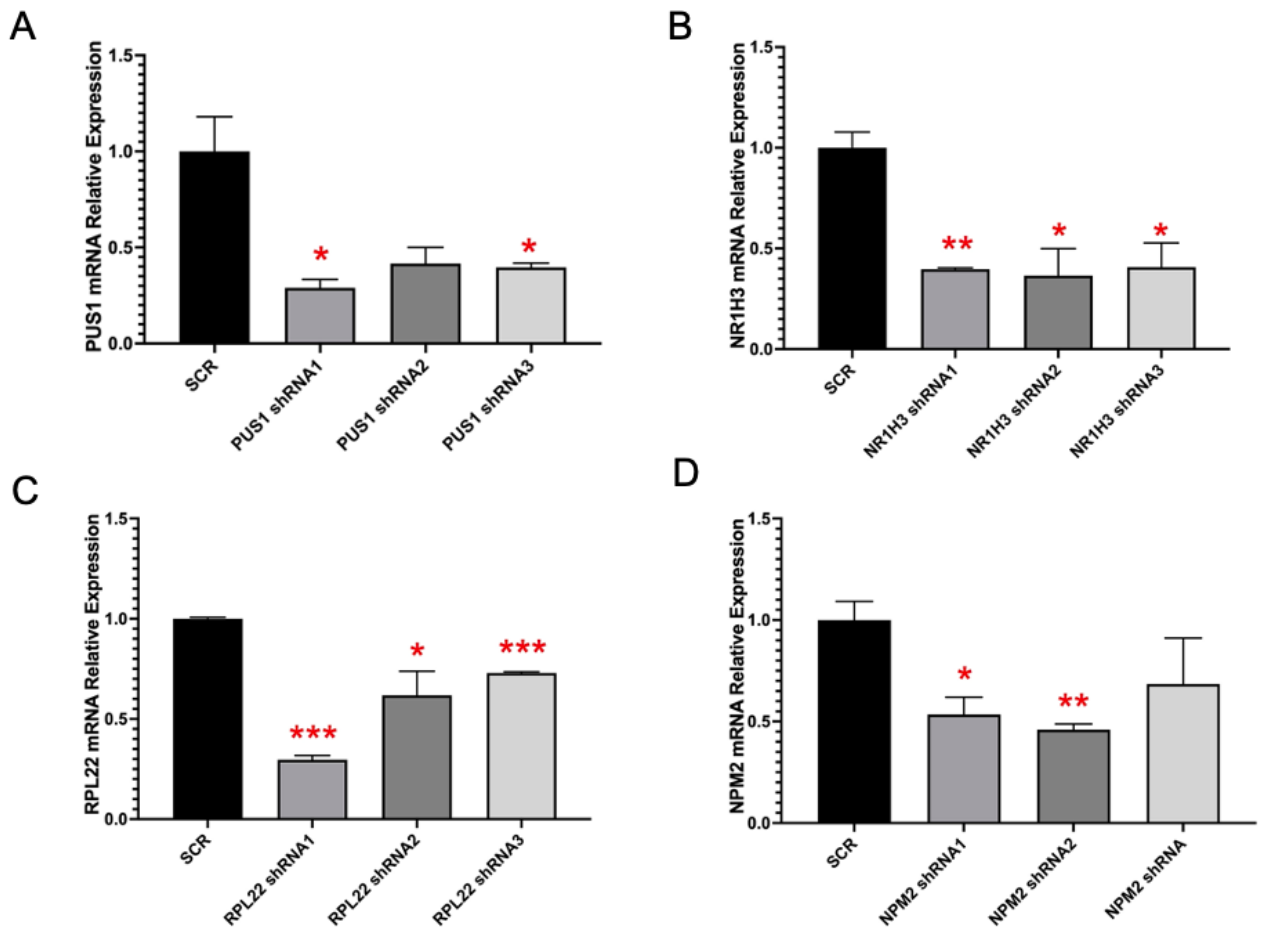


Figure 9: Validation of shRNA knockdown efficiency by qPCR. K562 cells overexpressing CRNDE204 after transfection with shRNA plasmids against PUS1, NR1H3, RPL22 and NPM2 for 24 hours before sorting. Total RNA was extracted from the sorted cells and the expression of PUS1, NR1H3, RPL22 and NPM2 mRNA was assessed by qPCR. Two technical replicates were performed for each condition, and the bar graph presented the mean fold change with SD. Significance was calculated by unpaired two-tailed t-test. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$

To evaluate the phenotypic effect of shRNA knockdown, we transfected the shRNA constructs and did erythroid or megakaryocyte differentiation assays to see the phenotypic effect after the knockdown, which is further quantified by fold change comparing with the scramble control. Erythroid differentiation assay showed that knockdown of all the binding partners can inhibit erythroid differentiation compared with the scramble control, as seen in the decrease of

fold change in CD235a fluorescence intensity. This result suggested all of the partners could be promoting erythropoiesis (Figure 10).

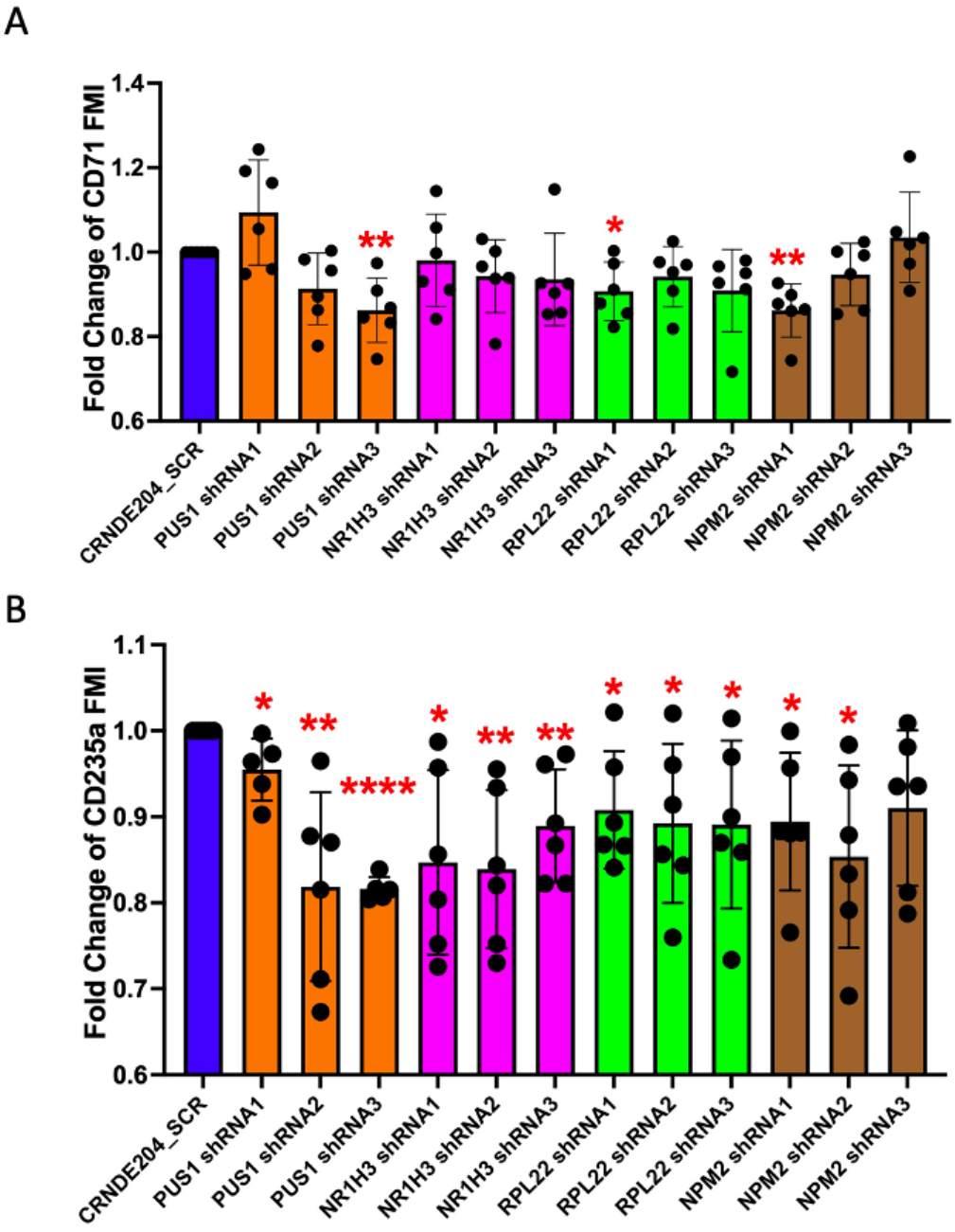


Figure 10: The effect of PUS1, NR1H3, RPL22 and NPM2 knockdown in erythroid differentiation. Three different shRNA constructs of each protein target were designed. Erythroid differentiation was used to assess the knockdown efficiency and the cells were analyzed by flow cytometry to compare the FMI of CD71 and CD235a. (A) Bar graph shown the fold change of the FMI of CD71 compared with the scramble control. (B) Bar graph shown the fold change of the FMI of CD235a compared with the scramble control. Significance was calculated by paired two-tailed t-test. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$.

For megakaryocyte differentiation, we observed a significant decrease in differentiation as measured by CD41/CD61 FMI after knocking down using PUS1 shRNA2. No significant change in CD41/CD61 FMI with NR1H3 knockdowns. In contrast, an increase in differentiation was observed in shRNA knockdown of RPL22 using shRNA2 and shRNA3. In knocking down of NPM2, shRNA1 showed significant knockdown effect, whereas shRNA2 and shRNA3 promoted megakaryocyte differentiation (Figure 11). This suggests PUS1 binds to *CRNDE204* and might be a positive regulator of myelopoiesis, whereas NR1H3 only binds to *CRNDE204* but does not have a direct effect in regulating myelopoiesis. On the other hand, current observation of RPL22 and NPM2 knockdown suggests these two binding partners might not regulate myelopoiesis directly.

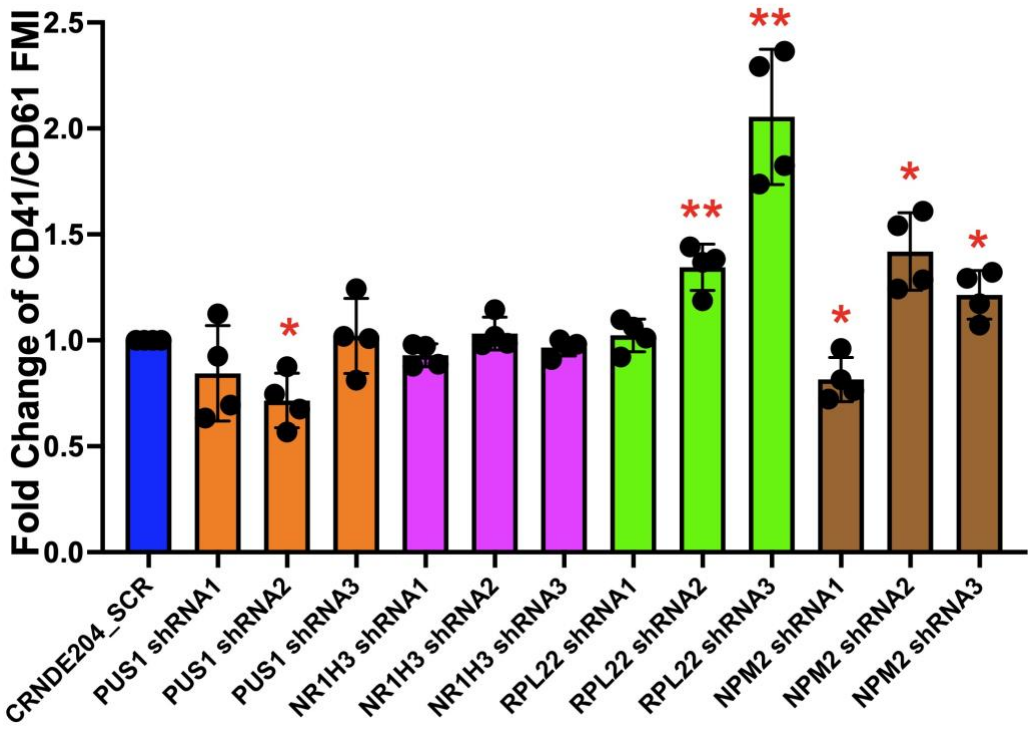


Figure 11: The effect of PUS1, NR1H3, RPL22 and NPM2 knockdown in megakaryocyte differentiation. Three different shRNA constructs of each protein target were designed. Megakaryocyte differentiation assay was used to assess the knockdown efficiency and the cells were analyzed by flow cytometry to compare the

FMI CD41/CD61. Bar graph shown the fold change of the FMI of CD41/CD61 compared with the scramble control. Significance was calculated by paired two-tailed t-test. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$.

Taken together, our results on shRNA knockdown and differentiation assays showed knocking down all the protein binding partners had significant effect in inhibiting erythroid differentiation. However, only PUS1 was shown to be the crucial regulator for myelopoiesis by phenotypic assays. Based on the observations, we hypothesize *CRNDE204* binds to PUS1 and this binding relationship regulates erythropoiesis and myelopoiesis directly.

Chapter 3: Discussion and Perspectives

3.1 Discussion

LncRNA *CRNDE*, originally identified in colorectal cancer, has been reported to involved in many types of cancer. However, little research has been done on identify the difference between *CRNDE* isoforms and their roles in cancer. We found the expression of *CRNDE* isoforms are silenced in SDS patients, and among these isoforms we identified *CRNDE201* and *CRNDE204*. Hematopoietic differentiation assays have shown *CRNDE204* can promote differentiation and proliferation towards mature erythrocytes and megakaryocytes, whereas *CRNDE201* suppress differentiation and proliferation. We proposed the role of *CRNDE201* and *CRNDE204* differs because the difference in their splicing pattern and structure. Based on Venna fold structure prediction, we saw *CRNDE204* has a unique structure which may include a specific binding motif for transcription factors such as PUS1 and NR1H3. *CRNDE201* lack this minimum binding motif, which cause it to lose ability to interact with important transcription factors in hematopoiesis. RNA-seq and IPA further support *CRNDE204* regulates erythropoiesis and myelopoiesis in several ways, and NR1H3 is identified as upstream regulator as important as GATA1. Our Y3H and DATNA together identified PUS1 as erythropoiesis regulator and myelopoiesis regulator under *CRNDE204* transcription pathway.

PUS1 is a protein that belongs to pseudouridine synthase (PUS) family proteins, and recent research has unveiled multiple functions of PUS family proteins beyond post-transcriptional modification of uridine. A recent study by Martinez et.al identified PUS family pseudouridine synthases, including PUS1, PUS7 and RPUSD4, is responsible for pseudouridination of pre-mRNAs co-transcriptionally. The positions of pseudouridnine modification were enriched near splicing sites and overlapped with the binding sites of multiple

RNA binding proteins, which suggested the role of PUS proteins in regulating alternative splicing[83]. Another study by Guzzi et.al. pointed out the role of PUS7-mediated pseudouridination in regulating translation in hESCs and also identified a potential binding motif on multiple types of RNAs, including tRNAs and non-coding RNAs. Further research revealed that PUS7 mediated pseudouridination is also important in regulating hematopoietic stem cell commitment, and loss of PUS7 is found in MDS with chromosome 7 defects, which leads to severe differentiation defects, significant decrease in BFU-E and increase of dysplastic cells[84].

Splice factor mutations are commonly seen in hematological disorders as well as other cancer types, and we speculated mutation in spice factors can drive isoform switch in *CRNDE*. The splice factor SF3B1 encodes for subunit 1 of splicing factor 3b protein complex, and it forms U2 snRNP with splicing factor 3a complex and 12s RNA[85]. The function of this complex is to bind to pre-mRNA splicing sites and regulate splicing event. Mutation in splicing factors is very common in hematopoietic malignancies such as AML and MDS. Recent study by Bamopoulos et al has discovered mutations in SRSF2, SF3B1 and U2AF1 affect the overall survival of AML patients. Mutation of SF3B1 is also found in non-proliferative hematopoietic disorders such as sideroblastic anemia (SA), a type of hematopoietic disorder categorized by the presence on ring sideroblasts caused by iron disposition in mitochondria[86]. In particular, there are multiple evidence showing the possibility of *CRNDE* isoform switching in cancers, specifically those with SF3B1 mutation. Magurie et.al. found out that SF3B1 K700E mutation in breast cancer disrupted the normal splicing of multiple RNAs, including *CRNDE*[87]. Similarly, Hintzsche et al revealed SF3B1 R625 mutation found in mucosal melanoma also caused alternative splicing of *CRNDE*[88]. Interestingly, the *CRNDE* splice variant under SF3B1 mutation state has truncated exon 4, which resembles the structure of *CRNDE201*.

We propose the balance between *CRNDE201* and *CRNDE204* is controlled by the splicing factors and mutation in splicing factor may disrupt the balance between the production of two isoforms and may favor the production of one isoform over the other, as SF3B1 mutations tend to favor the production of *CRNDE201* over *CRNDE204* and cause anemic symptoms. We propose reconstitution of *CRNDE204* can reverse the disease phenotype and can potentially cure the disease. Moreover, overexpression of *CRNDE204* does not cause phenotypic change in wild type cells, which makes *CRNDE204* as a potential target for gene therapy. We believe reintroducing *CRNDE204* could be a promising treatment for MDS patients to rescue the bone marrow failure symptoms.

3.2 Limitations

This study utilized the K562 cell line for study the effect of *CRNDE204* in regulating hematopoiesis. Although K562 cell line is widely used in study the development of erythroid and myeloid cells, the intrinsic mutations within the cell line can change the transcriptomics of the cells comparing with the normal bone marrow progenitor cells. RNA-seq and protein-seq data of K562 cells and K562 cells overexpressing *CRNDE* isoforms had differential expression of genes, which can mask the effect of knockdown efficiency in shRNA knockdown experiments.

Another limitation of this study is to the lack of a SDS mouse model to study the effect of *CRNDE* isoforms *in vivo*. This can be due to the difficulty in targeting Sbds gene in mouse. Complete knock-out of Sdbs gene in mouse is embryonic lethal, which suggests Sbds gene is crucial for early development[89]. Hematopoietic-specific Sbds knock-out mice had been designed, but the mice developed severe liver damage so was not suitable for further studies. Other

conditional knock-out models targeting the pancreas successfully recapitulated the pancreas dysfunction in SDS patients, but it failed to show hematopoietic symptoms[31]. Zambetti et. al. developed a model of SDS in mouse that can recapitulate the neutropenia symptoms in human patients by transplanting Sbds knockout myeloid progenitor cells in embryo to irradiated adult mouse. After transplantation, the adult mouse had severe neutropenia, which mimics the human phenotype[90]. Although this mouse model can be used to study the effect of Sbds knockdown in vivo, another challenge would be to overexpress *CRNDE* isoforms in the hematopoietic progenitor cells in the same mouse. Taking these factors together, it is still challenging to study the effect of *CRNDE* in vivo, and future study can focus on developing a reliable in vivo model system to study the effect of SBDS knockdown and the effect of overexpressing *CRNDE* simultaneously.

3.3 Perspectives and future directions

Based on our findings on the role of *CRNDE204* overexpression in K562 cells, we proposed a model (Figure 12) of *CRNDE204* expression in normal and abnormal hematopoiesis in SDS model. *SBDS* mutation in SDS patients could silence the expression of *CRNDE204*, which caused decreased differentiation and proliferation of hematopoietic cells. When reintroducing *CRNDE204* to *SBDS* knock-out cells, the differentiation and proliferation of hematopoietic cells returns back to normal. Because the expression of *CRNDE* is upregulated in several cancers, it will be worth investigating the effect of overexpressing *CRNDE* in healthy cells to see if *CRNDE* overexpression can cause hyperproliferation phenotype. Future investigation can utilize human primary cell lines (CD34+ progenitor cells) or an *ex vivo* system

to see overexpressing *CRNDE204* binding partners will have an effect on erythroid or myeloid differentiation and to observe the safety dosage of *CRNDE* overexpression in normal cells.

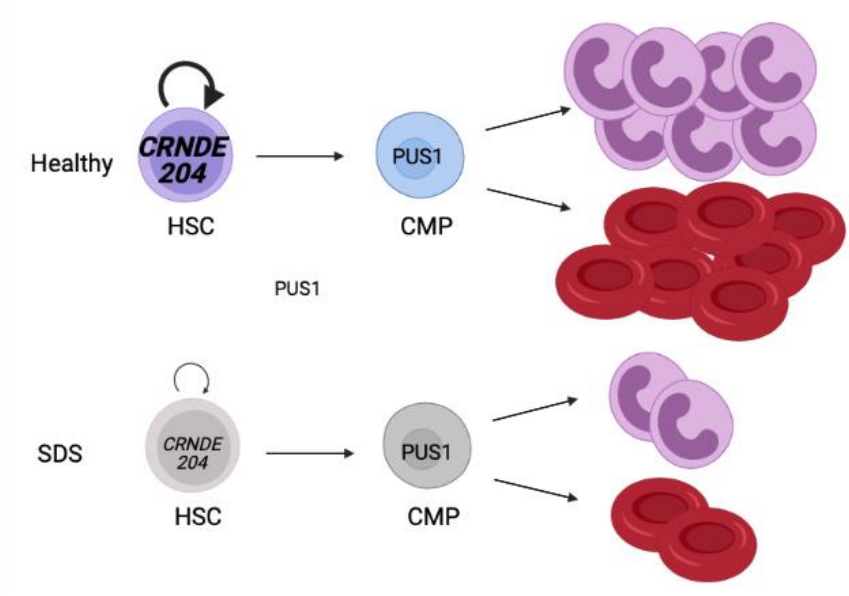


Figure 12: Proposed model of the role of *CRNDE204* in regulating hematopoiesis. The expression of *CRNDE204* is normal in healthy people, and it can bind to PUS1 to regulate the transcriptional programs involved in CMPs and promote the production of mature erythrocytes and myeloid cells. In SDS patients, *CRNDE204* is lowly expressed, and the regulatory network is disrupted, which results in anemia and neutropenia.

Additionally, we have identified PUS1 as the binding partner of *CRNDE204* that can regulate key erythropoiesis and myelopoiesis processes, and NR1H3 as the binding partner of *CRNDE204* that can regulate key erythropoiesis processes. However, the role RPL22 and NPM2 in regulating erythropoiesis and myelopoiesis and the relationship with *CRNDE204* still needs further investigation. We hypothesize the low expression level of RPL22 and NPM2 in K562 cells can make it difficult see the knockdown effect. Future research could focus on utilizing CD34+ cell line and observe the knockdown effect as a supplement to our current results in K562 cells.

From a clinical perspective, SDS patients have increased risk of developing myelodysplastic syndrome (MDS) or progression to AML, and the expression of lncRNA

CRNDE are upregulated in MDS and AML[91]. The reason why SDS and other ribosomopathies are associated with bone marrow failure with increased risk of developing hematological malignancy is still unclear. Future research could focus on understand the links between ribosomopathies and bone marrow failures, and the role of lncRNAs in regulating the process. Furthermore, the molecular mechanisms for the progression towards malignancies is also worth investigating. Understanding these mechanisms could provide more insights of treating ribosomopathies and allow clinicians to assess the risk of developing malignancies.

Chapter 4: Bibliography

1. Jagannathan-Bogdan, M. & Zon, L.I. Hematopoiesis. *Development* **140**, 2463-2467 (2013).

2. Ivanovs, A. *et al.* Human haematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323-2337 (2017).

3. Boiers, C. *et al.* Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell* **13**, 535-548 (2013).

4. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897-906 (1996).

5. Pucella, J.N., Upadhaya, S. & Reizis, B. The Source and Dynamics of Adult Hematopoiesis: Insights from Lineage Tracing. *Annu Rev Cell Dev Biol* **36**, 529-550 (2020).

6. Zivot, A., Lipton, J.M., Narla, A. & Blanc, L. Erythropoiesis: insights into pathophysiology and treatments in 2017. *Mol Med* **24**, 11 (2018).

7. Gregory, C.J. & Eaves, A.C. Human marrow cells capable of erythropoietic differentiation in vitro: definition of three erythroid colony responses. *Blood* **49**, 855-864 (1977).

8. Lee, S.H. *et al.* Isolation and immunocytochemical characterization of human bone marrow stromal macrophages in hemopoietic clusters. *J Exp Med* **168**, 1193-1198 (1988).

9. Gifford, S.C., Derganc, J., Shevkoplyas, S.S., Yoshida, T. & Bitensky, M.W. A detailed study of time-dependent changes in human red blood cells: from reticulocyte maturation to erythrocyte senescence. *Br J Haematol* **135**, 395-404 (2006).

10. Hattangadi, S.M., Wong, P., Zhang, L., Flygare, J. & Lodish, H.F. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood* **118**, 6258-6268 (2011).

11. Gao, J., Chen, Y.H. & Peterson, L.C. GATA family transcriptional factors: emerging suspects in hematologic disorders. *Exp Hematol Oncol* **4**, 28 (2015).

12. Barbarani, G., Fugazza, C., Strouboulis, J. & Ronchi, A.E. The Pleiotropic Effects of GATA1 and KLF1 in Physiological Erythropoiesis and in Dyserythropoietic Disorders. *Front Physiol* **10**, 91 (2019).

13. Takai, J. *et al.* The Gata1 5' region harbors distinct cis-regulatory modules that direct gene activation in erythroid cells and gene inactivation in HSCs. *Blood* **122**, 3450-3460 (2013).

14. Crispino, J.D. & Horwitz, M.S. GATA factor mutations in hematologic disease. *Blood* **129**, 2103-2110 (2017).

15. Duddu, S., Chakrabarti, R., Ghosh, A. & Shukla, P.C. Hematopoietic Stem Cell Transcription Factors in Cardiovascular Pathology. *Front Genet* **11**, 588602 (2020).

16. Tsai, F.Y. & Orkin, S.H. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* **89**, 3636-3643 (1997).

17. Suzuki, M. *et al.* GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. *Genes Cells* **18**, 921-933 (2013).

18. Samsel, L. & McCoy, J.P., Jr. Imaging flow cytometry for the study of erythroid cell biology and pathology. *J Immunol Methods* **423**, 52-59 (2015).

19. Huang, H. & Cantor, A.B. Common features of megakaryocytes and hematopoietic stem cells: what's the connection? *J Cell Biochem* **107**, 857-864 (2009).

20. Bohnsack, K.E. & Bohnsack, M.T. Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J* **38**, e100278 (2019).

21. Robledo, S. *et al.* The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA* **14**, 1918-1929 (2008).

22. Kampen, K.R., Sulima, S.O., Vereecke, S. & De Keersmaecker, K. Hallmarks of ribosomopathies. *Nucleic Acids Res* **48**, 1013-1028 (2020).

23. Aspesi, A. & Ellis, S.R. Rare ribosomopathies: insights into mechanisms of cancer. *Nat Rev Cancer* **19**, 228-238 (2019).

24. Nakhoul, H. *et al.* Ribosomopathies: mechanisms of disease. *Clin Med Insights Blood Disord* **7**, 7-16 (2014).

25. Marcel, V., Nguyen Van Long, F. & Diaz, J.J. 40 Years of Research Put p53 in Translation. *Cancers (Basel)* **10** (2018).

26. Boocock, G.R. *et al.* Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet* **33**, 97-101 (2003).

27. Austin, K.M. *et al.* Mitotic spindle destabilization and genomic instability in Shwachman-Diamond syndrome. *J Clin Invest* **118**, 1511-1518 (2008).

28. Rujkijyanont, P. *et al.* SBDS-deficient cells undergo accelerated apoptosis through the Fas-pathway. *Haematologica* **93**, 363-371 (2008).

29. Ball, H.L. *et al.* Shwachman-Bodian Diamond syndrome is a multi-functional protein implicated in cellular stress responses. *Hum Mol Genet* **18**, 3684-3695 (2009).

30. Burwick, N., Coats, S.A., Nakamura, T. & Shimamura, A. Impaired ribosomal subunit association in Shwachman-Diamond syndrome. *Blood* **120**, 5143-5152 (2012).

31. Bezzerri, V. & Cipolli, M. Shwachman-Diamond Syndrome: Molecular Mechanisms and Current Perspectives. *Mol Diagn Ther* **23**, 281-290 (2019).

32. Nelson, A.S. & Myers, K.C. Diagnosis, Treatment, and Molecular Pathology of Shwachman-Diamond Syndrome. *Hematol Oncol Clin North Am* **32**, 687-700 (2018).

33. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488 (2007).

34. Marchese, F.P., Raimondi, I. & Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol* **18**, 206 (2017).

35. Hajjari, M. & Salavaty, A. HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol Med* **12**, 1-9 (2015).

36. McHugh, C.A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236 (2015).

37. Mercer, T.R., Dinger, M.E. & Mattick, J.S. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**, 155-159 (2009).

38. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**, 126-130 (2008).

39. Beltran, M. *et al.* A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* **22**, 756-769 (2008).

40. Liu, B. *et al.* A cytoplasmic NF-kappaB interacting long noncoding RNA blocks IkappaB phosphorylation and suppresses breast cancer metastasis. *Cancer Cell* **27**, 370-381 (2015).

41. Wang, P. *et al.* The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* **344**, 310-313 (2014).

42. Tang, Q. & Hann, S.S. HOTAIR: An Oncogenic Long Non-Coding RNA in Human Cancer. *Cell Physiol Biochem* **47**, 893-913 (2018).

43. Amodio, N. *et al.* MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *J Hematol Oncol* **11**, 63 (2018).

44. Ji, P. *et al.* MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**, 8031-8041 (2003).

45. Chen, Q., Zhu, C. & Jin, Y. The Oncogenic and Tumor Suppressive Functions of the Long Noncoding RNA MALAT1: An Emerging Controversy. *Front Genet* **11**, 93 (2020).

46. Kwok, Z.H., Roche, V., Chew, X.H., Fadieieva, A. & Tay, Y. A non-canonical tumor suppressive role for the long non-coding RNA MALAT1 in colon and breast cancers. *Int J Cancer* **143**, 668-678 (2018).

47. Luo, M. *et al.* Long non-coding RNAs control hematopoietic stem cell function. *Cell Stem Cell* **16**, 426-438 (2015).

48. Alvarez-Dominguez, J.R. *et al.* Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood* **123**, 570-581 (2014).

49. Paralkar, V.R. *et al.* Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood* **123**, 1927-1937 (2014).

50. Brazao, T.F. *et al.* Long noncoding RNAs in B-cell development and activation. *Blood* **128**, e10-19 (2016).

51. Kotzin, J.J. *et al.* The long noncoding RNA Morrbid regulates CD8 T cells in response to viral infection. *Proc Natl Acad Sci U S A* **116**, 11916-11925 (2019).

52. West, K.A. & Lagos, D. Long Non-Coding RNA Function in CD4(+) T Cells: What We Know and What Next? *Noncoding RNA* **5** (2019).

53. Zhou, Y., Zhu, Y., Xie, Y. & Ma, X. The Role of Long Non-coding RNAs in Immunotherapy Resistance. *Front Oncol* **9**, 1292 (2019).

54. Delas, M.J. *et al.* lncRNA requirements for mouse acute myeloid leukemia and normal differentiation. *Elife* **6** (2017).

55. Graham, L.D. *et al.* Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. *Genes Cancer* **2**, 829-840 (2011).

56. Ding, J. *et al.* Long noncoding RNA CRNDE promotes colorectal cancer cell proliferation via epigenetically silencing DUSP5/CDKN1A expression. *Cell Death Dis* **8**, e2997 (2017).

57. Ma, X. *et al.* Overexpressed long noncoding RNA CRNDE with distinct alternatively spliced isoforms in multiple cancers. *Front Med* **13**, 330-343 (2019).

58. Ellis, B.C., Molloy, P.L. & Graham, L.D. CRNDE: A Long Non-Coding RNA Involved in CanceR, Neurobiology, and DEvelopment. *Front Genet* **3**, 270 (2012).

59. Lin, M. *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One* **6**, e23356 (2011).

60. Zheng, J. *et al.* CRNDE affects the malignant biological characteristics of human glioma stem cells by negatively regulating miR-186. *Oncotarget* **6**, 25339-25355 (2015).

61. Ellis, B.C., Graham, L.D. & Molloy, P.L. CRNDE, a long non-coding RNA responsive to insulin/IGF signaling, regulates genes involved in central metabolism. *Biochim Biophys Acta* **1843**, 372-386 (2014).

62. Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391-403 e319 (2019).

63. Lozzio, B.B. & Lozzio, C.B. Properties and usefulness of the original K-562 human myelogenous leukemia cell line. *Leuk Res* **3**, 363-370 (1979).

64. Andersson, L.C., Jokinen, M. & Gahmberg, C.G. Induction of erythroid differentiation in the human leukaemia cell line K562. *Nature* **278**, 364-365 (1979).

65. Tetteroo, P.A., Massaro, F., Mulder, A., Schreuder-van Gelder, R. & von dem Borne, A.E. Megakaryoblastic differentiation of proerythroblastic K562 cell-line cells. *Leuk Res* **8**, 197-206 (1984).

66. Inoue, T., Swain, A., Nakanishi, Y. & Sugiyama, D. Multicolor analysis of cell surface marker of human leukemia cell lines using flow cytometry. *Anticancer Res* **34**, 4539-4550 (2014).

67. Villeval, J.L. *et al.* Erythroid properties of K562 cells. Effect of hemin, butyrate and TPA induction. *Exp Cell Res* **146**, 428-435 (1983).

68. Shelly, C., Petruzzelli, L. & Herrera, R. PMA-induced phenotypic changes in K562 cells: MAPK-dependent and -independent events. *Leukemia* **12**, 1951-1961 (1998).

69. Huang, R. *et al.* Megakaryocytic differentiation of K562 cells induced by PMA reduced the activity of respiratory chain complex IV. *PLoS One* **9**, e96246 (2014).

70. Mello, F.V. *et al.* Maturation-associated gene expression profiles during normal human bone marrow erythropoiesis. *Cell Death Discov* **5**, 69 (2019).

71. Ferrari, F. *et al.* Genomic expression during human myelopoiesis. *BMC Genomics* **8**, 264 (2007).

72. Tie, R. *et al.* Interleukin-6 signaling regulates hematopoietic stem cell emergence. *Exp Mol Med* **51**, 1-12 (2019).

73. Gunes, C. *et al.* Comparative RNAi Screens in Isogenic Human Stem Cells Reveal SMARCA4 as a Differential Regulator. *Stem Cell Reports* **12**, 1084-1098 (2019).

74. Schuldenfrei, A. *et al.* HMGA1 drives stem cell, inflammatory pathway, and cell cycle progression genes during lymphoid tumorigenesis. *BMC Genomics* **12**, 549 (2011).

75. SenGupta, D.J. *et al.* A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci U S A* **93**, 8496-8501 (1996).

76. Bamopoulos, S.A. *et al.* Clinical presentation and differential splicing of SRSF2, U2AF1 and SF3B1 mutations in patients with acute myeloid leukemia. *Leukemia* **34**, 2621-2634 (2020).

77. Li, Y. *et al.* Identification and characterization of lncRNA mediated transcriptional dysregulation dictates lncRNA roles in glioblastoma. *Oncotarget* **7**, 45027-45041 (2016).

78. Furuyama, K. & Kaneko, K. Iron metabolism in erythroid cells and patients with congenital sideroblastic anemia. *Int J Hematol* **107**, 44-54 (2018).

79. N, A.G. & Castrillo, A. Liver X receptors as regulators of macrophage inflammatory and metabolic pathways. *Biochim Biophys Acta* **1812**, 982-994 (2011).

80. Lavergne, J.P., Marzouki, A., Reboud, J.P. & Reboud, A.M. Reconstitution of the active rat liver 60 S ribosomal subunit from different preparations of core particles and split proteins. *FEBS Lett* **236**, 345-351 (1988).

81. Fahl, S.P., Wang, M., Zhang, Y., Duc, A.C. & Wiest, D.L. Regulatory Roles of Rpl22 in Hematopoiesis: An Old Dog with New Tricks. *Crit Rev Immunol* **35**, 379-400 (2015).

82. Li, J. *et al.* Nucleophosmin regulates cell cycle progression and stress response in hematopoietic stem/progenitor cells. *J Biol Chem* **281**, 16536-16545 (2006).

83. Martinez, N.M. *et al.* Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect splicing. *bioRxiv*, 2020.2008.2029.273565 (2020).

84. Guzzi, N. *et al.* Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. *Cell* **173**, 1204-1216 e1226 (2018).

85. Dolatshad, H. *et al.* Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* **29**, 1798 (2015).

86. Ducamp, S. & Fleming, M.D. The molecular genetics of sideroblastic anemia. *Blood* **133**, 59-69 (2019).

87. Maguire, S.L. *et al.* SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J Pathol* **235**, 571-580 (2015).

88. Hintzsche, J.D. *et al.* Whole-exome sequencing identifies recurrent SF3B1 R625 mutation and comutation of NF1 and KIT in mucosal melanoma. *Melanoma Res* **27**, 189-199 (2017).

89. Zhang, S., Shi, M., Hui, C.C. & Rommens, J.M. Loss of the mouse ortholog of the shwachman-diamond syndrome gene (Sbds) results in early embryonic lethality. *Mol Cell Biol* **26**, 6656-6663 (2006).

90. Zambetti, N.A. *et al.* Deficiency of the ribosome biogenesis gene Sbds in hematopoietic stem and progenitor cells causes neutropenia in mice by attenuating lineage progression in myelocytes. *Haematologica* **100**, 1285-1293 (2015).

91. Ma, X. *et al.* Oncogenic role of lncRNA CRNDE in acute promyelocytic leukemia and NPM1-mutant acute myeloid leukemia. *Cell Death Discov* **6**, 121 (2020).