EXPLORING NOVEL REGULATORY ROLES OF LONG NON-CODING RNAS ASSOCIATED WITH MYELOPROLIFERATIVE NEOPLASMS

WONG NONTHAPHAT

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University
Department of Health Technology and Informatics

Exploring novel regulatory roles of long non-coding RNAs associated with myeloproliferative neoplasms

WONG Nonthaphat

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

August 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text

_____(Signed)

WONG Nonthaphat (Name of student)

Abstract

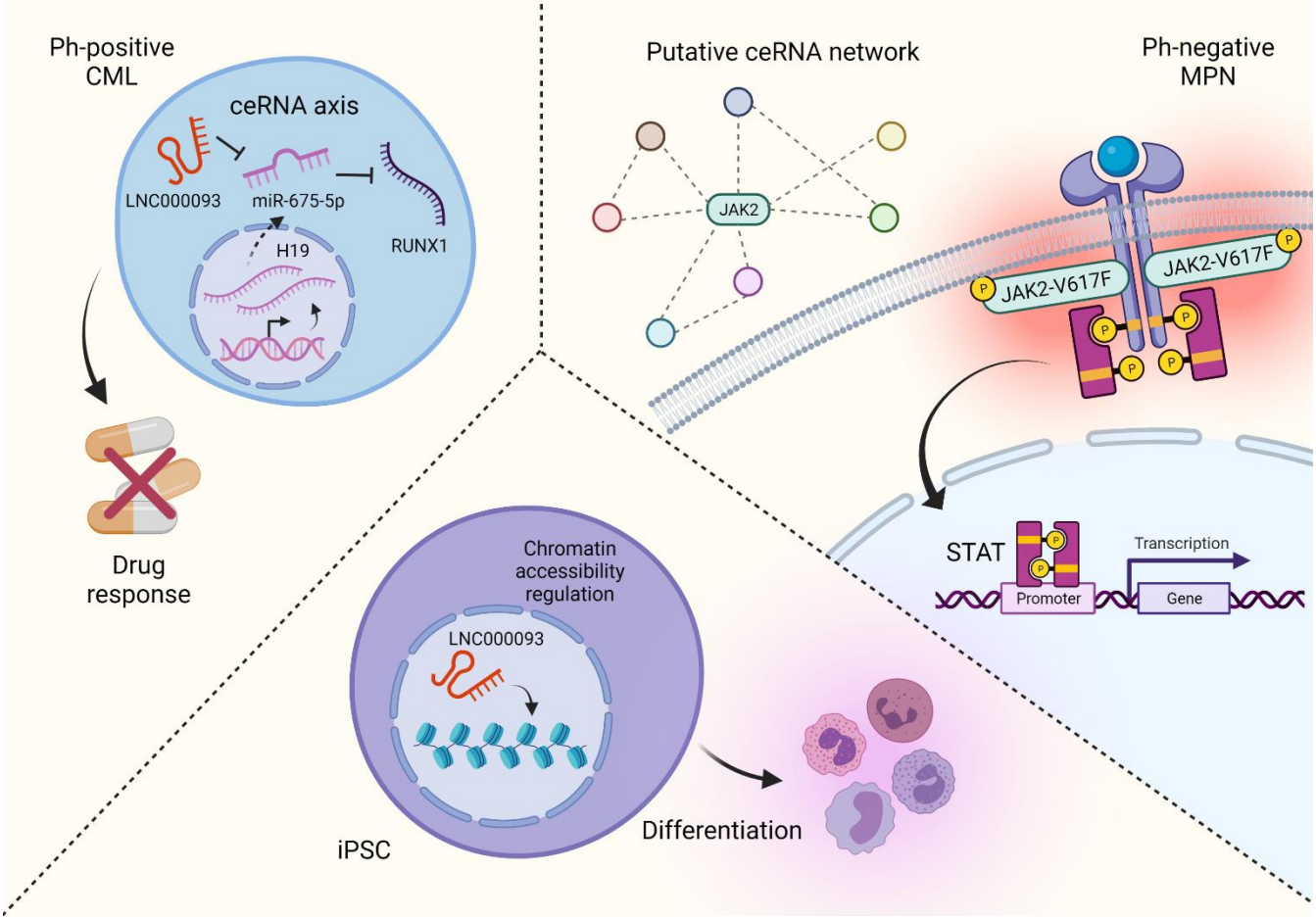
Myeloproliferative neoplasms (MPNs) are a group of heterogeneous diseases which primarily include chronic myeloid leukemia (CML), polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF). CML is characterized by the presence of Philadelphia chromosome (Ph) which encodes a dysregulated tyrosine kinase. In contrast, other Ph-negative MPNs are associated with notable mutations in *JAK2*, *MPL* and *CALR* genes with *JAK2*-V617F being the most prevalent one. In recent years, accumulating evidence has demonstrated the involvement of non-coding RNAs, especially long non-coding RNAs (lncRNAs), in the pathogenesis of various human diseases. However, the functional characteristics of these non-coding elements in MPNs are still largely unknown. The present study aimed to explore novel mechanisms that account for the regulatory roles of the lncRNA- and microRNA (miRNA)-axes in classical MPNs and drug response of CML.

To identify *JAK2*-V617F-associated lncRNAs, a qPCR array screening was performed using HEL cells. My data demonstrated that BANCR was the most downregulated lncRNA species after JAK2 inhibition. Subsequent experiments revealed its expression was specifically regulated by the *JAK2*-V617F signaling. Furthermore, RNA sequencing was performed to search for novel lncRNAs and explore their potential functions. Bioinformatics prediction analysis revealed putative interaction networks between the lncRNA, miRNA and mRNA in the JAK2 signaling that warrants further validation and investigation.

For the investigation of CML drug resistance, a novel lncRNA named LNC000093, which showed the largest downregulation in imatinib-resistant (IMR) K562 cells, was identified by RNA sequencing. Expression studies revealed a negative correlation between LNC000093 and H19/miR-675 expression, and subsequent *in silico* analysis as well as luciferase reporter assays demonstrated their interaction via direct binding. Furthermore, the potential interaction of H19/miR-675 and LNC000093 with RUNX1 was investigated to reveal their interrelationship. Taken together, my findings demonstrated that LNC000093 served as a competing endogenous RNA (ceRNA) to compete for miR-675-5p and indirectly regulate RUNX1 to mediate imatinib resistance of CML. Moreover, the potential regulatory role of LNC000093 in cell differentiation was also demonstrated using the induced pluripotent stem cell (iPSC) model. LNC000093 expression was substantially increased during iPSC differentiation, and the extent of differentiation was reduced after deletion of LNC000093 by CRISPR-Cas9. An altered chromatin accessibility profile of iPSCs was also revealed by scATAC-seq data analysis.

To conclude, this study has explored the molecular basis of MPNs in the non-coding area based on *in vitro* models that possess distinct genetic characteristics of Ph-positive or Ph-negative MPNs. Further translational research tools that can confirm the clinical relevance of my findings are required to enable the use of potential lncRNA biomarkers in the diagnosis and treatment of MPNs.

Graphical Abstract



For the investigation of Ph-negative MPNs, novel lncRNAs and putative interacting networks related to *JAK2*-V617F signaling were identified. In the study of CML drug resistance, the LNC000093-H19/miR-675-5p-RUNX1 ceRNA axis was shown to be involved in the drug response to imatinib. In iPSCs, LNC000093 functions in the nucleus and regulates chromatin accessibility, which in turn affects differentiation process.

Publications Arising from the Thesis

Research article

Wong, N. K., Luo, S., Chow, E. Y., Meng, F., Adesanya, A., Sun, J., Ma, H. M., Jin, W., Li, W., Yip, S. P., Huang, C. L. (2021). The Tyrosine Kinase-Driven Networks of Novel Long Non-Coding RNAs and Their Molecular Targets in Myeloproliferative Neoplasms. *Frontiers in Cell and Developmental Biology*, *9*, 643043.

Wong, N. K., Zhang, S., Wong, R. C., Yang, J. H., Gorospe, M., Yip, S. P., Huang, C. L. Long non-coding RNA LNC000093 Modulates Epigenetic Heterogeneity in Early Embryoid Bodies Differentiated from Induced Pluripotent Stem Cells. *Cells* (to be submitted)

Review article

Wong, N. K., Huang, C. L., Islam, R., & Yip, S. P. (2018). Long non-coding RNAs in hematological malignancies: translating basic techniques into diagnostic and therapeutic strategies. *Journal of Hematology & Oncology*, *11*(1), 131.

Conference abstract

Wong, N. K., Meng, F., Adesanya, A., Yip, S. P., & Huang, C. L. (2021). The Novel H19/miR-675-5p/LNC000093-Mediated Non-Coding RNA Pathway Regulates Treatment Response in BCR-ABL1-Positive CML Cells. *The FASEB Journal*, 35.

Lin, W.K., **Wong, N. K.**, Zhang, S., Wong, C. Y., Yip, S. P. & Huang, C. L. (2021). A Putative LncRNA-miRNA-mRNA Interaction Network Is Identified in Targeting JAK2-V617F-positive Myeloproliferative Neoplasms. 2021 IEEE International Conference on Bioinformatics and Biomedicine

Acknowledgments

I would like to express my sincere thanks to Dr. Huang Chien-Ling, my chief supervisor of this thesis study. Great thanks for her professional guidance and excellent supervision in the theoretical works. I am also grateful for her kind advice and encouragement throughout the research. I would also like to thank my co-supervisor Prof. Yip Shea Ping for his valuable and constructive suggestions which significantly improved my concepts in research works.

Besides, I am very grateful for all the support from my research team members, including Dr. Luo Shumeng, Dr. George Leung, Fei, Adenike, Lealem, Jiahong, Zhiwei, Shijing, Regina, Niaz, Petros, Lord and Riham. My thanks are also extended to all my friends and colleagues who helped and encouraged me a lot, including Dr. Henry Hung, Dr. Cara Cheung, Dr. Kelvin Wu, Dr. Marco Tam, Dr. Ashley Li, Dr. Annie Lee, Dr. Francine Au, Dr. Martin Yeung, Dr. Chen Singuang, Dr. Raymond Hui, Dr. Edith Cheung, Dr. Timothy Ho, Dr. Zou Xiang, Dr. Gilman Siu, Dr. Cesar Wong, Dr. Franki Tse, Dr. Helen Law, Rashidul, Leslie, Jocelyn, Jessie, Winnie, Edwin, Wenqi, Evelyn, Eden, Natalie, Bonnie, Roger, Chris, Elsie, Joey, Helia, Kelmond, Hazel, Eddie, Jake, Chloe, Timothy, Hiuyin, Ivan, Jack, Hennie, Pinky, Kaming, Li Yue, Gerald, Sunny, Kingston, Iris, David, Pamela and Yanyan.

Last but not least, I want to express my immeasurable gratitude to my family and my beloved companion Reese Wan for their unlimited support and love all the time. Thank you, my love.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3’-UTR	3’ untranslated region
A	Alanine
ABCC2	Adenosine triphosphate binding cassette subfamily C member 2
ABCG2	Adenosine triphosphate binding cassette subfamily G member 2
ABL1	Ableson leukemia virus
AGO	Argonaute
AKT	Protein kinase B
ALL	Acute lymphoblastic leukemia
Allo-SCT	Allogeneic stem cell transplantation
ALS	Amyotrophic lateral sclerosis
AML	Acute myeloid leukemia
ASOs	Antisense oligonucleotides
ASXL1	ASXL transcriptional regulator 1
ATAC-seq	Assay for transposase-accessible chromatin with high-throughput sequencing
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
AU-rich region	Adenylate-uridylate-rich region
BAD	BCL2 associated agonist of cell death
B-ALL	B-cell acute lymphoblastic leukemia
BALR-6	B-ALL associated long RNA-6
BANCR	BRAF-activated non-protein coding RNA
BCAR4	Breast Cancer Anti-Estrogen Resistance 4
Bcl-xl	B-cell lymphoma-extra large
BCR	Breakpoint cluster region
BCR-ABL1	BCR-ABL1 fusion gene
BET	Bromodomain and extraterminal
bFGF	Basic fibroblast growth factor
BMP4	Bone morphogenetic protein 4
BSA	Bovine serum albumin
CALR	Calreticulin
Cas9	CRISPR-associated protein 9
CBR3	Carbonyl Reductase 3
CBR3-AS1	CBR3 antisense RNA 1
CD34+	Cluster of differentiation 34 positive
CDKN1A	Cyclin-dependent kinase inhibitor 1A
cDNA	Complementary DNA
CEL-NOS	Chronic eosinophilic leukemia, not otherwise specified
ceRNA	Competing endogenous RNA

CHEK2	Checkpoint kinase 2
CLIP	Crosslinking immunoprecipitation
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
CML-CP	Chronic myeloid leukemia - chronic phase
CNBP	Cellular nucleic acid binding protein
CNL	Chronic neutrophilic leukemia
CO_2	Carbon dioxide
CRISPR	Clustered regularly interspaced short palindromic repeats
CSF3R	Colony stimulating factor 3 receptor
CUX1	Cut like homeobox 1
CXCR4	C-X-C chemokine receptor type 4
D	Aspartic acid
ddPCR	Droplet digital PCR
DDX43	DEAD-box helicase 43
DEGs	Differentially expressed genes
DNA	Deoxyribonucleic acid
DNase-seq	DNase I hypersensitive sites sequencing
DNMT1	DNA-methyltransferase 1
DNMT3A	DNA-methyltransferase 3A
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DTT	Dithiothreitol
DUSP1	Dual specificity phosphatase 1
E	Glutamic acid
E6/PVA	Essential 6/polyvinyl alcohol medium
EBs	Embryoid bodies
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGO	Eosinophil granule ontogeny
eIF4E	Eukaryotic initiation factor 4E
eNOS	Endothelial nitric oxide synthase
EPO	Erythropoietin
EPOR	Erythropoietin receptor
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
ESCs	Embryonic stem cells
ET	Essential thrombocythemia
EZH2	Enhancer of zeste 2 polycomb repressive complex 2 subunit
F	Phenylalanine
FBS	Fetal bovine serum
FDA	Food and drug administration
FLT3	Fms-like tyrosine kinase 3
Flt3L	Flt3-ligand

Fmrp	Fragile X mental retardation protein
FOXO	Forkhead box class O
FPKM	Fragments per kilobase of transcript per million mapped read
G	Glycine
GAB2	GRB2-associated binding protein 2
GADD45A	Growth arrest and DNA damage-induced 45 A
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GATA2	GATA-binding factor 2
G-CSF	Granulocyte colony-stimulating factor
G-CSFR	Granulocyte colony-stimulating factor receptor
gDNA	Genomic DNA
GFI1B	Growth factor independent 1B transcriptional repressor
GLI2	GLI Family Zinc Finger 2
GO	Gene Ontology
GRB2	Growth factor receptor bound protein 2
GTP	Guanosine-5'-triphosphate
H3K27me3	methylation of histone H3 at lysine 27
H3K4me3	methylation of histone H3 at lysine 4
HCl	Hydrochloric Acid
HDAC	Histone deacetylase
HEL	Human erythroleukemia
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hOCT1	Human organic cation transporter type 1
HOTAIR	HOX transcript antisense RNA
HOTAIRM1	HOXA transcript antisense RNA, myeloid-specific 1
HOX	Homeobox
HOXA	Homeobox A Cluster
HOXA-AS2	HOXA cluster antisense RNA 2
HOXC	Homeobox C cluster
HOXD	Homeobox D cluster
HRP-conjugated	Horseradish peroxidase-conjugated
HSC	Hematopoietic stem cells
Hsp90	heat shock protein-90
I	Isoleucine
IDH1	Isocitrate dehydrogenase 1
IGF-1	Insulin-like growth factor 1
IL-11	Interleukin 11
IL-3	Interleukin 3
IMP2	Inositol monophosphatase 2
INF-α	Interferon-alpha
iPSCs	Induced pluripotent stem cells
IRIS	International randomized study of interferon and STI571

JAK	Janus kinase
JAK2	Janus kinase 2
K	Lysine
K562-IMR	Imatinib-resistant K562 cells
K562-IMS	Imatinib-sensitive K562 cells
KCl	Potassium chloride
KD	Kinase domain
KDEL motif	Lysine-aspartic acid-glutamic acid-leucine motif
KEGG	Kyoto encyclopedia of genes and genomes database
KLF4	Kruppel-like factor 4
L	Leucine
LAMA84-IMR	Imatinib-resistant LAMA84 cells
LAST	LncRNA-assisted stabilization of transcripts
LINC	Long intergenic non-protein coding RNA
linc-MD1	Long intergenic non-protein coding RNA - muscle differentiation 1
lncRNA	Long non-coding RNA
LSD1	Lysine-specific demethylase 1
LUCAT1	Lung cancer associated transcript 1
MALAT1	Metastasis associated lung adenocarcinoma transcript 1
MAML1	Mastermind like transcriptional coactivator 1
MAPK	Mitogen-activated protein kinase
M-BCR	Major breakpoint cluster region
m-BCR	Minor breakpoint cluster region
Mcl-1	Induced myeloid leukemia cell differentiation protein
MDR1	Multidrug resistance protein 1
MDS	Myelodysplastic syndromes
MEF2C	Myocyte enhancer factor 2C
MEG3	Maternally expressed 3
MEK	MAPK/ERK kinase
MgCl2	Magnesium chloride
MIAT	Myocardial infarction associated transcript
miRNA	Micro-RNA
MM	Multiple myeloma
MPL	Thrombopoietin receptor
MPN	Myeloproliferative neoplasms
MPN-U	MPN, unclassifiable
MREs	miRNA response elements
mRNA	Messenger RNA
MRP1	Multidrug resistance-associated protein 1
MT	Mutated type
mTOR	Mechanistic target of rapamyci
MyoD	Myoblast determination protein 1
N	Asparagine

NaCl	Sodium chloride
NBR2	Neighbor of BRCA1 gene 2
N-cap	N-terminal cap
ncRNA	Non-coding RNA
NEAT1	Nuclear enriched abundant transcript 1
NF1	Neurofibromin 1
NFE2	Nuclear factor, erythroid 2
NF-YA	Nuclear transcription factor Y subunit α
NGS	Next generation sequencing
OCT4	Octamer-binding transcription factor 4
OSKM	OCT4-SOX2-KLF4-c-MYC
$P190^{BCR/ABL1}$	190kDa BCR-ABL1 protein
$P210^{BCR/ABL1}$	210kDa BCR-ABL1 protein
$P230^{BCR/ABL1}$	230kDa BCR-ABL1 protein
PABP	Poly(A)-binding protein
PAM site	Protospacer adjacent motif site
PANDA	P21-associated ncRNA DNA damage-activated
PBS	Phosphate-buffered saline
PCA3	Prostate cancer antigen 3
PCR	Polymerase chain reaction
PDGFR	Platelet-derived growth factor receptor
Ph	Philadelphia
Ph^+ MPNs	Philadelphia-positive MPNs
PI3K	Phosphoinositide 3-kinase
PKC	Protein kinase C
PMF	Primary myelofibrosis
PMSF	Phenylmethylsulfonyl fluoride
PNUTS	Phosphatase 1 NUclear Targeting Subunit
PPARγ	Peroxisome proliferator-activated receptor γ
PRAL	P53 regulation associated lncRNA
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PSA	Prostate-specific antigen
PV	Polycythemia vera
PVA	Polyvinyl alcohol
PVDF	Polyvinylidene fluorid
Rac	Ras-related C3 botulinum toxin substrate
RBP	RNA-binding protein
Rho/GEF	Ras homolog gene family/ Guanine nucleotide exchange factors
RIN	RNA integrity number
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid

RNA-seq	RNA Sequencing
RNP	Ribonucleoprotein
rRNA	Ribosomal RNA
RT	Reverse transcription
RUNX1	RUNX family transcription factor 1
S	serine
scATAC-seq	Single cell assay for transposase-accessible chromatin with high-throughput sequencing
SCF	Stem cell factor
SD	Standard deviation
SDF-1	Stromal-derived-factor-1
SEM	Standard error of the mean
SETBP1	SET binding protein 1
SF3B1	Splicing factor 3b subunit 1
sgRNA	Single-guide RNA
SH	SRC homology domains
SHP2	Protein tyrosine phosphatase-2
SMARCB1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1
SMD	Staufen 1-mediated mRNA decay
SNHG5	Small nucleolar RNA host gene 5
SNHG6	Small nucleolar RNA host gene 6
SNIP1	Smad nuclear-interacting protein 1
SORL1	Sortilin related receptor 1
SOS	Son of Sevenless
SOX2	Sex determining region Y-box 2
SRA	Steroid receptor RNA activator
SRSF2	Serine and arginine rich splicing factor 2
STAT	Signal transducer and activator of transcription
STAU1	Staufen 1
SWI/SNF	Switch/sucrose nonfermentable
T	Threonine
TET2	Tet methylcytosine dioxygenase 2
TFR	treatment-free remission
THSseq	Transposome hypersensitive sites sequencing
TINCR	Terminal differentiation-induced ncRNA
TKI	Tyrosine kinase inhibitor
TP53	Tumor protein P53
TPO	Thrombopoietin
TrxG/MLL	Trithorax group/MLL protein complexes
TSS	Transcription start sites
TUNA	Tcl1 Upstream Neuron-Associated lincRNA
U2AF1	U2 small nuclear RNA auxiliary factor 1

UCA1	Urothelial cancer associated 1
V	Valine
VEGF	Vascular endothelial growth factor
W	Tryptophan
WHO	World Health Organization
WT	Wild-type
Xist	X-inactive specific transcript
Y177	Tyrosine 177
ZRSR2	Zinc finger CCCH-Type, RNA binding motif and serine/arginine rich 2
μ-BCR	Micro-breakpoint cluster region

Chapter 1 – Introduction

1.1 Myeloproliferative neoplasms

Myeloproliferative neoplasms (MPNs) are a heterogeneous group of clonal hematopoietic disorders, characterized by overproduction of one or more myeloid lineages (Barbui et al., 2018). The mutated hematopoietic stem cells harboring driver mutation expand clonally and undergo maturation to rise clonal of cells with maturation. According to the World Health Organization (WHO) classification system for tumors of the hematopoietic and lymphoid tissue published in 2016, MPNs consist of 7 subcategories including polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (PMF), chronic myeloid leukemia (CML), chronic neutrophilic leukemia (CNL), chronic eosinophilic leukemia, not otherwise specified (CEL-NOS) and MPN, unclassifiable (MPN-U) (Barbui et al., 2018). Among these 7 subcategories, PV, ET and PMF, collectively known as classical MPNs, and CML have their driver mutations being discovered in the past decades.

1.2 Classical MPNs

Classical MPNs is a collective name for PV, ET and PMF, which is the most frequent MPN subcategory. Like other MPN entities, classical MPNs are featured with excessive production of fully functional mature blood cells. PV is characterized by over-production of mainly erythroid lineage and can also associate with different degrees of hyperplasia in megakaryocytic and granulocytic lineage (Barbui et al., 2018). The annual incidence rate is around 0.84/100,000 with a 14 to 20 years median survival (Crisa et al., 2010; Passamonti et al., 2010; Tefferi et al., 2013). ET is characterized by over-production of mainly megakaryocytic lineage (Barbui et al., 2018). It is with an annual incidence rate of 1.03/100,000 (Titmarsh et al., 2014). ET patients usually have a pretty good prognosis and a long median survival of 20 years (Girodon et al., 2010;

Barbui et al., 2011). However, PMF, when compared to PV and ET, is a more heterogeneous and aggressive disorder with the essential features of bone marrow fibrosis and megakaryocytic hyperplasia. The median survival of PMF patients is much shorter and drops to around 2 to 6 years (Cervantes et al., 2009; Cervantes et al., 2012; Anderson and McMullin, 2014; Byun et al., 2017). Although PMF is with the worse prognosis among classical MPNs, the annual incidence rate is the lowest and down to 0.47/100,000 (Titmarsh et al., 2014). Despite the differences in clinical presentations, precise diagnosis of the three entities is quite challenging. This is attribute to the continuum and unclear boundaries between the 3 entities, especially between ET and PMF (Vainchenker and Kralovics, 2017; Barbui et al., 2018). Extensive studies have been done on genetic basis of classical MPNs. Until now, there are driver mutations being found in 3 genes - Janus kinase 2 (*JAK2*), thrombopoietin receptor (*MPL*) and calreticulin (*CALR*) (James et al., 2005; Pikman et al., 2006; Scott et al., 2007; Nangalia et al., 2013; Vainchenker and Kralovics, 2017). They are almost mutually exclusive to each other (Jia and Kralovics, 2020). These mutations all cause cytokine-independent constitutive activation of the signal transducer and activator of transcription (STAT) molecules and its downstream pathways that results in the transcription of different oncogenes (Grinfeld et al., 2017; Vainchenker and Kralovics, 2017). PI3K, mTOR and MAPK-related pathways are found to be activated as well (Levine et al., 2007).

1.2.1 *JAK2* mutations

The genetic basis of classical MPNs have long been unclear until a breakthrough discovery found in 2005 (James et al., 2005). A guanine-to-thymine somatic point mutation, namely *JAK2*-V617F, have been discovered at nucleotide 1894 in exon 14 of *JAK2* (James et al., 2005; Kralovics et al., 2005; Levine et al., 2005). The alteration in nucleobase results in missense mutation that changes the original valine (V) to phenylalanine (F) at position 617 of JAK2 protein. JAK2 kinase is a tyrosine kinase located in cytoplasm and is responsible for signal transduction of hematopoietic cytokine receptors (Lu et al., 2005). Erythropoietin (EPO) receptor (EPOR), granulocyte colony stimulating factor (G-CSF) receptor (G-CSFR) and MPL regulate production of hematopoietic cells in erythroid, granulocytic or megakaryocytic lineages respectively (Lu et al., 2005; Lu et al., 2008; Sangkhae et al., 2014). When these receptors interact with their ligands – EPO, G-CSF and thrombopoietin (TPO), the receptors dimerize and lead to auto- and trans-phosphorylation of JAK2 and the receptors themselves (Lu et al., 2005; Lu et al., 2008; Vainchenker and Kralovics, 2017). The activated JAK2 is then interact with downstream substrates including STAT and leads to transcriptional induction of targeted genes. However, the *JAK2*-V617F mutation dysregulate these processes (Saharinen et al., 2000; James et al., 2005). The mutation results in a missense mutation and leads to the valine-to-phenylalanine change of amino acid at position 617 (*JAK2*-V617F) located in the pseudokinase domain of JAK2. *JAK2*-V617F mutation results in loss of the autoinhibitory effect of the pseudokinase domain (Kralovics et al., 2005; Levine et al., 2005). Thus, the JAK2 kinase is instinctively activated in a cytokine-independent manner and then the activated JAK2 kinase in turn activates the downstream STAT pathway, leading to over-transcription of target genes and thus the disease phenotypes (Kralovics et al.,

2005; Levine et al., 2005; Vainchenker and Kralovics, 2017). Among all of the classical MPN cases, around 70% of them harbor *JAK2*-V617F mutation. It can be detected in 95% of PV and 60% of ET and PMF patients (Tefferi et al., 2009; Geyer and Orazi, 2016; Barbui et al., 2018).

Apart from *JAK2*-V617F mutation, mutations have been found in *JAK2* exon 12 later in 2007 (Scott et al., 2007). Most of these mutations are small in-frame indel affecting residues between 537 and 543 (Scott et al., 2007; Jia and Kralovics, 2020). This region is within the linker between the SH2 domain and pseudokinase domain. Like *JAK2*-V617F, *JAK2* exon 12 mutations result in constitutive activation of JAK2 kinase and thus the downstream STAT pathway and other related signaling cascades including PI3K/AKT pathway and MAPK/ERK pathway. Despite both *JAK2*-V617F and *JAK2* exon 12 mutations affecting similar pathways, *JAK2* exon 12 mutations are only harbored by 5% PV patients but not ET or PMF patients (Scott et al., 2007). They pose a more significant effect on erythroid lineage and tend to give a mild degree of leukocytosis and thrombocytosis (Passamonti et al., 2011). Although JAK2 mutations can constitutively activate various pathways, how it leads to particular phenotypes of particular entity is still uncertain (Jang and Choi, 2020).

1.2.2 *MPL* mutations

MPL is a TPO receptor mainly expressed in hematopoietic stem cells (HSC) and megakaryocytic lineage. Two major types of *MPL* mutations have been reported in *MPL* exon 10. Mutations at codon 515 located on the cytosolic domain has been reported to be the most frequent one (Pikman et al., 2006). Tryptophan (W) at position 515 is most commonly changed to leucine (*MPL*-W515L) or lysine (*MPL*-W515K)

although substitution of W with arginine (*MPL*-W515A), alanine (*MPL*-W515A) and glycine (*MPL*-W515G) have also been found (Pikman et al., 2006; Defour et al., 2016). Generally, the transmembrane domain of wild-type (WT) MPL prohibit MPL itself from autoactivation (Staerk et al., 2006). However, the mutation at 515 position leads to loss of the prohibition property and results in spontaneous MPL activation (Pikman et al., 2006). The other major mutation is substitution of serine at position 505 with asparagine (*MPL*-S505N) which has a much lower frequency than mutations at codon 515 (Ding et al., 2004; Passamonti et al., 2011). In general, *MPL* mutations are harbored by around 3% of ET patients and 5% of PMF patients (Pardanani et al., 2006).

1.2.3 *CALR* mutations

In spite of the discovery of mutations in *JAK2* and *MPL*, there were still a portion of MPN cases not being explained by any genetic mutations. Another breakthrough discovery has been reported in 2013 (Klampfl et al., 2013; Nangalia et al., 2013). The -1/+2 frameshift mutations in *CALR* exon 9 have been described to be pathogenically associated with ET and PMF free of any *JAK2* or *MPL* mutations. CALR is a calcium-binding chaperone protein in endoplasmic reticulum (ER) that regulates calcium homeostasis and controls quality of newly synthesized glycoproteins (Michalak et al., 2009). The frameshift mutations in CALR lead to a brand-new C-terminal sequence, losing the negatively charged WT lysine-aspartic acid-glutamic acid-leucine (KDEL) motif (Balligand et al., 2016; Chachoua et al., 2016; Elf et al., 2016; Marty et al., 2016; Nivarthi et al., 2016). The negatively charged KDEL motif plays an important role in retaining CALR molecules in ER (Araki et al., 2016). However, with the less negatively charged C-terminal in mutant CALR, the mutant CALR cannot disassociate from MPL after binding with MPL in ER. Thus, the mutant CALR-bound MPL is exported to be

expressed on the cell surface (Staerk et al., 2006). This results in MPL activation independent of TPO as well as activation of JAK2/STAT pathways (Marty et al., 2016).

CALR mutations account for 20-25% of ET cases and around 25% of the PMF cases (Tefferi et al., 2014). There are two major types of *CALR* mutation: Type I is a 52 bp-deletion (c.1092_1143del, p.L367fs*46) and Type II is a 5 bp-insertion (c.1154_1155insTTGTC, p.K385fs*47). They account for around 80% of the CALR mutated MPN cases (Rumi et al., 2014a; Rumi et al., 2014b). Type I *CALR* mutation results in loss of most part of exon 9 and calcium binding sites while Type II mutation loses less part of the exon 9 and is capable of retaining half of the negative charge (Klampfl et al., 2013). Based on these two changes in CALR structure, the rest of *CALR* mutations are categorized into Type I-like and Type II-like where Type I-like mutations are more predominant than Type II-like (Pietra et al., 2016).

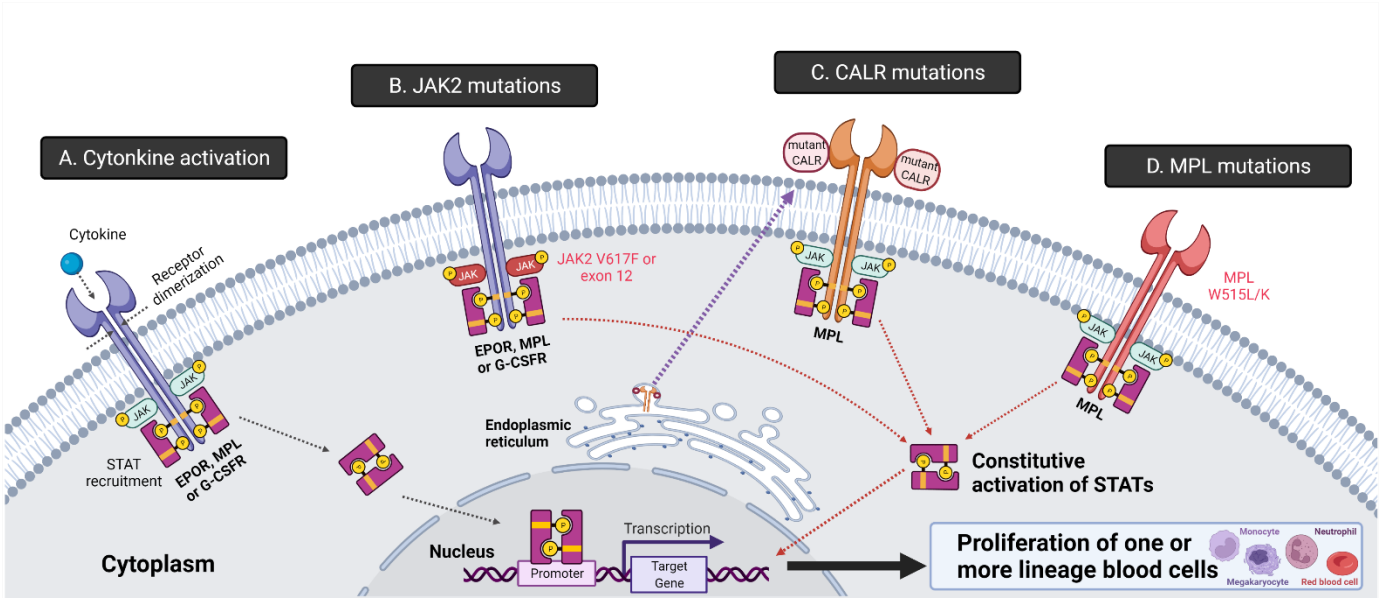


Figure 1.1 Constitutive activation of the JAK/STAT pathway resulted from mutated *JAK2, CALR, MPL*.

(A) General mechanism for cytokine activation of the JAK/STAT pathway. Ligand-receptor interaction results in dimerization of receptor subunits and transphosphorylation of JAK. The activated JAK then phosphorylates the receptors and STAT proteins. The phosphorylated STAT proteins are transported to the nucleus, resulting in their target gene transcription. Major somatic driver mutations of classical MPNs (B) *JAK2*, (C) *CALR* and (D) *MPL* result in JAK/STAT pathway activation. Figure adapted from Jang & Choi (2020).

1.2.4 Triple-negative MPN

The driver mutations described above are mutually exclusive and account for almost 100% of PV cases, ~90% of ET cases and ~95% of PMF cases (Rumi and Cazzola, 2017). Despite this, there are still around 10% of ET cases and 5% of PMF cases cannot be explained by any disease-driving mutations. These MPN cases fall into a subcategory of triple-negative MPN. This remains as a great unknown area of MPN development requiring more studies to discover the mechanisms behind.

1.2.5 Genetic factors associated with MPN heterogeneity

Despite the discovery of the driver *JAK2*, *MPL* and *CALR* somatic mutations, these driver somatic mutations are not able to give a complete explanation of the heterogeneity of classical MPNs (Cancer Genome Atlas Research et al., 2013; Papaemmanuil et al., 2016). With the development of next generation sequencing (NGS), there are several mutations that have been found to be associated with classical MPNs. Unfortunately, none of these mutations is specifically restricted to classical MPNs but can also be found in acute myeloid leukemia (AML) and myelodysplastic syndromes (MDS) (Cancer Genome Atlas Research et al., 2013; Lundberg et al., 2014). These mutations are not mutually exclusive. Indeed, accumulation of more mutations implies a worse prognosis and shorter median survival (Lundberg et al., 2014). The risk of leukemic transformation from MPNs increases as well. The cooperating mutations are usually found to be affecting genes crucial for epigenetic regulation (such as TET2, DNMT3A, ASXL1 and EZH2), mRNA splicing (such as SF3B1, SRSF2, U2AF1 and ZRSR2), transcription (such as TP53, CUX1, RUNX1 NFE2) and signal transduction (such as NF1, NRAS, KRAS and FLT3) (Stegelmann et al., 2010; Pasquier et al., 2014; Rampal et al., 2014). It seems there is a higher tendency of these mutations happening

in tumor suppressor genes and the mutations are usually loss of function. In conclusion, different combination of these cooperating mutations and driver mutations may result in different disease phenotypes (Vannucchi et al., 2013).

1.2.6 Mutant allele burden

JAK2-V617F is associated with PV, ET and PMF. However, how a single mutation can result in diseases with different phenotypes is still unclear. Mutant allele burden can be one of the possible explanations since it has been reported that the *JAK2*-V617F burden, which refers to the ratio of mutant to wild type *JAK2* in hematopoietic cells, varies between different classical MPN entities. The mutant allele burden is usually low in ET patients (~25%), median in PV patients (>50%), and highest in post-PV and post-ET myelofibrosis (~100%) (Rumi et al., 2014a).

1.2.7 MPN therapy

Since the symptoms and prognosis vary in different classical MPN entities, there are various treatment for classical MPNs. For example, since the median survival for low-risk PV and ET is quite long, the symptomatic treatments are usually prescribed. Low-dose aspirin and cytoreductive drugs like hydroxyurea are usually given to patients (Yogarajah and Tefferi, 2017). However, for PMF patients, since the disease is much more aggressive with worse prognosis and shortened median survival, more aggressive management is needed. Allogenic stem cell transplant can be an option for therapy (Lim et al., 2013). Before the transplantation, target drug therapy can be considered. Since *JAK2* mutation is the most frequent mutation in classical MPNs resulting in alteration of JAK/STAT pathway, JAK kinase inhibitor ruxolitinib has been approved for clinical use and can be prescribed to improve the survival rate and quality of life (Harrison et

al., 2012). Other promising JAK inhibitors like pacritinib and momelotinib that may cause less myelosuppression and reduced anemia were also subjected to clinical trials (Geyer and Mesa, 2014). However, monotherapy of JAK inhibitor does not benefit all the MPN patients and some patients could not receive such treatment due to cytopenia. The discovery of new molecular abnormalities in signaling pathways has led to the development of some non-JAK2 targeting agents such as histone deacetylase (HDAC) inhibitor, bromodomain and extraterminal (BET) inhibitor, heat shock protein-90 (Hsp90) inhibitor, etc. that have been under clinical trials and could be used alone or in combination with ruxolitinib (Geyer and Mesa, 2014; Economides et al., 2019).

1.3 Chronic myeloid leukemia

Chronic myeloid leukemia (CML), also known as Philadelphia (Ph)-positive MPNs (Ph^+ MPNs) or *BCR-ABL1*$^+$ MPNs, is a rare disease with an annual incidence of 1 to 2/100,000 (Sawyers, 1999). It accounts for around 15-20% leukemia cases and most commonly found in elderly population (Sawyers, 1999). CML is a three-phase disease with an initial chronic phase, followed by accelerated phase and the final blastic phase. Patients in chronic phase generally have no to mild symptoms and can remain in this phase for years (Soverini et al., 2018). When the disease progresses to accelerated and then blastic phase, the percentage of immature cells and the genetic instability increases. There is a higher possibility of drug resistance when the disease progresses (Sawyers, 1999; Shah, 2008).

To date, apart from morphological examination of blood cells, CML diagnosis is confirmed by molecular tests. The diagnostic hallmark harbored by CML cells was first described in 1960, that is the abnormally shortened Ph chromosome found in CML patients (Nowell, 1985). Thirteen years later in 1973 (Rowley, 1973), Ph chromosome

was found to be the result of a reciprocal chromosomal translocation between chromosome 9 and chromosome 22 – t(9,22) (Sessarego et al., 1983). Despite this, the exact genes involve in the t(9,22) translocation remained unclear until 1985 (Stam et al., 1985). It was reported that the proto-oncogene *ABL1* (Ableson leukemia virus) in chromosome 9 fuses with the *BCR* (breakpoint cluster region) gene in chromosome 22. This results in formation of *BCR-ABL1* fusion oncogene and production of a chimeric protein that leads to CML (Stam et al., 1985). Among all CML patients, 95% of them harbor the *BCR-ABL1* mutation, while the other 5% harbor a complex or variant translocations which result in the same *BCR-ABL1* mutation outcome (Sawyers, 1999).

1.3.1 *ABL1* and *BCR* gene

Human *ABL1* gene encodes a 145 kDa non-receptor tyrosine-protein kinase which is ubiquitously expressed (Laneuville, 1995). Depending on different alternative splicing of the first exon, 2 isoforms can be arisen (Laneuville, 1995). The ABL1 protein consists of several structure domains including a N-terminal "cap" (N-cap), 3 SRC homology domains (SH1 – SH3), 4 proline-rich regions, 3 nuclear localization signal domains, a nuclear exporting signal domain, a DNA-binding domain and actin-binding domains (Kipreos and Wang, 1992; McWhirter and Wang, 1993; Feller et al., 1994; Cohen et al., 1995; Deininger et al., 2000). The N-cap may or may not contain myristylation site depending on the alternative splicing (Laneuville, 1995). The tyrosine kinase activity of ABL1 protein is exhibited by SH1 domain, which is inhibited by the SH3 domain (Cohen et al., 1995). SH2 and SH3 domains are involved in the interaction with other proteins, while proline-rich regions involve in the interaction with other proteins containing SH3 domains (Cohen et al., 1995; Deininger et al., 2000). The native ABL1 protein regulates various biological processes such as cell growth and

apoptosis, oxidative stress and DNA damage response, cell adhesion and migration, integrin signaling, and so on (Soverini et al., 2018).

BCR gene encodes a 160 kDa protein carrying serine-threonine kinase property (Laneuville, 1995). The BCR protein is also ubiquitously expressed and consists of several structural domains including an oligomerization domain, serine/threonine kinase domains, a Ras homolog gene family/guanine nucleotide exchange factors (Rho/GEF) domain, a calcium binding domain and a Rac-GTPase activating domain (Diekmann et al., 1991; McWhirter et al., 1993; Denhardt, 1996; Deininger et al., 2000). The native BCR protein regulates neuronal RAC1 activity, macrophage motility and phagocytosis as well as keratinocyte adhesion and differentiation (Cho et al., 2007; Dubash et al., 2013).

1.3.2 *BCR-ABL1* fusion gene and chimeric protein

The translocation breakpoints of both *ABL1* and *BCR* gene are not in particular exact hotspot locations, but anywhere within a region (Melo, 1996; Deininger et al., 2000; Soverini et al., 2018). The translocation breakpoints of *ABL1* gene mostly fall in a large region ranging from upstream of the first exon 1b to downstream of second exon 1a (Melo, 1996). The breakpoints usually locate in the intron region. For *BCR* gene, the translocation breakpoints mostly locate in either the minor breakpoint cluster region (m-BCR) in intron 1 or major breakpoint cluster region (M-BCR) in intron 13 or 14. There is also a less common breakpoint cluster locates in intron 19, namely μ-BCR. Four *BCR-ABL1* isoforms can be derived from the fusion gene depending on the location of the breakpoints (De Klein et al., 1986; Clark et al., 1987; Wada et al., 1995; Soverini et al., 2018). Indeed, no matter where the exact breakpoint is located at *ABL1*

gene, the *BCR* gene eventually fuses with exon a2 of *ABL1* after the splicing of the primary transcript (Melo, 1996). The main factor that determines the final chimeric protein structure is the location of the breakpoints in *BCR* gene. When the breakpoint occurs at M-BCR, *ABL1* exon a2 fuses with either *BCR* exon e13 or e14 depending on the alternative splicing. Both e13a2 and e14a2 transcripts arise a 210 kDa chimeric protein ($p210^{BCR/ABL1}$) which can be detected in almost all the CML patients and approximately a third of the Ph-positive b-cell acute lymphoblastic leukemia (B-ALL) patients (Clark et al., 1987; Soverini et al., 2018). For other Ph-positive B-ALL, the *BCR* breakpoint locates in m-BCR which results in fusion between *ABL1* exon a2 and *BCR* exon e1. The e1a2 mRNA yields a 190 kDa BCR-ABL1 protein ($p190^{BCR/ABL1}$) (van Rhee et al., 1996). The rare μ-BCR breakpoint results in fusion of *ABL1* exon a2 and *BCR* exon e19. The formation of e19a2 transcript translates into a 230 kDa protein ($p230^{BCR/ABL1}$) which has been reported to be associated with CNL (Wada et al., 1995; Pane et al., 1996).

As described above, *ABL1* is implicated in various cellular processes. Thus, it should be strictly regulated. Nevertheless, the fusion with BCR results in deregulation of ABL1 protein which leads to deregulation of the related cellular processes. Among them, three major mechanisms are associated with the malignant transformation – (i) alteration of adhesion to stroma cells and extracellular matrix, (ii) mitogenic signaling activity as well as (iii) apoptosis inhibition (Soverini et al., 2018).

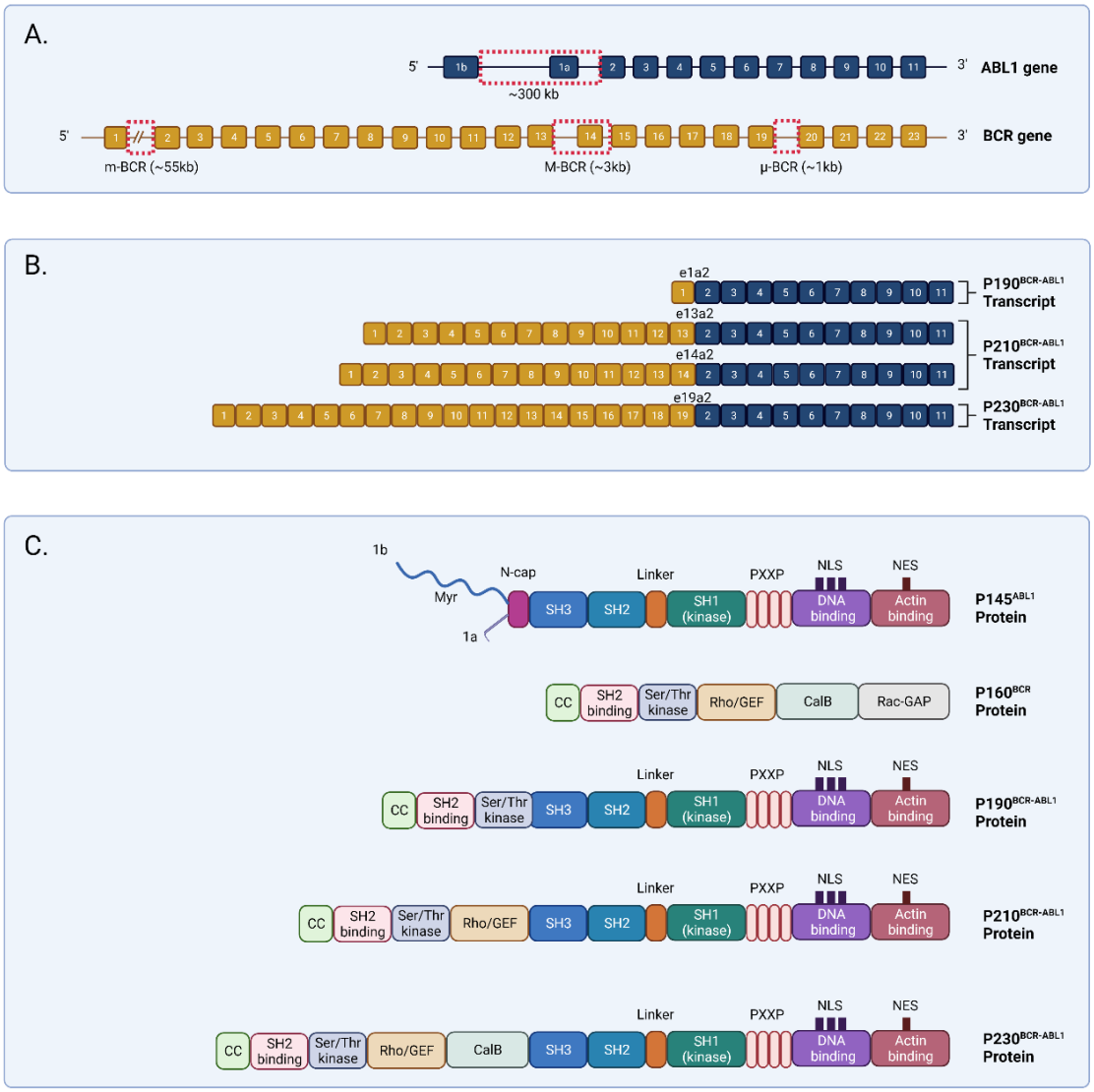


Figure 1.2 Breakpoints in *BCR* and *ABL1* genes and resulting transcripts as well as structures of native proteins and chimeric proteins.

(A) The translocation breakpoints of *ABL1* gene mostly fall in a large region ranging from upstream of the first exon 1b to downstream of second exon 1a. The translocation of *BCR* breakpoints mostly locate in either the minor breakpoint cluster region (m-BCR) in intron 1 or major breakpoint cluster region (M-BCR) in intron 13 or 14. There is also a less common breakpoint cluster (μ-BCR) locates in intron 19. (B) The most common fusion transcripts are e13a2 and e14a2 from translocation at M-BCR, which are both translated into the $p210^{BCR\text{-}ABL1}$ that has been found in typical CML and some Ph^+ ALL cases; e1a2 is produced from translocation at m-BCR, which is translated into $p190^{BCR\text{-}ABL1}$ that has been found in majority of Ph^+ ALL; e19a2 is produced from translocation at μ-BCR, which is translated into $p230^{BCR\text{-}ABL1}$ which has been found in

CNL. (C) A schematic diagram showing different domains of the ABL1, BCR and BCR-ABL1 proteins. Figure adapted from Soverini et al. (2018).

1.3.3 BCR-ABL1 oncogenic pathways

Normal ABL1 protein can translocate between nucleus and cytoplasm depending on the physiological needs (Cilloni and Saglio, 2012). Yet, BCR-ABL1 protein losses this property and mainly stays in cytoplasm. This allows the fusion protein interacting with proteins in cytoplasm taking part in oncogenic pathways (Pendergast et al., 1993). Apart from the change in localization, fusion with BCR also results in deregulation of tyrosine kinase activity of ABL1 due to loss of the N-cap of ABL1 and abrogation of SH3 inhibitory regulation on SH1 tyrosine kinase domain (Pendergast et al., 1991). Loss of N-cap results in exposure of SH1 domain to exhibit its tyrosine kinase activity. Together with dimer or tetramer formation facilitated by the oligomerization domain in BCR, the tyrosine kinase activity of ABL1 increased. In turn, various proteins implicated in different pathways can be deregulated, including adapter molecules, and proteins related to cytoskeleton and cell membrane organization (Deininger et al., 2000). BCR-ABL1 proteins are capable of autophosphorylation. This results in increase in amount of phosphotyrosine residues which can bind with SH2 domain-containing proteins (Pendergast et al., 1991). Those proteins are mainly cytoplasmic since BCR-ABL1 mainly retains in cytoplasm. Three major pathways are found altered in BCR-ABL1-transformed cells, including RAS/MAPK pathway, PI3K/AKT pathway, and JAK2/STAT pathway (Deininger et al., 2000).

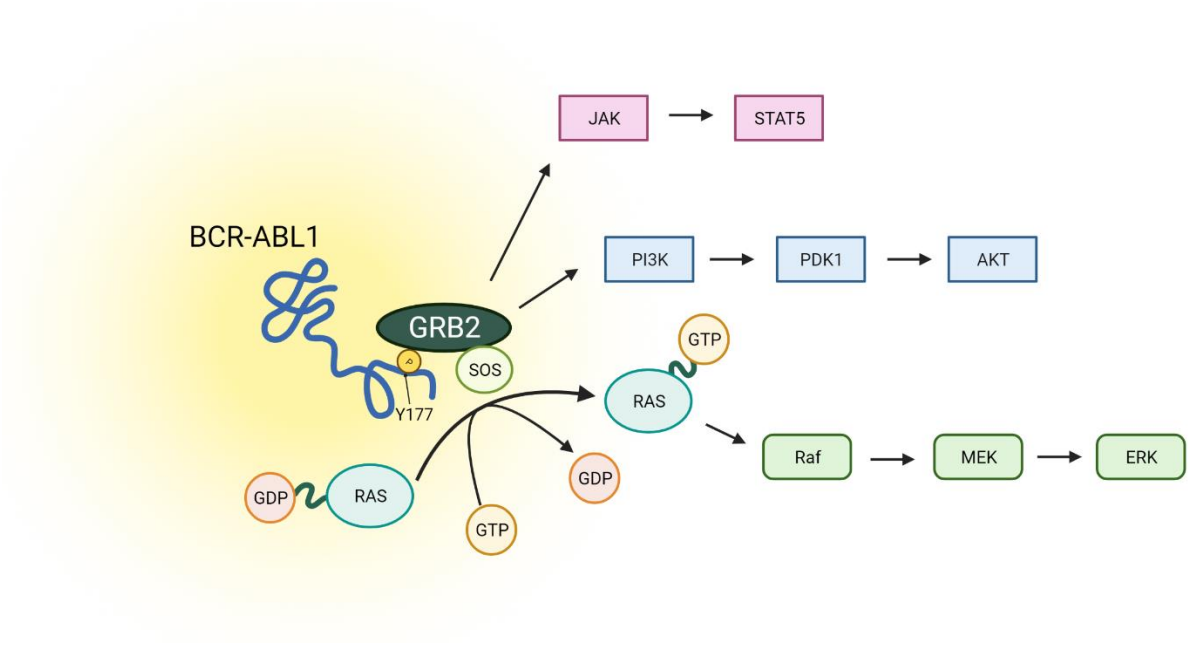


Figure 1.3 The signaling pathways activated by BCR-ABL1.

Tyr177 phosphorylation provides docking site for GRB2, forming the BCR-ABL1/GRB2 complex which recruits SOS and GRB2. The BCR-ABL1/GRB2/SOS complex converts Ras from GDP-bound inactivated form to GTP-bound active form. Ultimately, the Ras-Raf-MEK-ERK signaling cascade activation leads to gene transcriptions and, thus, abnormal cellular proliferation. Alternately, Ras/MAPK pathway can also be activated by BCR-ABL1/GRB2/GAB2 complex. The GRB2/GAB2/SOS complex can also activate PI3K and thus the downstream serine-threonine kinase AKT. Increased amount of phosphorylated tyrosine residues in BCR-ABL1 interact with SH2 domain in STAT and result in continuous activation of the STAT proteins in a cytokine/JAK-independent manner. Figure adapted from Li et al. (2017).

1.3.4 RAS/MAPK pathway

Phosphorylation of tyrosine 177 (Y177) in BCR is critical for BCR-ABL1-mediated leukemogenesis (Zhang et al., 2001; Li et al., 2017). The phosphorylated Y177 in SH2 domain provides a docking site for growth factor receptor bound protein 2 (GRB2) (Pendergast et al., 1991). The BCR-ABL1/GRB2 complex recruits both Son of Sevenless (SOS) and GRB2-associated binding protein 2 (GAB2) (Cortez et al., 1997). The BCR-ABL1/GRB2/SOS complex converts Ras from the inactive GDP-bound form to the active GTP-bound form (Hellmann, 1992). This activates the Ras/mitogen-activated protein kinase (MAPK) pathway. Activated Ras recruits Raf which initiates the signaling cascade through MAPK/ERK kinase (MEK) 1/2 and extracellular signal-regulated kinase (ERK). Alternately, Ras/MAPK pathway can also be activated by BCR-ABL1/GRB2/GAB2 complex when the complex binds with SH2 containing protein tyrosine phosphatase-2 (SHP2) through phosphorylated tyrosine residues on GAB2 (Sattler et al., 2002). Ultimately, the activation of Ras-Raf-MEK-ERK signaling cascade leads to gene transcriptions and, thus, abnormal cellular proliferation (Cilloni and Saglio, 2012).

1.3.5 PI3K/AKT pathway

Apart from RAS/MAPK pathway, the GRB2/GAB2/SOS complex can also activate phosphoinositide 3-kinase (PI3K), and thus, the downstream serine-threonine kinase AKT. AKT regulates various cellular processes including cellular survival, autophagy and cell proliferation (Skorski et al., 1995). It inhibits apoptosis by inactivation of pro-apoptotic proteins BAD and FOXO transcription factor, which results in reduction of transcription of cellular inhibitor p27 and another pro-apoptosis protein Bcl2 (Cilloni and Saglio, 2012). In addition, AKT upregulates mTOR which also involves in numbers

of cellular processes such as autophagy cell growth and proliferation (Cilloni and Saglio, 2012). Thus, through mTOR upregulation, FOXO and BAD inhibition, PI3K/AKT increases cellular survival and facilitates clonal expansion in CML patients.

1.3.6 JAK/STAT pathway

Signal transducer and activator of transcription (STAT) proteins are intracellular transcription factors that implicate in many cellular processes including cell apoptosis, proliferation and differentiation (Cilloni and Saglio, 2012). In normal cells, STAT proteins have to be activated by receptor-associated Janus kinase (JAK) exclusively after cytokine binding to receptors. However, in CML cells, autophosphorylation property of BCR-ABL1 results increase content of phosphorylated tyrosine residues which associate with SH2 domain in STAT1 and STAT5. This results in continuous activation of the STAT proteins in a cytokine/JAK-independent manner (Ilaria and Van Etten, 1996). Thus, STAT continuously translocate to nucleus, leading to transcription of target genes. Ultimately, CML cells exhibit an increase in proliferation and a reduction in apoptosis.

1.4 TKI treatment for CML

A decade ago, the treatment of CML was limited to nonspecific biological agents such as hydroxyurea, busulfan, and interferon-alpha (IFN-α)(Silver et al., 1999). The INF-α treatment reduced disease progression and improved patient survival, but it had a modest efficacy and a variety of side effects. Combined treatment with IFN-α and cytarabine may be more effective and was once viewed as the gold standard therapy, however, none of these methods could result in molecular remission. Currently, allogeneic stem cell transplantation (allo-SCT) is the only curative therapy option, but

it is associated with morbidity and mortality risks. Thus, allo-SCT is mainly recommended for young patients with good organ functions and performance status, and in the presence of an appropriate donor (Jabbour and Kantarjian, 2020).

With the improved understanding of the molecular basis of CML, the therapeutic landscape for CML radically changed due to the development of small molecule tyrosine kinase inhibitors (TKIs), which potently block BCR-ABL1 tyrosine kinase activity by impairing the interaction between the BCR-ABL1 oncoprotein and adenosine triphosphate (ATP), thereby effectively reduce the cellular proliferation of malignant clones (An et al., 2010). In general, TKIs could reduce CML colonies without interfering the apoptosis pathway and growth of normal colonies. As a result of such targeted therapy approach, CML has become a manageable disease with the 10-year survival rate remarkably improved from about 20% to 80-90% (Huang et al., 2012; Hochhaus et al., 2017).

1.4.1 Imatinib

Imatinib (as known as STI571 or Gleevec) is the first generation of TKI being discovered in 1996 and is the first TKI approved for the treatment of patients with CML-CP by the U.S. Food and Drug Administration (FDA) in 2001 (Buchdunger et al., 1996; Cohen et al., 2002). Imatinib inhibits BCR-ABL1 kinase activity by competitive inhibition of the ATP-binding site called SH1 domain, which prevents the subsequent phosphorylation of proteins that are involved in transduction of cell signaling. However, the specificity of imatinib is moderate, it also demonstrated inhibitory activity against the C-KIT (also referred to as stem cell factor receptor or CD117) tyrosine kinase and platelet-derived growth factor receptor (PDGFR) (Druker and Lydon, 2000).

A landmark clinical trial study for imatinib is the International Randomized Study of Interferon and STI571 (IRIS) (O'Brien et al., 2003). The study was conducted with randomly assigned 1,106 patients in chronic phasic CML to receive 400 mg imatinib daily, or IFN-α combined with cytarabine. With a median follow-up of 19 months, imatinib significantly improved patient outcomes compared to IFN-α plus cytarabine. Notably, the complete cytogenetic response rates for patients receiving imatinib and IFN-α plus cytarabine were 74% and 9%, respectively. Afterwards, another six-year follow up study of imatinib treatment showed 83% event-free survival rate and 88% overall survival rate, which underscored the safety and efficacy of imatinib as the highly effective first-line therapy for the CML patients (Hochhaus et al., 2009). Nonetheless, approximately 30% of patients either developed drug resistance or responded poorly to imatinib treatment with the standard dose 400 mg per day (Rosti et al., 2017). Therefore, different strategies were applied such as increasing the imatinib dose to 600-800 mg per day, combined therapy with IFN-α or pegylated interferon, which significantly improved the outcome compared with imatinib treatment alone. Still, some patients were intolerant to the toxicity and adverse effects of imatinib, hence second and third generation TKIs, such as dasatinib, nilotinib, bosutinib and ponatinib, have been developed for effective treatment of CML patients with imatinib resistance (Jabbour and Kantarjian, 2020).

1.4.2 Development of imatinib resistance

Imatinib resistance could be developed over time course of treatment due to the occurrence of mutations in the BCR-ABL1 oncoprotein, which is the most common cause of acquired resistance to targeted therapies. Since stringent binding is required for the BCR-ABL1 blockage by imatinib, mutations at the binding domain could result in substitution of amino acids, and hence the conformation change of domain would reduce the binding affinity for imatinib (Talati and Pinilla-Ibarz, 2018). Indeed, most mutations that confer resistance to imatinib could be overcome by second generation TKIs, but novel mutations have emerged to render drug resistance and a prominent point mutation in *BCR-ABL1* is the T315I mutation that allows resistance to almost all available TKI except the third generation TKI ponatinib (Wolfe and Rein, 2021). This mutation is caused by the substitution of thymine for cytosine at position 944 of the ABL gene and leads to the production of threonine instead of isoleucine. Such substitution eliminates an oxygen molecule that is vital for the hydrogen bonding between imatinib and ABL kinase, meanwhile, steric hindrance is also created to prevent the binding of imatinib (Bixby and Talpaz, 2009). Resistance can also occur due to gene amplification of *BCR-ABL1* and eventually leading to the overexpression of fusion protein, which is observed in 20% of patients who have imatinib therapy relapse (Hochhaus et al., 2002).

Besides point mutations in BCR-ABL1 as a frequent cause of TKI resistance, there are other accounts of imatinib resistance that are independent of BCR-ABL1. For example, it is known that imatinib is transported out of cells by MDR1 efflux pump protein and into cells by an active uptake mechanism via hOCT1 influx transporter, hence the differential function or expression of these transporters would result in a reduced

intracellular concentration of imatinib and consequently such pharmacodynamic properties contribute to the resistance (Thomas et al., 2004). Additional to abnormal expression of regulatory pumps controlling efflux and influx of TKIs, activation of alternative survival pathways to maintain viability and growth may also be responsible for primary or secondary resistance, despite BCR-ABL1 kinase activity is suppressed. For instance, BCR-ABL1-independent STAT3 activation has been shown to result in TKI resistance via both intrinsic pathway and extrinsic mechanism induced by bone marrow-derived factors, leading to increased levels of its target genes *Bcl-xl*, *Mcl-1*, and survivin (Bewry et al., 2008; Eiring et al., 2015). On the other hand, phosphatidylinositol-3 kinase (PI3K) is known to be one of downstream signaling molecules upon activated signal transduction by BCR-ABL1 and involve in proliferation of CML cells.

Later research has further demonstrated the important compensatory role of PI3K/AKT/mTOR-pathway in mediating CML cell survival during the early stage development of imatinib resistance prior to the emergence of strong resistance due to mutations in BCR-ABL1 (Burchert et al., 2005). Forkhead box O1 (FOXO1), a downstream transcription factor in the PI3K signaling pathway, has also been shown to be elevated in TKI-resistant cell lines as well as relapsed CML patients lacking BCR-ABL1 kinase domain mutations (Wagle et al., 2016). A study utilizing large-scale RNAi screening to identify genes associated with imatinib resistance independent of *BCR-ABL1* mutations has found an increased activation of RAF/MEK/ERK signaling pathway due to the increased expression of a protein kinase C (PKC) family member called PKCη, leading to inhibition of CML cell apoptosis and enhanced proliferation (Ma et al., 2014). In the study, it has also revealed an upregulation of Protein kinase C

eta (PRKCH) gene which encodes PKCη in CML patients with BCR-ABL1-independent imatinib resistance, and CML stem cells containing high expression of PRKCH could contribute to their intrinsic resistance. Then, further testing with the combination of imatinib and trametinib, which is a MAPK/ERK kinase (MEK) inhibitor, showed synergistic effect against imatinib-resistant CML cells and had minimal effects on normal hematopoietic stem cells (Ma et al., 2014).

In addition, mutations in epigenetic regulators, such as ASXL1, DNMT3A, IDH1, and SETBP1, could also contribute to BCR-ABL1-independent TKI resistance (Kim et al., 2017). In fact, it has been shown that an increased risk of poor treatment outcome is associated with the presence of these mutations at diagnosis. Also, the presence of mutations in epigenetic regulators are associated with the progression of CML to blast crisis phase (Giotopoulos et al., 2015; Branford et al., 2018). The mechanism by which these mutations cause TKI resistance in CML is not entirely clear at this moment, but there has been increasing evidence showing that epigenetic dysregulation is responsible for TKI resistance and leukemic clone escape (Loscocco et al., 2018).

Other epigenetic regulators that are frequently deregulated in cancers and leukemia are the polycomb repressive complex 1 (PRC1) and polycomb repressive complex 2 (PRC2), which predominantly regulate gene expression by trimethylation of lysine 27 on histone H3 (H3K27me3) (Crea et al., 2012; Scott et al., 2016). PRC2 dysregulation has been revealed in CML cell lines, primary CML cells, or murine models, and an aberrant expression level of PRC2 is also associated with TKI response and progression of CML to blastic phase (Pizzatti et al., 2010; Bozkurt et al., 2013).

Another epigenetic regulator involved in CML resistance is microRNA (miRNA), a family of small non-coding RNAs that could regulate gene expression through direct binding to their target mRNAs. A recently study utilized a miRNA microarray to compare the expression of miRNAs between K562 cell line and healthy controls and demonstrated a large amount of downregulated miRNAs in K562, which indicated the potential involvement of miRNA in CML (Rokah et al., 2012). Furthermore, the downregulation extent of several miRNAs, such as miR-29 cluster, miR-23a, and miR-451, in drug-naïve chronic phase CML patients can be evaluated to distinguish patients who are responding to imatinib or not (Fallah et al., 2015; Loscocco et al., 2019). Particularly, an inverse relationship between miR-451 and BCR-ABL1 expression has been observed in a subset of patients responding to imatinib, indicating miR-451 may potentially target BCR/ABL1 (Lopotová et al., 2011; Scholl et al., 2012).

1.4.3 New therapy approaches to deepen remission

To better achieve depth and durability of response, prevent the development of TKI resistance, limit the toxicity of TKIs, and promote treatment-free remissions, new treatment approaches are needed. Previously, frontline treatment was aimed at achieving a cytogenetic remission coupled with a molecular remission to improve overall survival of patients (Hanfstein et al., 2012). Recently, treatment-free remission (TFR) has been accepted as a goal of frontline treatment, as does the goal of achieving a deep molecular response that enables discontinuation of TKI treatment and improves the quality of life of patients.

Currently, the estimated discontinuation rate for imatinib and second-generation TKIs are 30–40% and 40–50% of patients, respectively, but there is no approach to predict who will not achieve a deep remission. Since TFR in CML therapy is a relatively new concept, no approved treatments are currently available to increase chances of reaching deep molecular responses, nor treatment adjuncts that may be added if patients are not achieving the desired result. However, the current ultimate goal is to enhance the number of patients eligible for treatment discontinuation, and even achieve a cure, hence several new approaches are being investigated to reach these goals.

For example, A breakthrough therapy designation has been granted to asciminib in February 2021 by the U.S. FDA for treating patients with CML who have previously been treated with two TKIs or more. Also, asciminib has a breakthrough status for treating CML patients possessing the T315I mutation. Instead of competing with ATP-binding sites in active conformation of BCR-ABL1, asciminib specifically binds to the myristoylation pocket and inhibit ABL1 activity by trapping it in the inactive

conformation (Wylie et al., 2017). Currently, asciminib is being evaluated in patients who have failed to respond to two or more TKIs (Rea et al., 2021). Nonetheless, due to the effectiveness and safety of asciminib, it may be potentially used even as first-line therapy to boost the speed and depth of the drug response. A combination of an TKI targeting ATP-site with asciminib may also prevent the development of resistance in the presence of point mutations in the TKI binding sites. Currently, a non-randomized Phase 2 study in Germany is confirming the drug effects by administering asciminib in combination with imatinib, nilotinib, or dasatinib in various doses (Javidi-Sharifi and Hobbs, 2021).

In addition, a combination of TKI and other drugs targeting signaling pathways or molecules other than BCR-ABL1-dependent pathway is another trend that under investigation. For instance, a combination of ruxolitinib, a JAK1/JAK2 inhibitor, and nilotinib was under phase I clinical trial to evaluate the safety and tolerability in CML patients since an activation of JAK/STAT pathway can enhance the survival of CML cells during TKI treatment (Sweet et al., 2018). Pre-clinical studies have also shown that blockage of JAK/STAT signaling could restore the sensitivity of CML cells to BCR-ABL inhibition (Nair et al., 2012). Combination of TKI with other drugs such as interferon, immune check point inhibitors, or pioglitazone are also being investigated in different phases of clinical trials (e.g. NCT03831776, NCT03516279, NCT02767063), and all these studies aim to achieve a deeper molecular response by targeting both BCR-ABL1-dependent and BCR-ABL1-independent signaling at the same time (Javidi-Sharifi and Hobbs, 2021).

Overall, the issue of drug resistance not only hinders the achievement of disease remission but poses a more far-reaching consequence where CML progresses from a mild chronic phase to the more aggressive lethal accelerated phase and blastic phase CML or even acute leukemia (Bhat et al., 2020). To tackle the drug resistance problem, it is crucial to understand the mechanisms and pathways of resistance development and disease progression, especially those are BCR-ABL-independent, hence more effective treatment with combination of TKI and specific drug can be applied. Despite much effort has been put in studying the underlying mechanisms in the past decades, most of the studies have been focusing on the messenger RNA expression and their encoded proteins (Bhat et al., 2020). Various mutations have been found to be associated with disease progression and TKI resistance mechanisms, but the resulting abnormalities seem to be quite diverse. Since only a small portion of human transcriptome is covered by protein-coding genes, more research has been moved on to studying another significant portion of the transcriptome, i.e., the non-coding RNA (ncRNA), including long non-coding RNA (lncRNA) and microRNA (miRNA) (Soltani et al., 2017; Bhat et al., 2020). With greater coverage of transcriptome studies, there can be a deeper insight in pathophysiology of CML progression and drug resistance, which is essential for discovery of new diagnostic markers and therapeutic targets.

1.5 MicroRNA

MicroRNAs (miRNAs) are relatively well-studied non-coding RNAs that are expressed in a wide range of eukaryotic cells from mammals to plants and even viruses (Ambros, 2001; He and Hannon, 2004). In general, miRNAs are approximately 22 nucleotides in length and their primary function is to repress the expression of cellular mRNAs although it has recently been shown that miRNAs are also able to stimulate mRNA expression through binding to AU-rich region at the 3'end of target transcript (Vasudevan et al., 2007; Vasudevan, 2012) or epigenetically activate gene transcription by serving as an enhancer trigger (Xiao et al., 2017a). Extensive studies have shown that miRNAs play crucial regulatory roles in the development of tissues and cells as well as progression of various diseases (Vishnoi and Rani, 2017). The biogenesis of mature miRNAs is initiated once the miRNA genes are transcribed by RNA polymerase II which yields a long primary transcript that possesses a hairpin-like structure, also referred to as pri-miRNA (Michlewski and Caceres, 2019). Then, the premature-miRNA undergoes nuclear and cytoplasmic processing that are executed by DROSHA and the endoribonuclease DICER respectively, resulting in a miRNA duplex. One of the miRNA strands then associates with RNA-induced silencing complex (RISC) along with an Argonaute (AGO) protein, resulting in the mature functional miRNA (Michlewski and Caceres, 2019).

Generally, miRNAs regulate gene expression post-transcriptionally through the complementary binding to the target mRNAs. Such binding guides the effector proteins of the RISC towards target mRNA and causes its degradation (Thomson et al., 2011). Absolutely complementary binding between miRNA and mRNA is very infrequent, but as little as 6 bp matching can also be sufficient for their interaction and lead to gene

suppression. Such interaction usually occurs between the 5'-end of the miRNA (nucleotides 2-7 from 5' and denoted as 'seed' region) and the 3' untranslated region (3'-UTR) of the mRNA (Doench and Sharp, 2004; Brennecke et al., 2005). Consequently, a partial hybrid of the mRNA and the miRNA 'seed' region or other sections of the miRNA is formed. This results in the inhibition of gene expression by a combination of RNA degradation and translational repression. RNA degradation is promoted by either endonucleolytic cleavage catalyzed by AGO2, or deadenylation and exonucleolytic attack (Bartel, 2004). Since considerable research reported alteration of miRNA expression during development and progression of human diseases, miRNA could serve as a promising molecular biomarker for diagnosis or prognosis (Ullah et al., 2014).

1.6 Long non-coding RNA

Long non-coding RNA (lncRNA) is a newly evolving class of non-coding RNAs that are conventionally defined as longer than 200 nucleotides in length without protein-coding capacity due to lack of significant open reading frames, hence they would not be translated into proteins but act as a functional transcript instead. In the past, lncRNA were once considered as an existence of transcriptional noise with low evolutionary conservation (Rinn and Chang, 2012). However, increasing amount of studies have provided strong evidence that lncRNAs are a crucial contributor to diverse physiological and pathological conditions through their critical regulatory roles in different layers (Batista and Chang, 2013; Yang et al., 2014). LncRNAs share some similarities with mRNAs. Most lncRNA transcripts are transcribed by RNA polymerase II and they undergo processing including splicing, 5' end capping and polyadenylation. However, lncRNA expression level is in generally lower, and is much more specific to particular cell or tissue types (Batista and Chang, 2013; Alvarez-Dominguez et al., 2014; Fatica and Bozzoni, 2014b).

Research work in the last decade has demonstrated the diverse roles of lncRNAs in regulation of different biological processes by mediating gene expression at epigenetic, transcriptional or posttranscriptional levels via various mechanisms, such as recruitment and assembly of chromatin modifiers, organization of chromosome structure, mRNA splicing, translational regulation, etc. (Batista and Chang, 2013; Garitano-Trojaola et al., 2013; Schaukowitch and Kim, 2014). The multifaceted regulatory functions of lncRNAs are made possible due to their high versatility to interact with a variety of different molecules including DNAs, other RNAs or protein complexes. Notably, lncRNAs specifically express and function across different cell or

tissue types, meaning a particular lncRNA may be expressed and function in one cellular phenotype but not work in the same way in others (Cabili et al., 2011; Cabili et al., 2015). All these special features of lncRNAs created additional levels to the complexity of precise gene expression control.

According to NONCODEv5, approximately 96,000 human lncRNAs have been identified to date (Fang et al., 2018). Among them, only about 18,000 lncRNA genes (19%) have been annotated by the GENCODE Consortium (Version 33) and 32,000 records of human disease association with experimental support have been obtained (Harrow et al., 2012; Fang et al., 2018). It clearly reveals that only a small portion of lncRNAs have been functionally characterized and the vast majority remains unknown, thus a lot of effort is still required to expand our understanding of lncRNAs.

1.6.1 Archetypes of lncRNA functional mechanism

On account of the widespread attention to lncRNAs and the advancement in technology, researchers have gained a better understanding of lncRNAs regarding their mechanisms of action. Initially, it was proposed that the primary function of lncRNAs is gene expression regulation and four dominant ways of lncRNAs interacting with their targets were suggested, including signals, decoys, guides and scaffolds as shown in Figure 1.4 (Wang and Chang, 2011). First, lncRNAs could function by serving as signaling molecules in response to diverse stimuli in specific cell or tissue types (Mathy and Chen, 2017). It therefore activates or negatively regulates the downstream transcription of other genes at very specific time and space according to different stimuli. For example, lncRNA PANDA can be activated by the interaction between p53 and CDKN1A as a result of DNA damage response. Activated PANDA prolongs the survival of tumor

cells by preventing the expression of apoptotic genes via interaction with the nuclear transcription factor Y subunit α (NF-YA) (Hung et al., 2011).

LncRNAs could also function as molecular decoys and negatively regulate gene expression by binding to target molecules such as transcription factors or chromatin modifying enzymes and impede them from binding to promoter of other genes, resulting in a suppression of downstream transcription and hence gene expression. Functionally opposite to decoys, lncRNAs could act as guides so that they bind to target effector proteins such as ribonucleoprotein complex, and then direct their localization to specific target site. Then, the ribonucleoprotein complex could result in activation or repression of gene transcription either in *cis* or in *trans* via chromatin modification. For instance, lncRNA Fendrr interacts with PRC2 and TrxG/MLL complexes and modulates histone modification H3K27me3 and H3K4me3 at specific promoter sites, resulting in regulation of genes controlling pluripotency and cell differentiation in mouse embryonic cells (Grote et al., 2013).

Finally, lncRNAs could act as scaffolds to provide a platform upon which multiple effector proteins or subunits can be assembled simultaneously in order to function together. These complexes could then cause activation or repression of transcription via epigenetic modification of chromatin. One of the most studied scaffold lncRNAs is Xist, which is expressed by one of the two X chromosomes that is supposed to be inactivated in females (Bousard et al., 2019; Colognori et al., 2019). Xist functions by recruiting the PRC1 and PRC2 complexes, which leads to the formation of heterochromatin state via histone methylation and then suppresses the gene expression of this X chromosome, leading to a dose compensation role in mammals. The above-mentioned regulatory

mechanisms are the predominant ways of action first discovered in lncRNAs that are mainly related to gene transcriptional control but increasing research has elucidated the mode of action of lncRNAs could be very diverse, which will be described in Chapter 1.6.2 & 1.6.3.

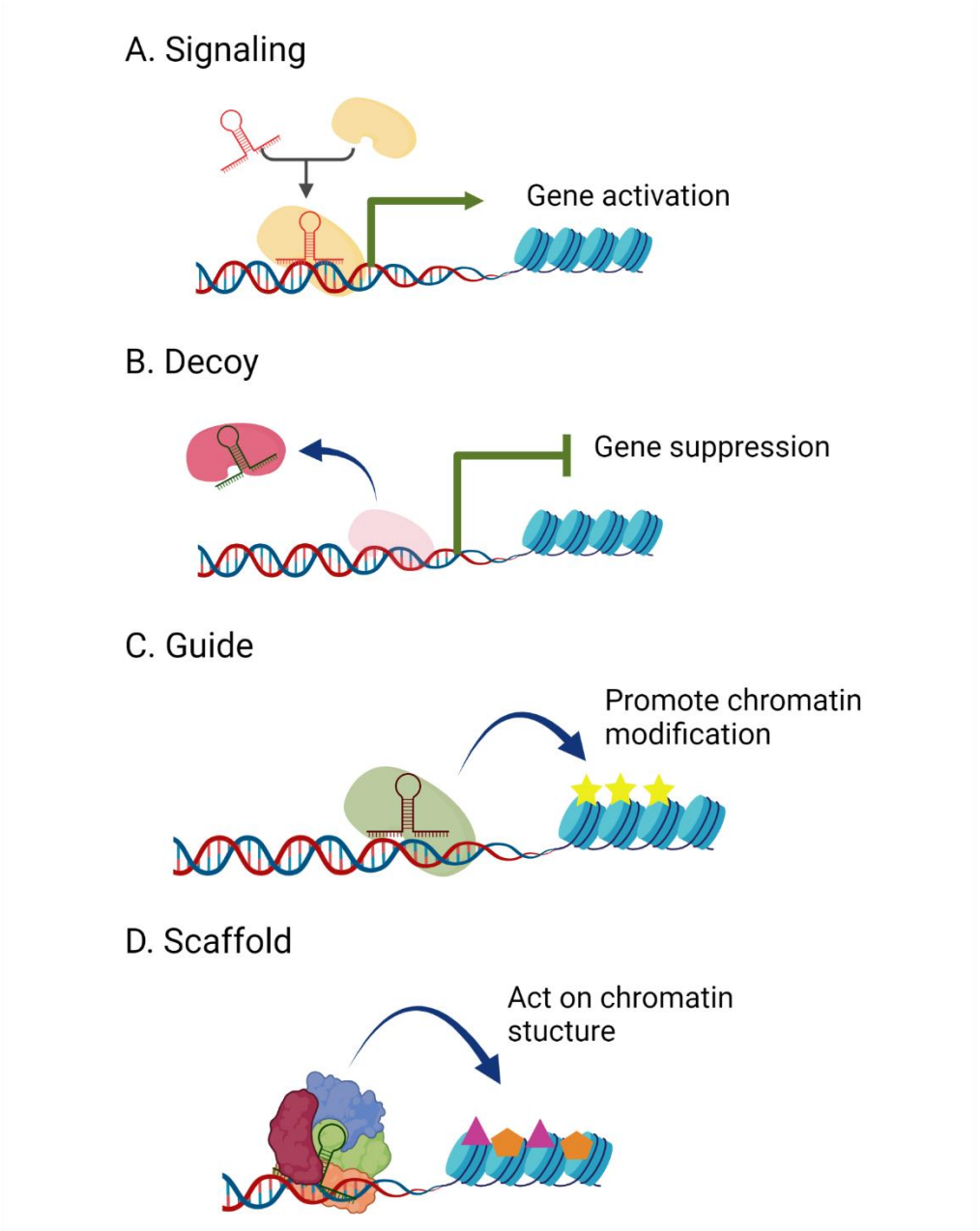


Figure 1.4 The four archetypes of lncRNA mechanism.

(A) As a signal molecule, lncRNA mediates the transcription of downstream genes in response to specific stimuli. (B) As a decoy, lncRNA binds to functional proteins to block their regulation on other genes. (C) As a guide molecule, lncRNA carries functional protein molecules and locate them to their target sites to perform gene regulations. (D) As a scaffold molecule, lncRNA can be a platform for different types of molecular complexes to assemble on it and work in cooperation. Figure adapted from Wang & Chang (2011).

1.6.2 Functions of lncRNAs in the nucleus

Distinct functions of lncRNAs depend on the cellular compartments where they are localized. A significant amount of lncRNA is preferentially localized in the nucleus, which regulate the expression of other genes mainly at epigenetic and transcriptional levels. Moreover, some nuclear lncRNAs play a vital role in the organization of nuclear structures (Engreitz et al., 2016; Sun et al., 2018). For epigenetic regulation of chromatin organization also known as the chromatin remodeling, lncRNAs usually involves the modification of histone proteins which in turn affect chromosome structure by acting as molecular scaffold of particular chromatin-regulatory complexes or histone-modifying machinery for their integration (Engreitz et al., 2016). Chromatin remodeling could impact DNA accessibility for transcription factors to the target promoters and hence affect the gene transcription process.

In fact, epigenetic modifiers are the most frequent protein partners interacting with lncRNAs that have been identified so far, including Polycomb Repressive Complex (PRC), MLL/TrxG complex, DNA-methyltransferase 1 (DNMT1), histone demethylase LSD1, DNA demethylation regulator GADD45A, histone methyltransferase G9a, and so on; among which the members of the Polycomb Repressive Complex 2 (PRC2) and histone methyltransferases have gained a lot of awareness (Han and Chang, 2015; Hanly et al., 2018). As an example, HOTAIR is a well-studied lncRNA and is transcribed from *HOXC* locus located on chromosome 12. It exerts regulatory function by directly binding to EZH2, which is a component of the PRC2, and then the assembled PRC2 could silence a cluster of tumor suppressor genes on the *HOXD* locus epigenetically by regulation of histone methylation (Rinn et al., 2007; Tsai et al., 2010). Recently, a study has reported that HOTAIR could promote

kidney cancer progression by altering the chromatin structure and enhance the transcription of the *SNAIL* gene through interaction with ARID1 and SMARCB1, which are subunits of the chromatin remodeling complex SWI/SNF (Imai-Sumida et al., 2020).

Some lncRNAs modulate chromatin organization without direct interaction with chromatin modulators or the chromatin, but through the interplay with other protein molecules. For instance, lncRNA BCAR4 indirectly influences p300-dependent histone-acetylation and activates the Hedgehog/*GLI2* transcriptional program to promote cell migration in breast cancer cells (Xing et al., 2014). By interacting with the RNA-binding protein (RBP) called SNIP1, BCAR4 releases the suppression of SNIP1 on p300 to promote the H3K18 acetylation of *GLI2* promoter. Meanwhile, BCAR4 recruits another RBP called PNUTS to the acetylated H3K18 and leads to the activation of *GLI2*. Overall, BCAR4 influences the epigenetic regulation of *GLI2* transcription by interacting with different RBPs (Xing et al., 2014).

For transcriptional regulation, lncRNAs exert their functions mainly by serving as a decoy or guide molecule. Transcription starts with the binding of RNA polymerase II to the gene promoter region together with transcription factors. As a decoy, lncRNAs could repress gene expression by direct binding to RNA polymerase II or transcription factor complexes, and result in the intervention of their binding to the promoter region. On the contrary, lncRNAs could also guide RNA polymerase II to the promoter of specific genes for binding and thus facilitate gene transcription (Faust et al., 2012). For instance, lncRNA LEENE guided and facilitated the recruitment of RNA polymerase II to the promoter region of endothelial nitric oxide synthase (*eNOS*) and enhance the

transcription of eNOS mRNA (Miao et al., 2018). Indeed, some lncRNAs could even serve as transcription factors themselves. The lncRNA GAS5 folds into a DNA-like structure by forming hairpin so that it could bind to the DNA-binding domain of the glucocorticoid receptor, and inhibits the transcription of glucocorticoid-responsive genes, influencing the metabolic activities and cell survival during starvation (Kino et al., 2010).

Besides, lncRNAs are also involved in the regulation of alternative splicing of pre-mRNA in the nucleus, which is an important process to increase transcriptome and proteomic complexity in higher eukaryotes. The serine/arginine (SR) splicing factors are crucial regulators to modulate cell- or tissue-type-specific alternative splicing dependent of their concentration and phosphorylation state (Long and Caceres, 2009). LncRNAs participate in splicing via two main ways, that are interaction with splicing factors or forming RNA-RNA duplexes with pre-mRNAs (Liu et al., 2021). For example, MALAT1 is a highly nucleus-restricted lncRNA localized at the nuclear speckles and it functions as a molecular sponge to titrate and dilute the cellular pool of SR splicing factors, influencing the distribution of splicing factors in the nucleus. MALAT1 also regulates the phosphorylation of SR proteins to affect their activity and hence controls alternative splicing of pre-mRNA (Tripathi et al., 2010; Tripathi et al., 2013). Another lncRNA 51A could bind to intron one of a protein-coding gene Sortilin Related Receptor 1 (*SORL1*) and mask the canonical splicing sites, thereby driving a splicing shift of SORL1 pre-mRNA to an alternatively spliced protein isoforms, leading to a potential in Alzheimer's disease (Ciarlo et al., 2013).

Finally, lncRNAs are also known to maintain and nucleate specific nuclear domains, which are non-membranous structures enriched with a unique set of RNAs and proteins. Eukaryotic cells possess several nuclear domains, including the nucleolus, nuclear speckles and paraspeckles. The lncRNA NEAT1 is known to nucleate the formation of paraspeckles that contain certain RBPs. NEAT1 is essential for the structural integrity of paraspeckles as experiments have shown the knockdown or knockout of NEAT1 would result in the dispersion of paraspeckle proteins (Clemson et al., 2009; Sunwoo et al., 2009). Furthermore, the paraspeckle integrity requires ongoing transcription process of NEAT1, which suggests the formation of paraspeckles is coupled to the biogenesis of NEAT1 (Mao et al., 2011). The major functions of nuclear enriched lncRNAs are summarized in Figure 1.6.2, also showing some cytoplasmic functions that will be described in Chapter 1.6.3.

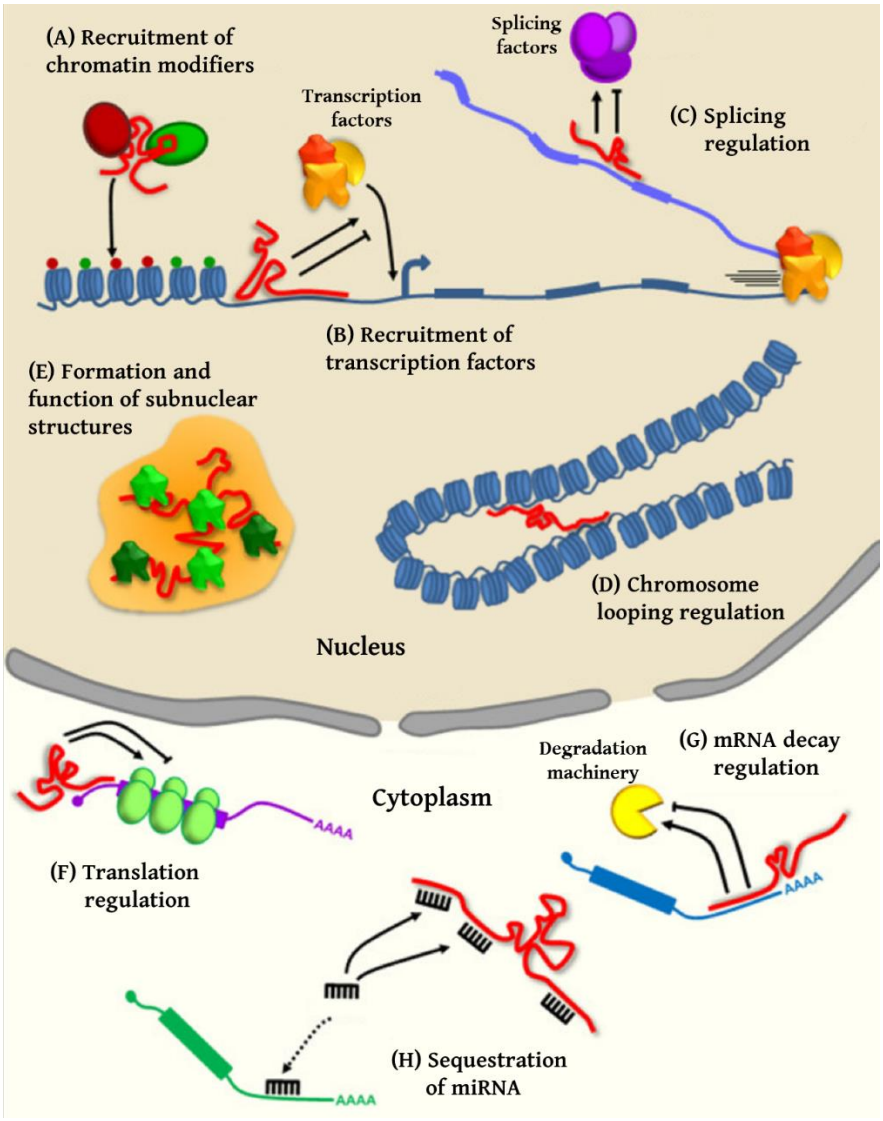


Figure 1.5 Major molecular functions of lncRNAs.

(A) LncRNAs (in red) recruit chromatin modifiers to mediate histone modification. (B) LncRNAs recruit transcription factors and core components to regulate transcription process. (C) LncRNAs modulate alternative splicing events. (D) LncRNAs regulate the organization of DNA through chromatin looping (E) LncRNAs serve as structural components for the formation of nuclear bodies. (F) LncRNAs regulate translation process of mRNAs. (G) LncRNAs modulate the degradation process of mRNAs. (H) LncRNAs act as molecular sponges to sequester miRNA, thus de-repressing the expression of the mRNAs targeted by the miRNA. Figure adapted from Neguembor, Jothi, & Gabellini, (2014).

1.6.3 Functions of lncRNAs in the cytoplasm

In the cytoplasm, lncRNAs regulate gene expression mainly at post-transcriptional, translational and post-translational levels, ranging from the regulation of mRNA stability by preventing or inducing their degradation, to the regulation of protein translation (Carrieri et al., 2012; Yoon et al., 2012; Fatica and Bozzoni, 2014a; Noh et al., 2018). Some lncRNAs could act as sponges for the competition of miRNAs, or even house miRNAs within their transcripts (Cai and Cullen, 2007; Zhang et al., 2012; Wang et al., 2013b).

A lncRNA example of regulating RNA stability is lncRNA LAST, which acts as a stabilizer to promote the stability of Cyclin D1 mRNA by interacting with cellular nucleic acid binding protein (CNBP) (Cao et al., 2017). Besides, certain lncRNAs could also regulate mRNA stability through Staufen 1 (STAU1)-mediated mRNA decay (SMD) by pairing with the 3'UTR of target mRNA to form a double-stranded RNA during SMD. STAU1 recognizes and binds to the dsRNA and results in degradation of the target mRNA. This class of lncRNA has been named as half-STAU1-binding site RNAs (1/2-sbsRNAs) (Gong and Maquat, 2011). Moreover, cytoplasmic lncRNA could also directly interact with STAU1 and the lncRNA terminal differentiation-induced ncRNA (TINCR) is a good illustration. A study revealed the direct binding of TINCR to STAU1 protein, and the TINCR-STAU1 complex has been found to mediate the stabilization of certain mRNA of somatic tissue differentiation genes (Kretz et al., 2013). At the same time, TINCR interacts with a variety of mRNA encoding proteins for cell differentiation via a 25-nucleotide motif called "TINCR box", and renders their stability (Kretz et al., 2013).

LncRNAs could also engage in the regulation of translation processes through interaction with ribosomes. For instance, lncRNA GAS5 interacts with translation initiation factor eIF4E by direct binding through the RNA binding motifs on eIF4E. Meanwhile, GAS5 also binds to c-Myc mRNA, and the overall interactions result in the inhibition of c-Myc protein translation (Hu et al., 2014). Another lncRNA, lincRNA-p21 interacts with partially complementary mRNAs, and enhances the recruitment of translation suppressors Fmrp and Rck, which lead to the suppression of mRNA translation by inhibition of ribosome binding (Yoon et al., 2012). In addition to controlling the translation process of specific gene targets, lncRNAs could also regulate the translation machinery in general. The lncRNA BC1, mainly expressed in germ cells and neurons, is an example. BC1 could inhibit the assembly of translation initiation complexes. It has been found that BC1 transcript could interact with eIF4A and PABP and hence repress translation in Xenopus oocytes (Kindler et al., 2005; Wang et al., 2005).

Furthermore, lncRNAs could influence protein stability at post-translational level by regulating ubiquitination and degradation through ubiquitin-proteasome system. Ubiquitin molecules attach to a substrate protein through ubiquitin conjugation, then the marked protein would undergo degradation and proteolysis into amino acids via the proteasome (Glickman and Ciechanover, 2002). An example that lncRNA regulates protein stability is HOTAIR, which functions as scaffold to interact with E3 ubiquitin ligases containing the RNA-binding domains, Mex3b and Dzip3. This enhances the ubiquitination of their corresponding substrate proteins, Snurportin-1 and Ataxin-1 respectively, thus accelerating their degradation, and reducing the protein levels, and prevents premature senescence *in vitro* (Yoon et al., 2013).

Recently, more attention has been paid to the interaction between lncRNAs and miRNAs, and a concept of competing endogenous RNA (ceRNA) has been introduced (Figure 1.7), which refers to the interplay between different RNA transcripts through the competition for the same pool of miRNAs (Salmena et al., 2011). LncRNAs can indirectly regulate gene expression by binding to miRNAs through miRNA binding sites known as miRNA response elements (MREs) that should be highly complementary to the seed region of target miRNA. Hence, lncRNAs can compete with mRNAs that possess shared MREs for particular miRNA and impede their RISC-mediated degradation to de-repress their expression (Chan and Tay, 2018a). In other words, lncRNAs act as molecular sponge of miRNAs and sequester them to prevent their functions, thereby rescue the expression of their target genes.

For example, lncRNA SNHG5 could enhance ABCC2 expression by binding to miR-205-5p and subsequently promotes imatinib resistance (He et al., 2017). Another well-known proto-oncogenic lncRNA, MALAT1, could sequester miRNA-328 by serving as a miRNA sponge and hence facilitates the drug response of CML cells to imatinib (Wen et al., 2018). To conclude, cytoplasmic lncRNAs mainly function in post-transcriptional, translational and post-translational regulations, which is summarized in Figure 1.6, and a predominant role for lncRNA that has gained increasing awareness is to act as ceRNAs to crosstalk with miRNAs and mRNAs, which is illustrated in Figure 1.7.

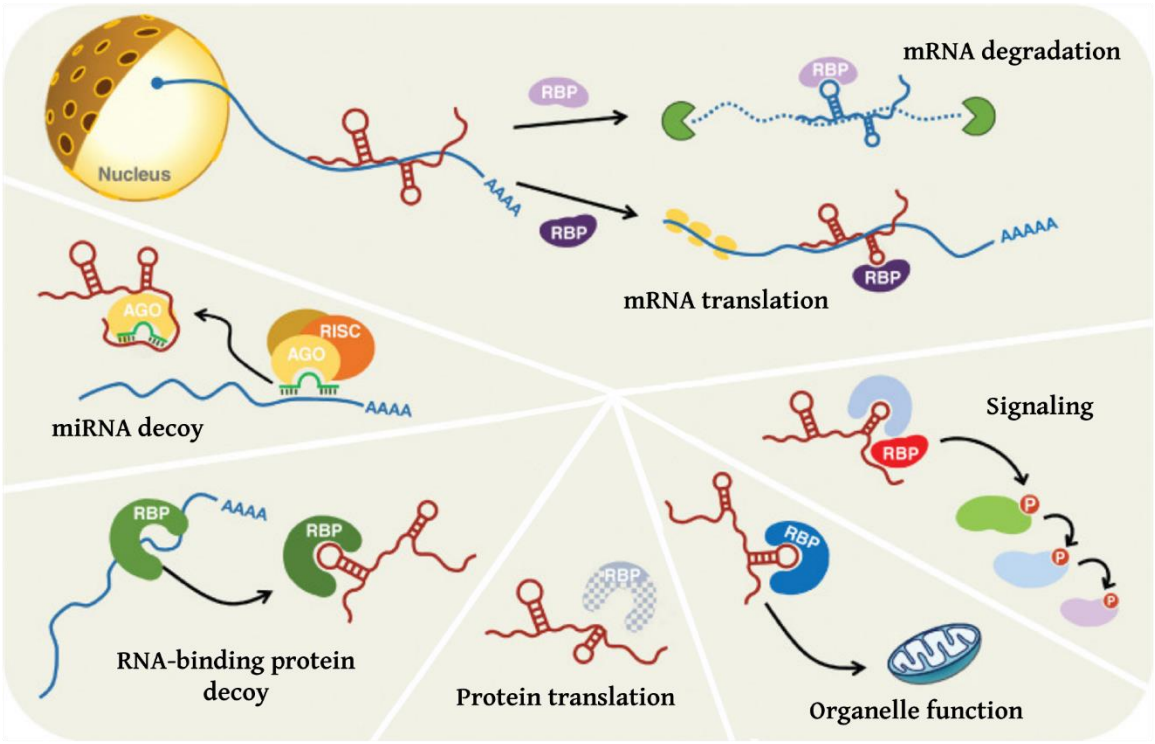


Figure 1.6 Different levels of gene expression regulation by cytoplasmic lncRNAs. Cytoplasmic lncRNAs regulate gene expression at post-transcriptional, translational and post-translational levels via different mechanisms. The major types of mechanisms include the regulation of mRNA stability, modulating translation process and protein stability, and acting as miRNA sponge to mediate ceRNA networks. Figure adapted from Noh et al. (2018).

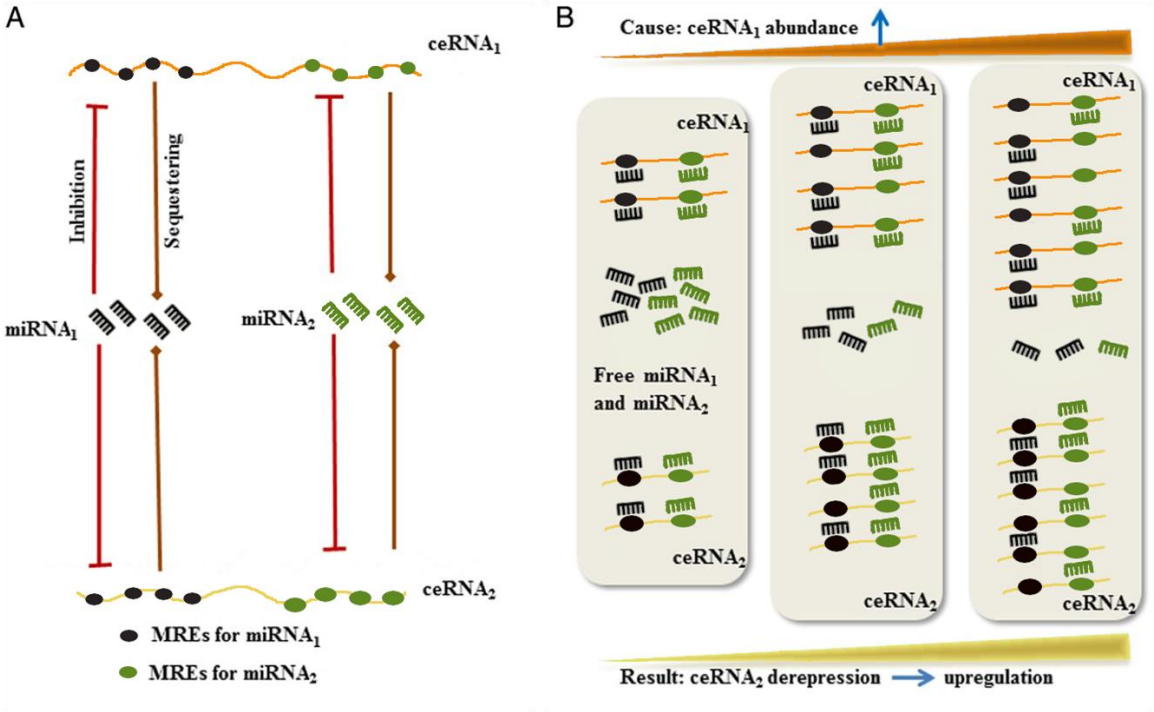


Figure 1.7 Schematic diagram showing the mechanism of ceRNA interplay.

(A) Both ceRNA$_1$ and ceRNA$_2$ contain MREs which allows binding of miRNA$_1$ and miRNA$_2$. Hence, ceRNA$_1$ and ceRNA$_2$ share a common pool of miRNAs that are miRNA$_1$ and miRNA$_2$. (B) When there is an increased abundance of ceRNA$_1$, more miRNAs are bound by ceRNA$_1$, resulting in a de-repression of ceRNA$_2$, and vice versa. Figure adapted from Qi et al. (2015).

1.6.4 LncRNAs and human diseases

As discussed above, lncRNAs have extremely diverse roles in regulation of gene expression through every molecular layer including pre-transcription, transcription, post-transcription, translation and post-translation. Since all these molecular processes are crucial to maintain normal cellular functions, such as cell proliferation, invasion, migration and apoptosis, it is expected that the dysfunction or abnormal expression of lncRNAs could consequently result in the development of various types of human diseases and cancers. Accumulating evidence has demonstrated the recognition of lncRNAs in different human disorders, including leukemia and cancers in lung, liver, breast, prostate, bladder, etc. (Shi et al., 2013; Neguembor et al., 2014; Isin and Dalay, 2015; Sanchez Calle et al., 2018; Zhang et al., 2019).

For example, MEG3 is a renowned lncRNA expressed in most normal tissues and has been considered as important regulators in diverse biological processes but it is also highlighted in numerous human disorders (Zhang et al., 2003; Al-Rugeebah et al., 2019). MEG3 could activate p53, which is a tumor suppressor protein that regulates cell cycle, cell senescence and apoptosis of damaged cells, thereby inhibiting the tumorigenesis and progression of different cancers (Alexandrova and Moll, 2012). A loss of MEG3 expression has been found in different primary tumor cells as well as cancer cell lines, including lung cancers, colorectal cancers, hepatocellular cancers and gliomas (Braconi et al., 2011; Wang et al., 2012; Lu et al., 2013; Yin et al., 2015). This example clearly showed that aberrant expression of certain lncRNAs in specific cell or tissue types could lead to related pathogenesis.

It is not surprising that aberrant lncRNA expression also contributes to different types of leukemia and hematological disorders, such as UCA1 and HOTAIR in acute myeloid

leukemia (AML) (Hughes et al., 2015; Wu et al., 2015; Xing et al., 2015), BALR-6 and HOXA-AS2 in acute lymphoblastic leukemia (ALL) (Rodriguez-Malave et al., 2015; Zhao et al., 2019), DLEU1/2 and MIAT in chronic lymphocytic leukemia (CLL) (Garding et al., 2013; Sattari et al., 2016). Definitely, there are also some lncRNAs implicated in CML. For instance, the expression level of lncRNA MEG3 has been reduced in patients with the advanced phases of CML, which are the accelerated and blast phases, possibly due to the methylation of MEG3 promoter and histone deacetylation (Li et al., 2018b). Ectopic expression of MEG3 could result in inhibited proliferation and apoptosis of CML cells. Its inhibitory effects were promoted through interaction with its target miRNA, miR-147, and suppression of the JAK/STAT signaling pathway in which overactivation of this pathway is linked to various tumor formations. Treatment using a novel histone deacetylase inhibitor, chidamide and a demethylation drug, 5-azacyidine, has resumed the expression of MEG3 in CML (Li et al., 2018b). Besides, MEG3 has also been found to discourage CML cell proliferation and trigger apoptosis via downregulation of another target miRNA, miR-21 (Li et al., 2018a). Moreover, further downregulation of MEG3 has been observed in imatinib-resistant CML cells, as compared to imatinib sensitive cells. By suppressing miR-21, overexpression of MEG3 has downregulated the expression of multidrug resistant transporters, such as MDR1, MRP1 and ABCG2 (Zhou et al., 2017). Therefore, MEG3 can be a potential biomarker or therapeutic target for imatinib-resistant CML.

1.6.5 The role of lncRNAs in cell differentiation

In addition to being implicated in pathological conditions, lncRNAs play vital roles in normal physiological functions. Abundant research has revealed that lncRNAs participate in stem cell differentiation towards different cellular lineages, such as

adipocytes, bone cells, muscle cells, neurons and blood cells (Chen et al., 2020). During adipogenesis, the lncRNA SRA (steroid receptor RNA activator) has been reported to upregulate the expression of a major regulator of adipogenesis, PPARγ. It is the first lncRNA being identified to induce adipogenic differentiation and silencing of SRA has resulted in suppressed differentiation of pre-adipocyte cell line (Lanz et al., 2003). In osteogenesis, lncRNA H19 has been found to promote the formation of bones by upregulating the expression of osteogenic related genes via interacting with miR-22 and miR-141, and attenuating their inhibitory effects. As these two miRNAs could inhibit the expression of β-catenin while the Wnt/β-catenin signaling is necessary for osteoblast development, H19 stimulates bone formation by counteracting the inhibition of Wnt/β-catenin signaling (Liang et al., 2016).

For myogenesis, a muscle-specific lncRNA linc-MD1 stimulates muscle generation by acting as molecular sponge for the miR-133 and miR-135 and reduces their inhibitory regulation on the expression of myogenic transcription factors, such as MAML1 and MEF2C (Cesana et al., 2011). Another example lncMyoD has a highly specific expression in the early differentiation of myoblasts, and it is located upstream of the MyoD gene. The absence of lncMyoD could suppress the terminal differentiation of myoblasts by influencing the cell cycle processes. LncMyoD expression is activated by MyoD and it could directly bind to the protein IMP2, and suppress IMP2-mediated translation of genes responsible for cell cycling and cell differentiation (Gong et al., 2015). Regarding neurogenesis, TUNA is a lncRNA exclusively expressed in central nervous system and it is involved in both maintenance of pluripotency in neuronal stem cells and stimulation of neural differentiation. Depletion of this lncRNA has led to the

reduced expressions of various markers for neural progenitors and globally affected hundreds of genes related to neural lineage differentiation (Lin et al., 2014).

Researchers have also revealed that lncRNAs are highly involved in hematopoiesis, especially the development of myeloid lineage. The lncRNA EGO is highly expressed in mature eosinophils and it plays a crucial role in eosinophil development. Knockdown of EGO reduced the expression of certain basic proteins and neurotoxins that are required for eosinophil development. This suggests that EGO functions to promote the eosinophil differentiation in hematopoiesis (Wagner et al., 2007). Besides, lncRNA could also regulate the production of monocytes and granulocytes. HOTAIRM1 is a myelopoiesis-associated regulatory lncRNA that is highly upregulated in myeloid progenitors during granulocytic differentiation, and it is transcribed in antisense to the HOXA genes. Functional studies demonstrated that HOTAIRM1 depletion repressed HOXA1 and HOXA4 expression and hence attenuated the transcription of CD11b and CD18, which are myeloid differentiation related genes, and thereby impaired myeloid cell differentiation (Zhang et al., 2009).

The involvement of lncRNAs in stem cell differentiation does not only restrict to the previously mentioned cell types only. There are considerable numbers of lncRNAs that have been investigated to be involved in stemness or differentiation regulation, which are summarized in Figure 1.8. Since the stemness property and differentiation of pluripotent stem cells are regulated by complicated and multilayered networks including lncRNAs, understanding of such regulatory network is essential for future applications of pluripotent stem cells in clinical applications.

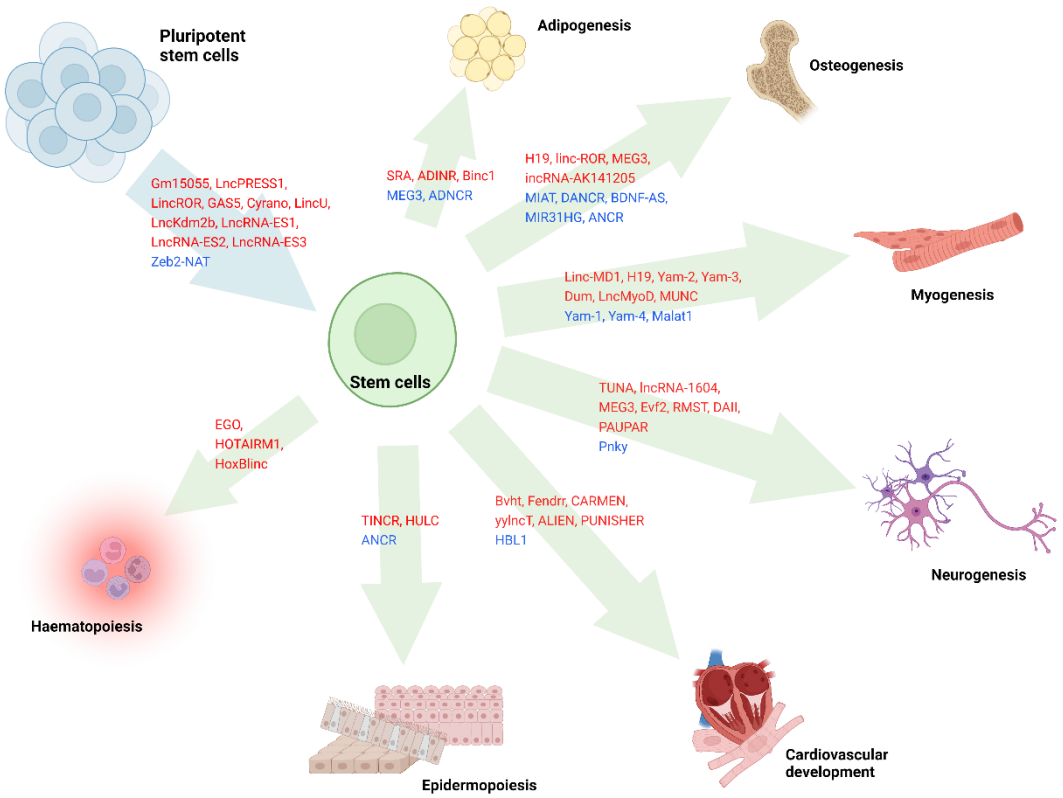


Figure 1.8. A summary of lncRNAs involved in pluripotent stem cell differentiation.

Examples of lncRNAs that are involved in differentiation of different cell lineages are shown. Red and blue colors indicate their positive and negative regulation, respectively. Figure adapted from Chen et al. (2020).

1.7 Induced pluripotent stem cells

Induced pluripotent stem cells (iPSCs) refer to the reprogrammed somatic cells which possess pluripotency similar to embryonic stem cells that are characterized by their unlimited self-renewal and proliferative abilities, as well as the capacity of differentiating into the three primary germ cell layers including endoderm, mesoderm and ectoderm, and hence virtually all cell types of the human body (Takahashi and Yamanaka, 2013; Braganca et al., 2019). Following the introduction of the iPSC technology in 2006 by Takahashi and Yamanaka's team, leading to an unprecedented breakthrough in stem cell biology and regenerative medicine, iPSCs have been widely applied for the development of stem cell transplantation therapy, drug discovery, disease modelling, and basic science investigations (Takahashi and Yamanaka, 2013; Aboul-Soud et al., 2021). Theoretically, iPSCs can be generated through reprogramming of any somatic cell types using suitable approach to integrate a combination of reprogramming factors into the somatic cells. It has been reported that many types of somatic cells are useful sources in producing iPSCs, of which fibroblasts, peripheral blood cells, and cord blood cells remain the most common because of their accessibility and low invasiveness for collection (Seki et al., 2010; Singh et al., 2015b; Kim et al., 2016).

The first iPSCs were produced by employing a cocktail of four reprogramming factors, also known as Yamanaka factors, consisting of octamer-binding transcription factor 4 (OCT4), sex determining region Y-box 2 (SOX2), Kruppel-like factor 4 (KLF4) and c-MYC (collectively OSKM), via retroviral transduction into adult fibroblasts (Takahashi and Yamanaka, 2006). To date, various approaches including different combination of reprogramming factors, the usage of chemicals, miRNAs, peptides and proteins have

been proposed for the generation of iPSCs (Malik and Rao, 2013). A recent research finding has shown a synergistic effect to OSKM-mediated iPSC reprogramming by introducing NANOG and LIN28 through the activation of LIN41 and Wnt/β-catenin signaling, which enhances the development of initial iPSC colonies (Wang et al., 2019). The same study also revealed an addition of a histone methyltransferase inhibitor iDOT1L could further reinforce the reprogramming efficiency.

Besides the reprogramming cocktails, different delivery approaches have been investigated to show different reprogramming efficiencies, and they can be divided into two main categories, integrating and non-integrating methods (Wang et al., 2021). In the past, reprogramming factors for generating iPSCs were mainly introduced by retroviral transduction that may result in genomic integration of the delivered transgenes. However, it raised the concerns about genome integrity and led to the exploration and development of non-integrating methods for delivering reprogramming factors. A recent non-integrating approach for generating iPSCs is to express reprogramming factors via delivery of episomal DNA vectors, which are non-viral and non-integrating plasmid-based approach, and therefore are safe to use and inexpensive (Yu et al., 2009; Singh et al., 2015a; Aboul-Soud et al., 2021). More recently, a new method of optimizing human iPSC reprogramming has been developed in which mRNAs or miRNAs are directly delivered to the cells instead of DNA or viruses, thereby preventing the risk of insertional mutagenesis due to viral intermediates (Wang et al., 2021). By a single transfection, the engineered single-strand RNAs consists of multiple reprogramming factors can be delivered into cells efficiently.

1.7.1 Disease modeling using iPSCs

It is crucial to gain comprehensive understanding of the underlying mechanisms related to the human disease pathogenesis for the development of therapeutic strategies. However, the lack of appropriate biological models for relevant diseases has created a significant obstacle for the development of effective therapies. Through decades of study, animal models have been extensively used to study human diseases *in vivo* to explore their mechanisms and develop therapies. Nevertheless, animal models cannot fully reflect the exact pathologies of human disorders due to the fundamental differences between different species genetically, resulting in a number of failed clinical trials when the treatment is switched from animals to humans (Sayed et al., 2016). For instance, in the treatment of amyotrophic lateral sclerosis (ALS), numerous medications that are previously shown to have therapeutic values in rodent ALS models, were reported to have no effects in human, indicating the importance of human originated disease models (Desnuelle et al., 2001; Shefner et al., 2004). Although using primary cells from patients as disease models can accurately reflect the relevant human diseases, the limited capacity of expansion hinders their application, especially neural cells and cardiac cells that can only be collected through invasive process (Sayed and Wu, 2017; Liu et al., 2018). Due to the ease of access and unlimited potential of expansion and differentiation that human iPSCs possess, they have become a promising alternative for disease modeling and drug discovery. This created unlimited supplies of personalized disease-specific cells and facilitated direct clinical implementation of precision medicine.

Currently, iPSC technology has been adopted in different studies about pathogenesis of inherited genetic diseases and development of novel targeted therapies, and recent

cellular models of iPSC have been reported to be established from both normal and malignant tissue cells (Carette et al., 2010; Singh et al., 2015a). For example, a study has demonstrated the generation of iPSC from $CD34^+$ peripheral blood cells of PV and PMF patient with *JAK2*-V617F mutation which can serves as an MPN disease model (Ye et al., 2009). Normal phenotypes, karyotype and pluripotency were observed in undifferentiated iPSCs derived from MPN patients. However, enhanced proliferation and skewed tendency to form erythroid colonies were observed upon directed hematopoietic differentiation (Ye et al., 2009). Another research also utilized PV patient-derived iPSCs combined with directed hematopoietic differentiation towards $CD34^+$ progenitors in the study. With this model and subsequent examination, the researchers revealed that *JAK2*-$V617F^+$ progenitors are strongly dependent on DUSP1 activity for proliferation and survival under stress condition, and high DUSP1 expression provides adaptation of the *JAK2*-V617F+ progenitors to inflammatory stress (Stetka et al., 2019). These collective findings evidently showed iPSC could serve as a powerful research tool on hematopoietic differentiation and the investigation of molecular mechanisms.

1.7.2 Challenges and considerations of iPSC model

In spite of the numerous advantages of iPSC technology, there are limitations and challenges that need to be overcome before maximization of their application. To begin with, the efficiency of reprogramming somatic cells into iPSCs is very low according to current protocols, typically less than 1%, and the reprogramming efficiency becomes even lower in aged cells or cells having a high tendency of cell divisions (Strassler et al., 2018). Secondly, iPSCs and ESCs have been reported to have different gene expression profiles. The full understanding of the effects of the differentially expressed

genes remains critical before iPSCs can be used to replace ESCs in different medical applications, especially cell therapies (Ho et al., 2011; Turhan et al., 2021).

Moreover, a remarkable issue of iPSC reprogramming is the use of lentiviral or retroviral vector for introducing the essential transcription factors into somatic cells, which could result in the random integration of viral DNA in any loci of the host cell's genome, leading to a potential disruption of the functional protein coding sequence, or promoter and enhancer regions (Fernandez Tde et al., 2013). The use of oncogene (e.g. c-MYC and KLF4) for the generation of iPSCs is another concern since it may result in a higher risk of cancer development. Despite the expression of the four mentioned transcriptional factors is silent in established iPSC cell lines, there has been study showing that reactivation or residual expression of these reprogramming factors could induce tumor formation in mice model (Saha and Jaenisch, 2009), thereby the suitability of iPSC for cell replacement therapy still needs further investigation. Besides, it has been suggested that iPSCs showed genomic instability in early passages as a result of genetic reprogramming, which require frequent genomic monitoring to ensure the stability of phenotype as well as clinical safety (Laurent et al., 2011).

In addition, it has been revealed that both the reprogramming approach and the choice of somatic cells for reprogramming could influence the differentiation capability of the resulting iPSCs. Attenuated hematopoietic differentiation capacity and abnormal hypermethylation may be acquired during the process of reprogramming when retroviral methods are used to produce iPSCs (Nishizawa et al., 2016). Basically, peripheral blood mononuclear cells and fibroblasts remain the gold standard for iPSC generation owing to the ease of reprogramming and their isolation method is non-

invasive. Nevertheless, it has been shown that iPSCs derived from blood cells are less likely to acquire abnormal methylation during reprogramming than that from fibroblasts and have higher capacity of hematopoietic differentiation when compared with those derived from other somatic tissues (Wattanapanitch, 2019). Therefore, this finding provides crucial insight into the selection of reprogramming strategies and starting cell types to produce iPSCs for differentiation study especially in hematopoietic lineage.

1.8 Scope of the current study

Classical MPNs are a rare group of hematological disease and has a relatively high survival rate compared to other leukemia. However, patients without effective treatment will eventually progress into more severe bone marrow failure or transform into acute leukemia. Nowadays, many molecular details of this disease, particularly the area of non-coding elements, are not clear yet and await exploration. Different from CML that some lncRNAs has been reported being involved in BCR-ABL1 tumorigenesis, there is still no lncRNA implicated in Ph-negative MPNs. Given that many known lncRNAs have been investigated to be associated with human diseases, it is very probable that they could also play a pivotal role in Ph-negative MPNs. Thereby, one of the study aims is to identify any lncRNAs that may be involved in the malignant signaling pathway of Ph-negative MPNs.

On the other hand, although the existing TKIs are quite effective for treatment of CML patients in chronic stage, the development of drug resistance is an obstacle for effective therapy. Recent research has demonstrated some lncRNAs are critically involved in the development of imatinib resistance in CML, such as HOTAIR (Wang et al., 2017), UCA1 (Xiao et al., 2017b), and MEG3 (Zhou et al., 2017), which suggested that there might exist an unique lncRNA network in the modulation of imatinib resistance. Therefore, this study also aims to focus on the identification and characterization of any novel functional lncRNAs associated with the development of imatinib resistance in CML, in order to make a positive impact on the early diagnosis and treatment of CML.

In view of the current lack of curative treatments for both CML and MPNs, it is indispensable to further improve our knowledge and understanding of the disease and

therapeutic mechanisms by studying novel regulators such as non-coding RNAs. Since lncRNAs provide an additional layer of gene expression control and play a crucial regulatory role in various physiological processes, they have the potential to be used in the future as biological markers for disease diagnosis and treatment. In summary, exploring the non-coding area of MPN diseases may confer deeper understanding of the molecular interactome in the aberrant signaling pathways and thus provide more insights into the pathogenesis and progression or treatment of these blood malignancies. Meanwhile, as lncRNAs have yet to catch up to miRNAs in terms of knowledge and understanding, this study may provide more valuable information to drive the field forward.

Chapter 2 – Materials and Methods

2.1 Leukemic cell culture and drug treatment

The human CML cell lines K562 (#CCL-243), LAMA84-IMS (#CRL-3347) and LAMA84-IMR (#CRL-3348) were purchased from the company American Type Culture Collection (ATCC). The Human erythroleukemia (HEL) cell line was kindly provided by Prof. SP Yip (HKPU, HK). The SET-2 cell line (#ACC608) was purchased from Leibniz Institute DSMZ. All the cells were cultured in RPMI-1640 (Gibco) supplemented with 10% fetal bovine serum (FBS), except SET-2 required 20% FBS, and 1% penicillin-streptomycin, and incubated at 37°C with 5% CO_2.

For the *in vitro* JAK2 inhibition assays, 1×10^5/mL HEL cells were treated with 1 μM ruxolitinib (Selleck Chemicals) for 48 hours. Imatinib-resistant K562 cells (K562-IMR) were generated in-house with increasing concentrations of imatinib (Selleck Chemicals) treatment starting from 0.5 μM. Cells were subcultured twice per week with 0.5 μM increment of drug dose. The final cell population with stable survival ability under 10 μM imatinib treatment was acquired and considered as K562-IMR.

For the cell viability assays, 5×10^5/mL cells were seeded in 12-well plates in triplicates, and then treated, respectively, with 3 μM, 5 μM and 10 μM imatinib for 72 hours. After washing with 1X PBS, trypan blue exclusion method was used to determine the cell viability with the aid of the Countess II Automated Cell Counter (ThermoFisher Scientific).

2.2 RNA isolation and RT-qPCR

Total RNA was isolated by Trizol/RNeasy hybrid method with DNase treatment by combining the use of TRIzol reagent (Life Technologies), RNeasy Mini Kit (QIAGEN) and RNase-Free DNase Set (QIAGEN). The quantity and quality of RNA were examined by NanoDrop spectrophotometer (ThermoFisher Scientific). miRNA was isolated with QIAzol™ lysis reagent (QIAGEN) and miRNeasy Isolation Kit (QIAGEN). The extracted miRNA was quantified by a Qubit 2.0 Fluorometer (ThermoFisher Scientific) with the use of Qubit microRNA assay kit (ThermoFisher Scientific).

First-strand cDNAs were generated using ReverAid First Strand cDNA Synthesis Kit (ThermoFisher Scientific). Reverse transcription (RT) for miRNA was performed with the miScript II RT Kit (QIAGEN). 0.5-1 μg total RNA from each sample was used for all RT reactions. The cDNA was used as the input template for real-time qPCR reaction, which was performed on ViiA7 real-time cycler (Applied Biosystems) with the QuantiNova SYBR Green PCR mastermix (QIAGEN) following the manufacturer's instruction. In brief, the PCR reaction was carried out with initial activation at 95°C for 2 min, and then 40 cycles of 95 °C for 30 s and 60 °C for 60 s. Melting curve analysis was also included by default setting in each run for monitoring the specificity of reaction. Relative gene expression change between samples was calculated by $2^{-\Delta\Delta Ct}$ method and GAPDH expression was used for normalization. All the primers used are listed in Table 2.1. For miRNA detection, specific miScript primer assay (QIAGEN) was used with miScript SYBR Green PCR kit (QIAGEN) according to the manufacturer's protocol.

For lncRNA qPCR array, cDNA was first diluted with nuclease-free water and then added to the RT2 SYBR Green Mastermix (QIAGEN). Next, 25 µl of the PCR mix was added to each well of the RT2 lncRNA PCR Array (QIAGEN). The qPCR reaction was performed on ViiA7 real-time cycler (Applied Biosystems) using the thermal cycling profile as follows: 95°C for 10 min; 40 cycles of 95°C for 15 s and 60°C for 1 min; default melting curve stage.

Table 2.1 List of primers used in the study.

Primer sequences (5' → 3')		
qPCR primer		
BANCR	Forward	TGA GCC TCT ATT GGA ATC AGC
	Reverse	GCC AGG GAT GAC TTG CGT ATA
CBR3-AS1	Forward	TGT GAG GGA GCG GGA GTC T
	Reverse	GCG TCT GGA TGA GAA GAG GAA A
LINC00261	Forward	CTG GGC AGA GAC TAC AAA ACA A
	Reverse	CCT TTG CTT TCC TCC AAG ACA A
LINC00887	Forward	ACA GAT GAG AGA GAA CTG ATG C
	Reverse	TCA TGT TTA AGA GGA GGC TGC T
LUCAT1	Forward	CAT GCT GAG CTA CAG AGT TTC G
	Reverse	GTT GGA TTC CTG GGT GTG GT
NBR2	Forward	TCG CTA CCT ATT GTC CAA AGC A
	Reverse	TCA AAT GAA ACT TTT ACC GAA ACT GG
H19	Forward	ATC GGT GCC TCA GCG TTC G
	Reverse	CTC TGT CCT CGC CGT CAC A
MDR1	Forward	ATG CTA TAA TGC GAC AGG AG
	Reverse	CCC AGT GAA AAA TGT TGC CA
LNC000093	Forward	GCA TTA CCT CTA TCC TAC CTG G
	Reverse	GGT TGT GTT TTA CTC CTC GCA GA
CD34	Forward	CCA CAA CAA ACA TCA CAG AAA CGA
	Reverse	GGT GGT GAA CAC TGT GCT GAT TA
CXCR4	Forward	TGC CCT CCT GCT GAC TAT TC
	Reverse	CCA ACC ATG ATG TGC TGA AAC TG
GATA2	Forward	GAA CCG ACC ACT CAT CAA GC
	Reverse	CGT CTG ACA ATT TGC ACA ACA GG
GAPDH	Forward	AGG TCG GAG TCA ACG GAT TTG
	Reverse	TGA AGG GGT CAT TGA TGG CAA CA

2.3 RNA-seq and data analysis

Total RNA was extracted as described in Chapter 2.2. The RNA integrity number (RIN), which indicates the quality of RNA, was determined using an Agilent 2100 Bioanalyzer (Agilent) with the use of Agilent RNA 6000 Nano kit (Agilent). All samples for RNA-seq have an RIN number > 9.5, which represented very high quality of the isolated RNA. Library preparation and poly(A)-enriched mRNA sequencing was performed by Groken Bioscience Ltd. (Hong Kong). In short, rRNAs removal was done with the Ribo-Zero™ magnetic kit, and cDNA libraries were prepared with the NEBNext Ultra Directional/non-Directional RNA Library Prep Kit (NEB). The next-generation sequencing was conducted in the Illumina HiSeq platform.

The raw data quality of RNA-seq was evaluated using FastQC (Andrews, 2010). Then, low-quality reads and adapters were trimmed by Fastp (Chen et al., 2018). After quality checking, sequencing reads were mapped to the human reference genome GRCh38 by Hisat2 (Kim et al., 2015). Aligned reads were assembled and identified to be known or novel transcripts using StringTie (Pertea et al., 2015) with the human reference annotation file (Cunningham et al., 2019). Novel transcripts were then filtered to be classified as lncRNAs based on the following criteria: (1) transcript length ≥ 200 bp, (2) transcript exon number ≥ 2, and (3) non-coding potential features supported by the following all 4 software applications, namely CNCI (Sun et al., 2013), CPAT (Wang et al., 2013a), CPC2 (Kang et al., 2017), and PfamScan (Mistry et al., 2007). Differentially expressed genes (DEGs) were identified by DESeq2 (Love et al., 2014). Functional annotation of DEGs were analyzed by the Metascape database (Zhou et al., 2019) to indicate their biological functions and pathways. The co-expression network was calculated from DESeq2 normalized gene read counts with a threshold of Pearson

correlation coefficient > 0.9 and adjusted *p*-value < 0.05. The network was then integrated and visualized using Cytoscape (Shannon et al., 2003). A flow chart summarized the RNA-seq data analysis up to the identification of DEGs is shown in Figure 2.1. The name of bioinformatics tools is labeled in blue. The transcript and DEG numbers shown in the flow chart are in regard to the RNA-seq for HEL cells (Chapter 3.2).

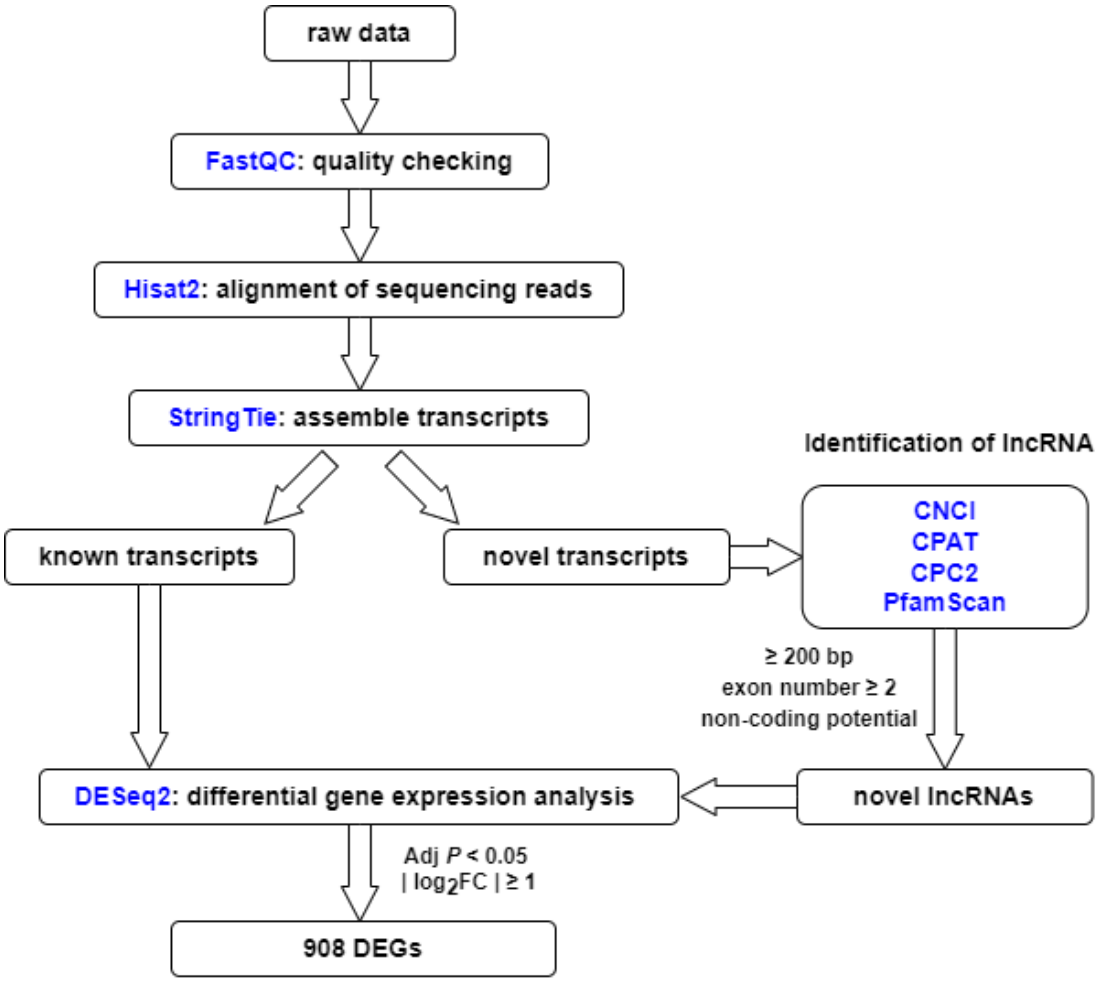


Figure 2.1 Flow chart showing the RNA-seq data analysis.

2.4 Whole-genome sequencing and copy number variant identification

DNA libraries for whole-genome sequencing (WGS) were prepared by the MGIEasy FS PCR-Free Library Prep Set (MGI) with DNA samples of eight MPN patients (patient information is shown in Table 2.2) following the manufacturer's manual. In brief, DNA samples were subjected to fragmentation and adapter ligation in the beginning of library construction. After heat-denaturation, splint oligonucleotides that are complementary to the 5' and 3' terminal ends of the single-stranded DNA were hybridized to both ends to form a nicked circle, which was then ligated to produce a single-stranded circle DNA. The remaining linear DNA was digested by exonuclease to complete the library for the subsequent sequencing with DNA-Nanoball technology (MGI). The sequencing was conducted using the MGISEQ-2000 platform with DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150).

For data analysis, remove low-quality reads were filtered after trimming the barcodes and sequencing adaptors using Fastp (Chen et al., 2018). Clean reads were mapped to human reference genome GRCh38 using BWA alignment tool to generate SAM files. The Genome Analysis Toolkit (GATK 4.1.6.0) was used to convert SAM files into compressed BAM files, sort the BAM files by genomic coordinates, and remove duplicates. Copy number variations (CNVs) were computed by CNVKit package (Talevich et al., 2016).

Table 2.2 Information of specimens collected from MPN patients.

MPN patients	**(*BCR-ABL1*-negative)**				
Sample No.	**WBC (10^9/L)**	**RBC (10^{12}/L)**	**PLT (10^9/L)**	**Mutation type**	**Interpretation**
008	9.84	5.12	45.5	JAK2 V617F	PMF/PV
012	21.8	2.32	57.2	-	PMF
014	7.21	2.68	713	CALR 52del (Type I)	PMF
027	32.7	2.92	506	CALR 52del (Type I)	PMF
118	4.97	2.77	200	CALR 39del	PMF
125	7.44	3.84	129	JAK2 V617F	PMF
126	19.3	3.28	298	CALR 52del + point mutation	PMF
144	7.71	2.76	84.1	JAK2 V617F	PMF

2.5 Cycle sequencing

Plasmids, cDNA and gDNA templates were sequenced using an ABI 3130 Genetic Analyzer Sequencer with the BigDye Terminator v1.1 Cycle kit (Applied Biosystems). In brief, each cycle sequencing reaction mixture in a total volume of 10 μl was prepared as follows: 2 μl of BigDye Terminator v1.1 reaction mix, 1 μM of either forward or reverse sequencing primer, DNA template (500 ng for plasmid, 50 ng for gDNA or cDNA) and nuclease-free water. The reaction mixture was denatured at 96°C for 1 min, followed by 40 cycles of 96°C for 10 s, 50°C for 30 s, and 60°C for 6 min. The products were precipitated by mixing with 95% ethanol and 3M sodium acetate (pH 4.6). After washing with 70% ethanol and drying by CentriVap Complete Vacuum Concentrator (Lfabconco), the DNA pellet was resuspended in Hi-Di formamide (Applied Biosystems) and incubated in dark for 20 min. The samples were denatured at 95°C for 3 min before being analyzed by ABI 3130 sequencer.

2.6 Cellular fractionation

A total of 1×10^6 cells were harvested and washed twice with cold PBS and then incubated in 300 μl of hypotonic buffer (50 mM HEPES, pH 7.5, 1 mM EDTA, 350 mM sucrose, 10 mM KCl, 1 mM DTT and 0.1% Triton X-100) on ice for 10 min. After centrifugation at 2,000 g for 5 min at 4°C, the supernatant was collected and considered as the cytoplasmic fraction. After one additional washing with hypotonic buffer, the remaining nuclear pellet was resuspended in 300 μl of lysis buffer (10 mM HEPES, pH 7.0, 0.5% NP-40, 5 mM $MgCl_2$, 100 mM KCl, 10 μM DTT and 1 mM PMSF) and incubated on ice for 10 min. The lysate was considered as nuclear fraction. Then, 1 mL TRIzol reagent was added to each fraction and total RNA was isolated according to Chapter 2.2.

2.7 Droplet digital PCR (ddPCR)

cDNA samples for ddPCR were synthesized by RT reaction as described in Chapter 2.2. ddPCR assays were performed in 20 µl reaction mixture for each sample, which composed of 50 ng cDNA template, 10 µl 2X ddPCR EvaGreen Supermix (Bio-Rad), 250 nM each of forward and reverse primers, and nuclease-free water. The reaction mix and droplet generation oil (Bio-Rad) were loaded in the DG8 cartridge (Bio-Rad), and droplet mixtures were generated using QX200 droplet generator (Bio-Rad). 40 µl of droplet mix was then gently transferred into a 96-well PCR plate. The sealed 96-well plate was subjected to following thermal cycling conditions in a Veriti thermocycler (Applied Biosystems): 95°C for 10 minutes, 40 cycles of 94°C for 30 s and 60°C for 60 s, followed by a signal stabilization step at 4°C for 5 min and 90°C for 5 min, and finally holding at 4°C until data acquisition. 2°C/s ramp rate was applied throughout the cycling process to ensure oil droplets reached the correct temperature. Then the plate was analyzed by the QX200 droplet reader (Bio-Rad).

2.8 Computational prediction of miRNA binding region

The computational prediction was performed using STarMir, which is a bioinformatics tool to predict potential binding sites for one or more miRNA on specific RNA sequence (mRNA or ncRNA) (Rennie et al., 2014). The prediction models were developed based on high throughput miRNA binding data from crosslinking immunoprecipitation (CLIP) experiments. STarMir is an online tool freely available at http://sfold.wadsworth.org/starmir.html .

2.9 Cell transfection and luciferase reporter assay

K562 cells were seeded in 6-well plates at a density of 1×10^6 cells/mL, then 50 nM miR-675-3p and miR-675-5p mimics (Dharmacon) were transfected into cells using Lipofectamine 2000 reagent (Invitrogen). The mirVana™ miRNA mimic (Invitrogen) was transfected into the negative control group in the same manner. After 24 hours of incubation, the culture medium was replaced with fresh RPMI1640 medium supplemented with 10% FBS. Cells were harvested at 48 hours post-transfection for extracting RNA or protein that would be used in subsequent assays.

Full-length LNC000093 sequence containing the wild-type miR-675-5p putative binding site (5'-UGCACC-3') or mutant binding site (5'-GUACAA-3') was synthesized and inserted into the parental luciferase reporter vector PGL3-CMV-LUC-MCS (Genomeditech, Shanghai, China). After confirming the sequence by Sanger sequencing, the luciferase vector was amplified and then extracted by the EasyPrep HY-Midi Plasmid Extraction Kit (Biotools, Taiwan). Both the luciferase vector and the Renilla luciferase control reporter vector were co-transfected into K562 cells in the presence of either miR-675-5p mimic or miR-675-3p mimic, or negative control

miRNA. Cells were harvested and lysed after 48 hours of transfection and luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega). Relative luciferase activity was calculated from the ratio of target Firefly to Renilla luciferase.

2.10 Protein extraction and western blotting

After harvesting cells with PBS washing twice, total protein was extracted using RIPA lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% NP-40) containing protease and phosphatase inhibitor. After incubation for 30 min on ice with occasional vortexing, the lysate was then centrifuged at 15,000 g for 30 min at 4°C. The extracted protein was quantified with the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. In total, 20-40 μg proteins were resolved by 8% or 10% Bis-Tris polyacrylamide gels at room temperature, and transferred to polyvinylidene fluoride (PVDF) membrane by wet transfer on ice. The PVDF membrane was incubated in blocking buffer (PBS + 5% non-fat dry milk + 0.05% Tween-20) for 1 hour at room temperature, and then probed with appropriate primary antibodies at 4°C overnight. After washing with PBS containing 0.1% Tween-20, the membrane was incubated with HRP-conjugated secondary antibodies at room temperature for 1 hour. Signals were visualized with enhanced chemiluminescence (ECL) substrates. Antibody information is listed in Table 2.3.

Table 2.3 List of antibodies used in the study.

Name	Species	Company	Cat. No.
Antibodies			
RUNX1	Mouse	Santa Cruz Biotechnology	sc-365644
β-actin	Mouse	Santa Cruz Biotechnology	sc-47778
JAK2	Rabbit	Cell Signaling Technology	3230
Phospho-STAT3	Rabbit	Cell Signaling Technology	9131
Phospho-STAT5	Rabbit	Cell Signaling Technology	9359
STAT3	Rabbit	Santa Cruz Biotechnology	sc-482
STAT5	Mouse	Santa Cruz Biotechnology	sc-74442
Anti-rabbit IgG, HRP-linked	Rabbit	Cell Signaling Technology	7074
Anti-mouse IgG, HRP-linked	Mouse	Cell Signaling Technology	7076

2.11 CRISPR-Cas9 genome editing

For CRISPR-Cas9-mediated deletion of H19/miR-675, a plasmid-based system was used. A pair of single-guide RNA (sgRNA) was designed to target the first exon of H19. The sgRNA oligonucleotides were separately cloned into pAll-Cas9.Ppuro (Academia Sinica, Taiwan) which is an all-in-one backbone vector that can express both sgRNA and Cas9 nuclease. Transfection was performed using the Neon Transfection system (Thermo Fisher Scientific). For each reaction, 3×10^5 K562 cells were collected, washed twice with 1X PBS, and centrifuged at 200 g for 5 min. After discarding supernatant, the cell pellet was resuspended in 10 μl resuspension Buffer R containing 1 μg of each H19-CRISPR-Cas9 vector and subjected to electroporation. The electroporated cells were immediately transferred to a 24-well plate containing prewarmed complete RPMI medium.

For LNC000093-deletion in iPSCs, the Alt-R CRISPR-Cas9 system (IDT) was applied. A pair of custom sgRNAs was designed to target the genomic regions flanking LNC000093 using proprietary algorithms from IDT. The synthetic sgRNAs were mixed with Alt-R® S.p. Cas9 Nuclease V3 (IDT) to form ribonucleoprotein (RNP) complexes, which were then incubated with Lipofectamine RNAiMAX reagent (Invitrogen) to assemble transfection complexes. In total, 10 nM RNP complexes were transfected into iPSCs at a density of 3×10^5/well in 24-well plates.

For all CRISPR-Cas9-edited cells, gDNA was extracted by FlexiGene DNA Kit (QIAGEN), and PCR as well as gel electrophoresis were further performed to confirm the deletion effect. In brief, a pair of primer flanking the deleted genomic region was designed for PCR, and it is expected that an extra PCR product with smaller size would

appear which represents the population with successful CRISPR-Cas9-deletion. The sequences of sgRNA and primers for validating CRISPR-deletion are listed in Table 2.4.

Table 2.4 SgRNAs for CRISPR-Cas9 system and primers for validation.

Target	sgRNA sequences (5' → 3')	PAM site
H19 sgRNA1	GCTAGGACCGAGGAGCAGGGTG	AGG
H19 sgRNA2	GATCGGTGCCTCAGCGTTCGGGC	TGG
LNC000093 sgRNA1	TCCCAACTCCCACGTTAGAG	TGG
LNC000093 sgRNA2	ACATTTTTGCTTAAGAACTT	CGG
PCR primer	**(5' → 3')**	
H19-CRISPR Forward	GGG CCA CCC CAG TTA GAA AA	
H19-CRISPR Reverse	GTC CTG CTT GTC ACG TCC AC	
LNC000093-CRISPR Forward	ATG TTG GTG TAT CTT GAG ATC CTC	
LNC000093-CRISPR Reverse	TCC CCA GTT GTA CTC CAT CTG T	

2.12 iPSC culture and differentiation

Human iPSCs (ALSTEM, CA, USA) were cultured in growth medium consisting of Essential 8 (E8) medium with completed supplements (Gibco). Upon 80% confluence, cells were subcultured into new culture plates pre-coated with Geltrex (Gibco) diluted with DMEM/F12 (Gibco) in a ratio of 1:50.

iPSCs were differentiated to embryoid body (EB) by spin EB formation method, and Essential 6/Polyvinyl Alcohol Medium (E6/PVA) was used to support the spontaneous differentiation. In brief, culture medium was discarded, and cells were rinsed once with warm PBS. Then, 0.5 mL per well ReLeSR (STEMCELL Technologies) was added and then removed within 1 min so that the cells were exposed to a thin film of ReLeSR solution. After incubation for 7-9 min at room temperature, 1 mL per well of E6/PVA medium was added, and colonies were detached by tapping the side of culture plate for 30-60 s. Then, iPSC aggregates were dissociated into single cells by gently pipetting with 1-mL autopipettes. Next, cells were counted and diluted to the appropriate cell densities with E6/PVA containing 10 μM Y-27632 (STEMCELL Technologies). Suspended iPSCs were seeded to a 96-well plate at a density of 8,000 cells/well in 100 μL medium. Next, cells were spun down to the bottom of the plate by centrifugation at 300 g for 5 min, and then incubated at 37°C with 5% CO_2.

For directed hematopoietic differentiation, EBs were generated as mentioned above for 4 days and then transferred to a 6-well plate containing differentiation medium consisted of StemPro™-34 SFM (Gibco™) supplemented with 2 mM glutamine, 50 μg/mL ascorbic acid, 150 μg/mL transferrin, 0.4 mM monothioglycerol and varying combination of cytokines in different days as follows. Days 4 to 9: 30 ng/mL BMP4,

50 ng/mL VEGF, 50 ng/mL SCF, 50 ng/mL TPO, 50 ng/mL Flt3L, 20 ng/mL bFGF; Days 9 to 14: 30 ng/mL BMP4, 50 ng/mL SCF, 50 ng/mL TPO, 10 ng/mL IL-3, 5 ng/mL IL-11, 2 U/mL EPO, and 125 ng/mL IGF. A schematic diagram of different differentiation schedule is shown in Figure 2.2. The information for the supplements used in iPSC differentiation is listed in Table 2.5.

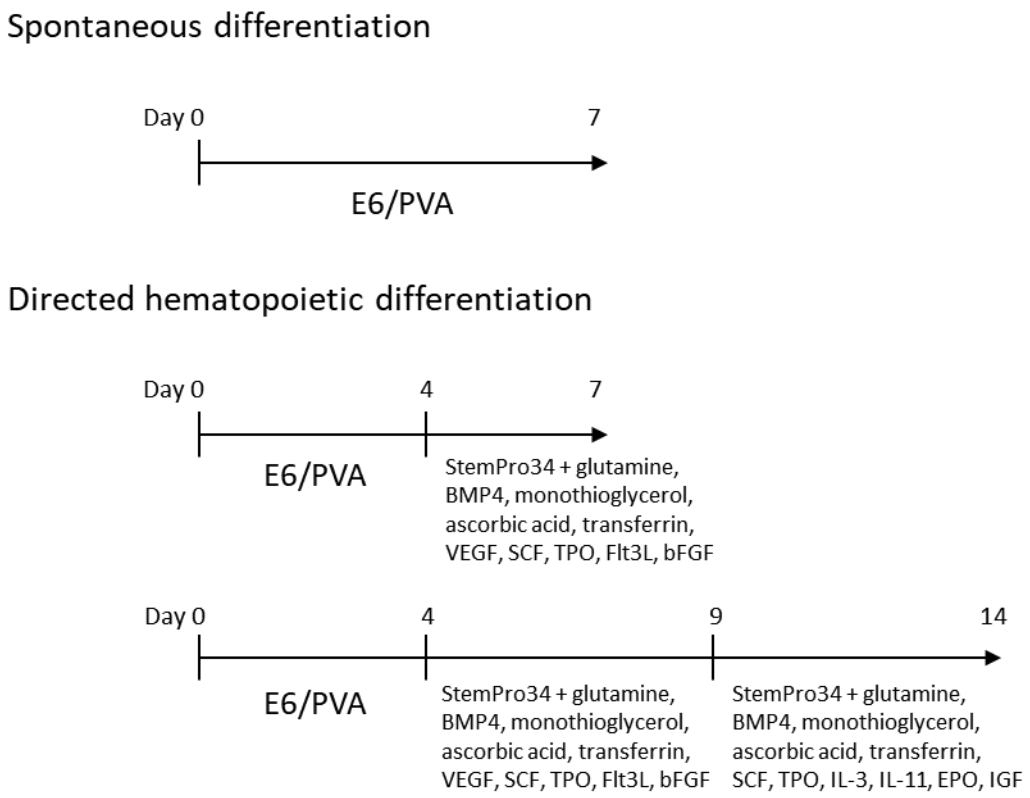


Figure 2.2 Schematic diagram showing the spontaneous and directed hematopoietic differentiation schedules of iPSCs.

Table 2.5 List of supplements used in iPSC differentiation.

Name	Company	Cat. No.
VEGF	PeproTech	100-20A
SCF	PeproTech	300-07
Flt3-L	PeproTech	300-19
IL-3	PeproTech	200-03
IL-11	PeproTech	200-11
EPO	PeproTech	100-64
IGF-1	PeproTech	100-11
BMP4	Prospec	CYT-361
TPO	Prospec	CYT-302
Transferrin	Prospec	PRO-315
Ascorbic acid	Sigma	A8960
1-thioglycerol	Sigma	M1753
b-FGF	PeproTech	100-18B
Polyvinyl alcohol (PVA)	Sigma	P8136

2.13 scATAC-seq and data analysis

The scATAC-seq libraries were prepared using the SureCell ATAC-Seq Library Prep Kit (Bio-Rad) and SureCell ATAC-Seq Index Kit (Bio-Rad) according to the manufacturer's instruction. The protocol started with cell lysis and tagmentation. Washed and pelleted cells and tagmentation buffers were pre-chilled and kept on ice. Cells were lysed with the Omni-ATAC lysis buffer (0.1% NP-40, 0.1% Tween-20, 0.01% digitonin, 10 mM NaCl, 3 mM $MgCl_2$ and 10 mM Tris-HCl pH 7.4) for 3 min on ice. The lysis buffer was diluted with ATAC-Tween buffer which contains 0.1% Tween-20. Lysed cells were collected and resuspended in Omni Tagmentation Mix which is formulated with ATAC Tagmentation Buffer and ATAC Tagmentation Enzyme and was buffered with 1× PBS + 0.1% BSA. The tagmentation mixture was agitated on a ThermoMixer at 37°C for 30 min. Then, tagmented nuclei were loaded onto a ddSEQ Single-Cell Isolator (Bio-Rad) and then bead barcoding and sample indexing were performed in a thermal cycler with the following PCR conditions: 37°C for 30 min, 85°C for 10 min, 72°C for 5 min, 98°C for 30 s, 8 cycles of 98°C for10 s, 55°C for 30 s and 72°C for 60 s, and a single 72°C extension for 5 min. Emulsions were disrupted and products were cleaned up using AMPure XP beads (Beckman Coulter). Barcoded amplicons were further amplified using with PCR conditions as follows: 98°C for 30 s, 9 cycles of 98°C for 10 s, 55°C for 30 s and 72°C for 60 s, and a single 72°C extension for 5 min to finish. PCR products were purified again using AMPure XP beads and quantified using Agilent Bioanalyzer (Agilent) with the High-Sensitivity DNA kit (Agilent). The sequencing of libraries was conducted in Illumina HiSeq platform.

The quality of ATAC-seq raw data was checked by FastQC (Andrews, 2010). TrimGalore was used to remove the adapters and low-quality reads. ATAC-seq reads

were aligned to the human reference genome (GRCh38) using BWA (Li, 2013). Peak calling was performed using MACS2 (Gaspar, 2018). Downstream analyses (clustering analysis, differential accessible peak analysis, gene score analysis) was performed by ArchR (Granja et al., 2021). The accessible peak-related genes were identified by BEDTools (Quinlan and Hall, 2010) and their The Gene Ontology terms were annotated by g:Profiler (Raudvere et al., 2019).

2.14 Statistical analysis

Statistical analysis was performed using GraphPad Prism version 9.0 (GraphPad Software, San Diego, California USA). All results were described as mean ± SD from at least triplicate data, unless otherwise stated. To evaluate differences between two groups, unpaired Student's t-test was performed. A p-value < 0.05 was considered as statistically significant.

Chapter 3 – Results

3.1 Identification of lncRNA involved in *JAK2*-V617F mutant cells

3.1.1 Differentially expressed lncRNA in HEL cells upon JAK2 inhibition

In order to identify lncRNA candidates involved in *JAK2*-V617F signaling, which is the most prevalent driver mutation of *BCR-ABL1*-negative MPNs, a total of 84 lncRNAs related to tumorigenesis were assessed by an RT-qPCR array to compare HEL cells with or without ruxolitinib treatment for 48 hours. (Figure 3.1.1A) The array data revealed six lncRNAs (BANCR, CBR3-AS1, LINC00261, LINC00887, LUCAT1, and NBR2) showing statistically significant differences in their expressions in response to JAK2 inhibition in HEL cells (Table 3.1.1).

Then, their expression changes were further validated by individual RT-qPCR with re-designed primers, and results showed the expression trend and levels are matched with RNA-seq data but LUCAT1 and NBR2 did not display a statistically significant difference (Figure 3.1.1B). Among the remaining four candidates, the lncRNA BANCR displayed the greatest expression difference after JAK2 inhibition in HEL cells, hence its involvement in *JAK2*-V617F signaling was further investigated.

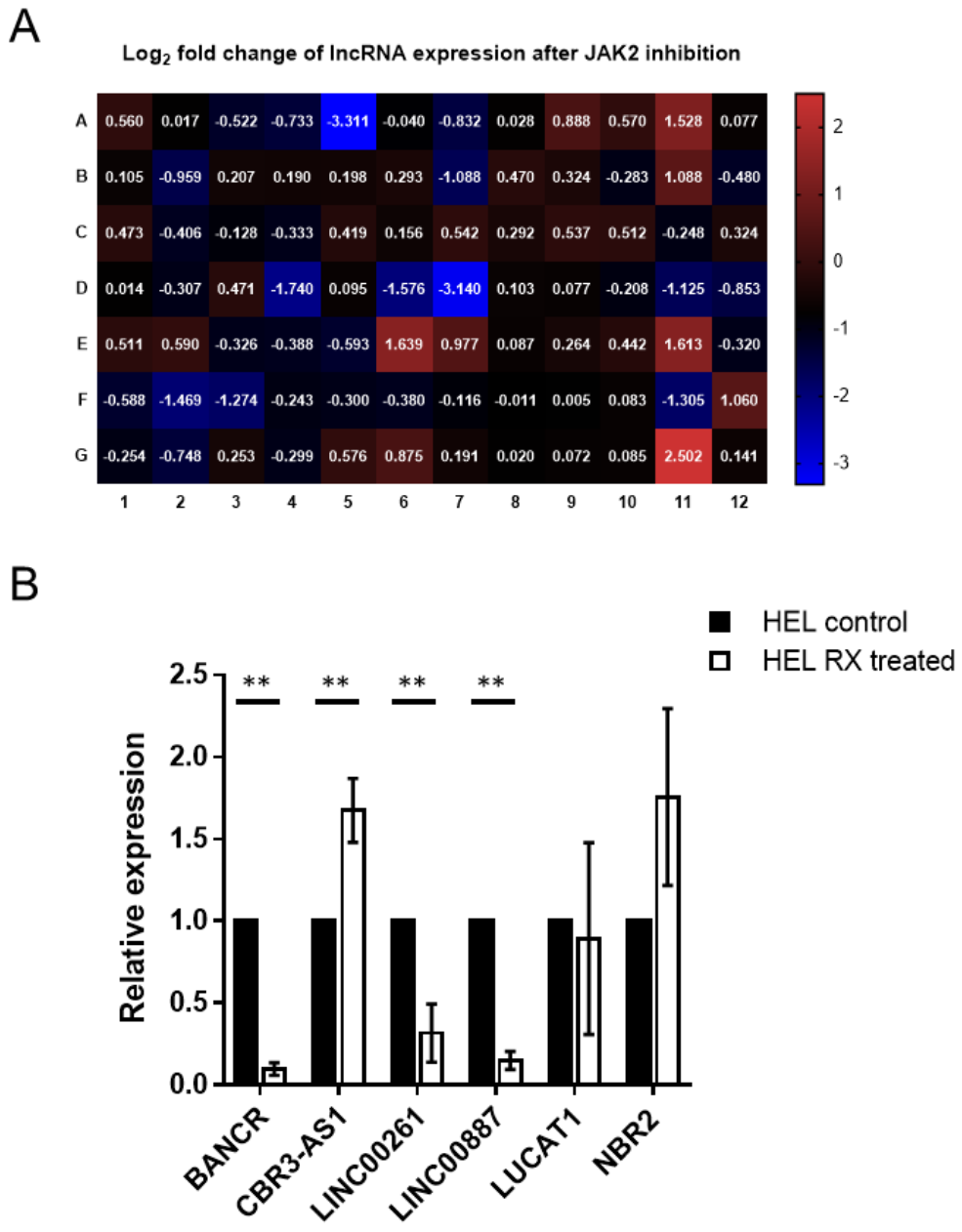


Figure 3.1.1 Differentially expressed lncRNA in HEL cells upon JAK2 inhibition.

(A) Real-time qPCR array assay revealed the expression change of 84 tumorigenesis-related lncRNAs in HEL upon 1 μM ruxolitinib treatment for 48 hours. (B) The expression of six differentially expressed lncRNAs selected from qPCR array results were further validated by individual RT-qPCR (n=4). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

Table 3.1.1 LncRNAs showing significant differential expression in HEL with ruxolitinib treatment.

Position*	Gene symbol	Fold change	Log_2 fold change	t-test *p*-value†	Description
A05	BANCR	0.101	-3.311	0.0051	BRAF-activated non-protein coding RNA
A09	CBR3-AS1	1.851	0.888	0.0043	CBR3 antisense RNA 1
D04	LINC00261	0.299	-1.740	0.0105	Long intergenic non-protein coding RNA 261
D07	LINC00887	0.113	-3.140	0.00046	Hypothetical LOC100131551
D12	LUCAT1	0.554	-0.853	0.0016	Lung cancer associated transcript 1 (non-protein coding)
E09	NBR2	1.201	0.264	0.0066	Neighbor of BRCA1 gene 2 (non-protein coding)

* This refers to the position shown in **Figure 3.1.1A**.

† The *p*-values were calculated using an unpaired Student's t-test on the replicate $2^{-\Delta CT}$ values for each gene in the treatment group compared to the control group.

3.1.2 Investigation of BANCR expression in *JAK2*-V617F signaling

RT-qPCR analysis revealed a dose-dependent expression change of BANCR over 0.1 µM, 0.5 µM and 1 µM ruxolitinib treatment (Figure 3.1.2A). The downregulation of BANCR after JAK2 inhibition was observed in both HEL and SET-2, which are *JAK2*-V617F^{+} cell lines, with 10.2-fold and 6.94-fold reduction, respectively (Figure 3.1.2B**)**. In K562 cell line, which is *BCR-ABL1*-positive but *JAK2* wild-type, the expression of BANCR did not show significant change after ruxolitinib treatment.

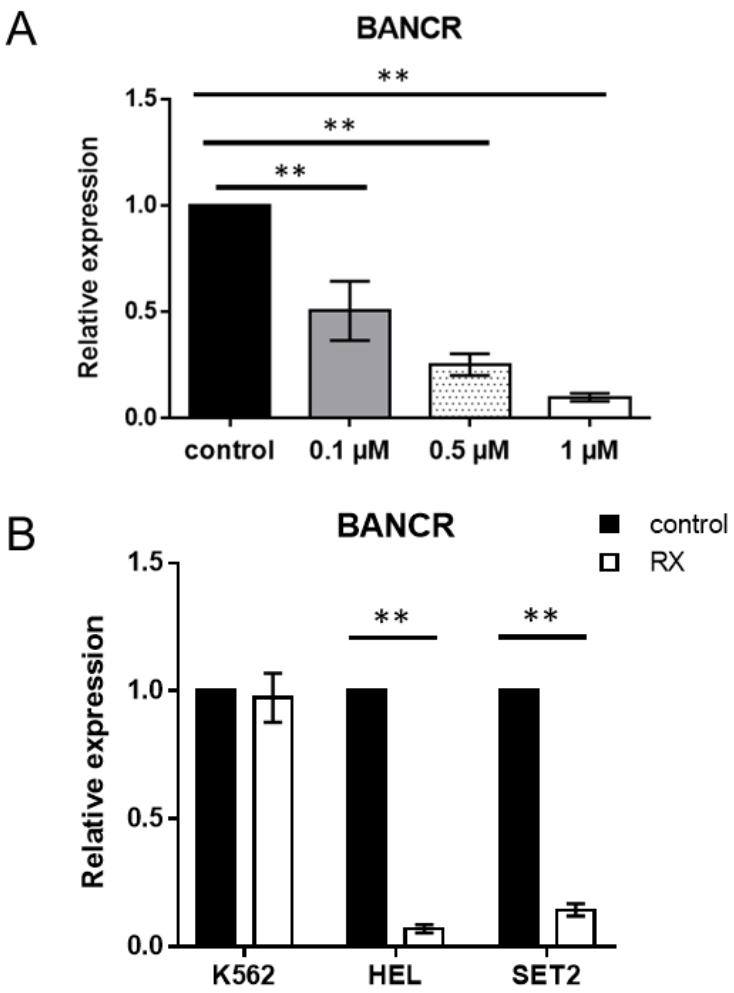


Figure 3.1.2 Investigation of BANCR expression in *JAK2*-V617F signaling.

(A) RT-qPCR analysis revealed the expression changes of BANCR in HEL cells upon different dosage of ruxolitinib treatment (n=4). (B) The expression changes of BANCR upon ruxolitinib treatment in K562, HEL and SET-2 were detected by RT-qPCR (n=3). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.1.3 Exogenous expression of *JAK2*-V617F in HEK293T cells

To confirm the expression of BANCR can be regulated by *JAK2*-V617F signaling, *JAK2*-V617F overexpression experiment was performed using HEK293T cells. After transfection of *JAK2*-V617F expression vector for 48 hours, total protein was extracted from the cells and western blotting showed an ectopic expression of JAK2 kinase and its downstream pathway represented by the phosphorylation of STAT3 and STAT5 (Figure 3.1.3A). Subsequent RT-qPCR detected an upregulation of BANCR by 4.18-fold in *JAK2*-V617F expressed HEK293T cells relative to the cells transfected with vector-control (Figure 3.1.3B). These results indicated the expression of BANCR could be modulated by *JAK2*-V617F signaling.

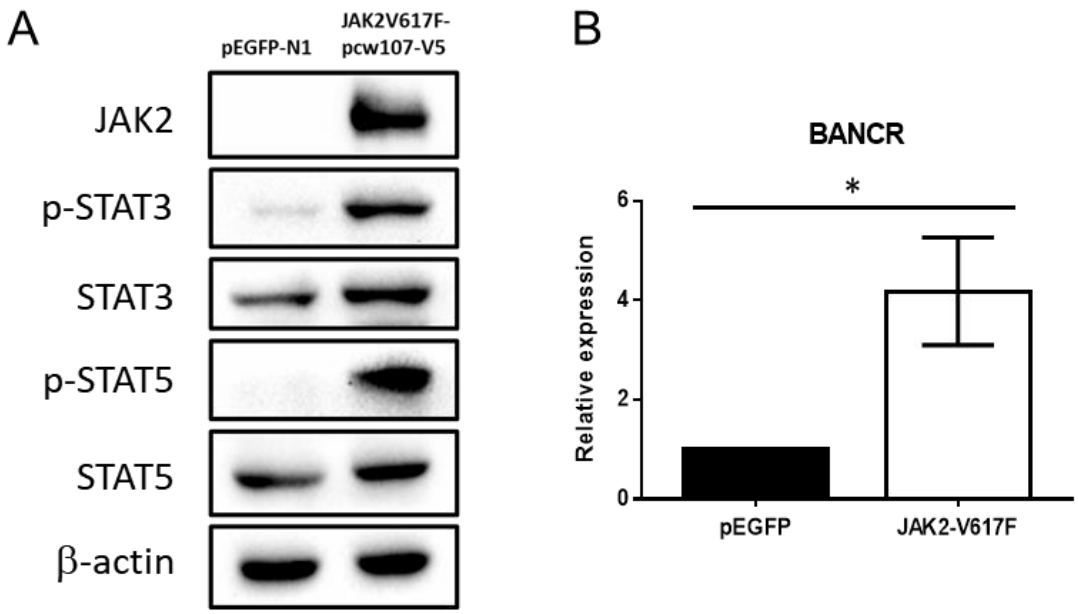


Figure 3.1.3 Exogenous expression of *JAK2*-V617F in HEK293T cells.

(A) Western blots revealed a strong activation of STAT3 and STAT5 phosphorylation in HEK293T cells transfected with *JAK2*-V617F overexpressing vector. (B) An upregulation of BANCR in HEK293T upon overexpression of *JAK2*-V617F was detected by RT-qPCR analysis (n=6). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.2 Identification of novel JAK2-associated lncRNA networks and putative ceRNA interaction

3.2.1 Gene differential expression pattern between HEL cells and ruxolitinib treated cells

To identify novel JAK2-associated lncRNAs and related networks, high throughput screening using RNA-seq was conducted with three batches of control and ruxolitinib-treated (48 hours) HEL cells. The sequencing data was corrected for batch effect from the replicates and differential gene expression analysis was performed using bioinformatics tool DEseq2 (Figure 2.1). The screening criteria was adjusted *p*-value < 0.05 and $\log_2$ fold change (log2FC) > 1 or < -1.

A total of 908 differentially expressed genes (DEGs) were identified, including novel transcripts (Figure 3.2.1A). A heatmap was constructed to visualize the 51 newly predicted lncRNAs, which showed significant differential expression in HEL cells upon JAK2 inhibition (Figure 3.2.1B). Among these 51 novel lncRNAs, 28 of them were upregulated and 23 of them were downregulated in ruxolitinib-treated HEL. The top 20 upregulated and downregulated lncRNAs including both annotated and unannotated ones are listed in (Table 3.2.1)

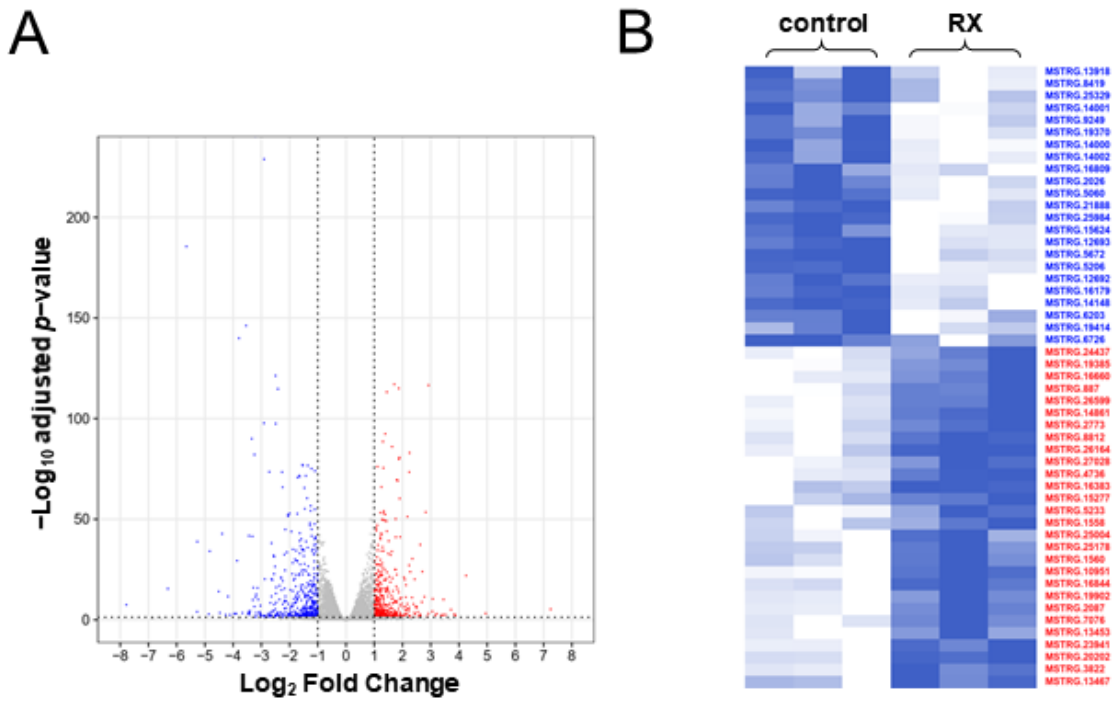


Figure 3.2.1 Gene differential expression pattern in HEL upon ruxolitinib treatment.

(A) The volcano plot showed differentially expressed genes in HEL upon ruxolitinib treatment with the selection criteria adjusted *p*-value<0.05 and $|\log_2 FC| \geq 1$. The red dots and blue dots represent upregulated and downregulated genes, respectively. (B) The heat map displayed novel lncRNAs showing differential expression ($|\log_2 FC| \geq 1$) in RNA-seq (n=3). LncRNAs labeled with blue and red color represent downregulation and upregulation, respectively.

Table 3.2.1 Top 20 up- and down-regulated lncRNA in HEL after ruxolitinib treatment.

Gene ID[†]	Gene	baseMean[††]	Log_2 Fold Change	lfcSE[†††]	*p*-value	p_{adj}
ENSG00000274370	AC144831.3	19.698287	2.914185	0.758311	0.000122	0.000693
ENSG00000244382	RP11-373I8.1	39.944185	2.474340	0.449706	3.75E-08	4.42E-07
ENSG00000240050	RP1-93H18.1	14.628234	2.257259	0.802638	0.004919	0.01781
ENSG00000223727	AC026188.1	13.691442	2.002816	0.711658	0.004888	0.017725
ENSG00000275527	CTD-3154N5.2	13.778523	1.996198	0.740527	0.007025	0.024195
ENSG00000278921	EPB41L4A-AS2	15.053160	1.929660	0.687477	0.005003	0.018077
ENSG00000261888	AC144831.1	25.308926	1.823062	0.590087	0.002005	0.008225
MSTRG.19902	MSTRG.19902	33.419314	1.790756	0.476815	0.000173	0.000951
MSTRG.4736	MSTRG.4736	2665.882866	1.757615	0.114254	2.12E-53	6.87E-51
ENSG00000266651	RP11-138I1.3	39.525033	1.727911	0.453195	0.000137	0.000774
MSTRG.25004	MSTRG.25004	338.792323	1.697930	0.265227	1.54E-10	2.66E-09
ENSG00000232470	RP11-313D6.3	49.670179	1.680342	0.429571	9.17E-05	0.000541
MSTRG.1558	MSTRG.1558	19.756380	1.629320	0.627209	0.009384	0.030864
MSTRG.20202	MSTRG.20202	894.430509	1.621538	0.127750	6.46E-37	9.16E-35
ENSG00000251434	RP11-315A17.1	27.901942	1.617880	0.592863	0.006354	0.022195
ENSG00000234690	AC073283.4	23.228258	1.609461	0.572161	0.004909	0.017779
ENSG00000231439	WASIR2	85.249281	1.593935	0.294263	6.07E-08	6.93E-07
ENSG00000258084	RP11-754N21.1	25.564300	1.559774	0.537576	0.003714	0.013999
ENSG00000232855	AF131217.1	18.738069	1.548156	0.629797	0.013964	0.043218
ENSG00000225173	XXbac-BPG308K3.5	28.007982	1.489438	0.539012	0.005722	0.020299
MSTRG.5206	MSTRG.5206	140.924638	-5.270468	0.388686	6.94E-42	1.34E-39
MSTRG.12693	MSTRG.12693	318.926452	-3.320190	0.237110	1.50E-44	3.42E-42
MSTRG.16179	MSTRG.16179	34.214206	-3.030161	0.522689	6.74E-09	8.93E-08
ENSG00000238062	SPATA3-AS1	11.359000	-2.713517	0.907172	0.002779	0.010914

MSTRG.5060	MSTRG.5060	45.454877	-2.680041	0.422096	2.16E-10	3.67E-09
MSTRG.14148	MSTRG.14148	153.166073	-2.641340	0.268349	7.35E-23	4.94E-21
MSTRG.25984	MSTRG.25984	242.337475	-2.514937	0.224062	3.10E-29	3.14E-27
MSTRG.14000	MSTRG.14000	173.362803	-2.484996	0.313933	2.46E-15	7.82E-14
MSTRG.14001	MSTRG.14001	23.901599	-2.435222	0.632455	0.000118	0.000676
MSTRG.14002	MSTRG.14002	26.118020	-2.431415	0.578305	2.62E-05	0.000176
ENSG00000274021	RP11-823E8.3	16.776498	-2.389231	0.719458	0.000897	0.004073
ENSG00000228340	MIR646HG	27.464887	-2.335150	0.564576	3.53E-05	0.000229
ENSG00000266378	RP11-214O1.3	11.772855	-2.253249	0.818848	0.005928	0.020936
ENSG00000266709	RP11-214O1.2	57.558749	-2.130532	0.402905	1.24E-07	1.34E-06
ENSG00000261685	RP11-401P9.4	24.502471	-2.047620	0.554804	0.000224	0.0012
MSTRG.13918	MSTRG.13918	24.637600	-2.046617	0.635149	0.001272	0.005505
MSTRG.9249	MSTRG.9249	13.452298	-1.986449	0.753081	0.008345	0.027992
MSTRG.2026	MSTRG.2026	25.312443	-1.984915	0.537325	0.000221	0.001186
MSTRG.6203	MSTRG.6203	90.275618	-1.984434	0.398111	6.21E-07	5.87E-06
MSTRG.21888	MSTRG.21888	30.217200	-1.983776	0.497453	6.67E-05	0.000407

† Gene ID starting with MSTRG are newly identified novel lncRNA from the RNA-seq data

†† the average of the normalized counts taken over all samples

††† standard error of the $\log_2$ fold change estimate

3.2.2 Gene ontology enrichment and gene co-expression analyses

The functional information of the DEGs was annotated by Metascape database. From pathway and process enrichment analysis, the top 20 clusters of the enriched biological pathways were identified (Figure 3.2.2A). The results revealed the DEGs may involve in biological functions or pathways including vascular functions (GO: 0001568), cytokine regulation (GO: 0001817) and receptor kinase signaling (GO: 0007178).

Then, gene co-expression network analysis was performed to disclose potential interrelationships between lncRNA and JAK/STAT signaling by correlating the expression patterns of them. A co-expression network was constructed with the differentially expressed lncRNAs including novel ones and JAK/STAT pathway genes. This demonstrated that a number of lncRNAs showed high correlation with certain genes in JAK/STAT signaling in terms of transcript expression, and hence potential interaction. Particularly, JAK2 showed the greatest number of co-expression relationship with all other JAK/STAT genes and lncRNAs in the network (Figure 3.2.2B).

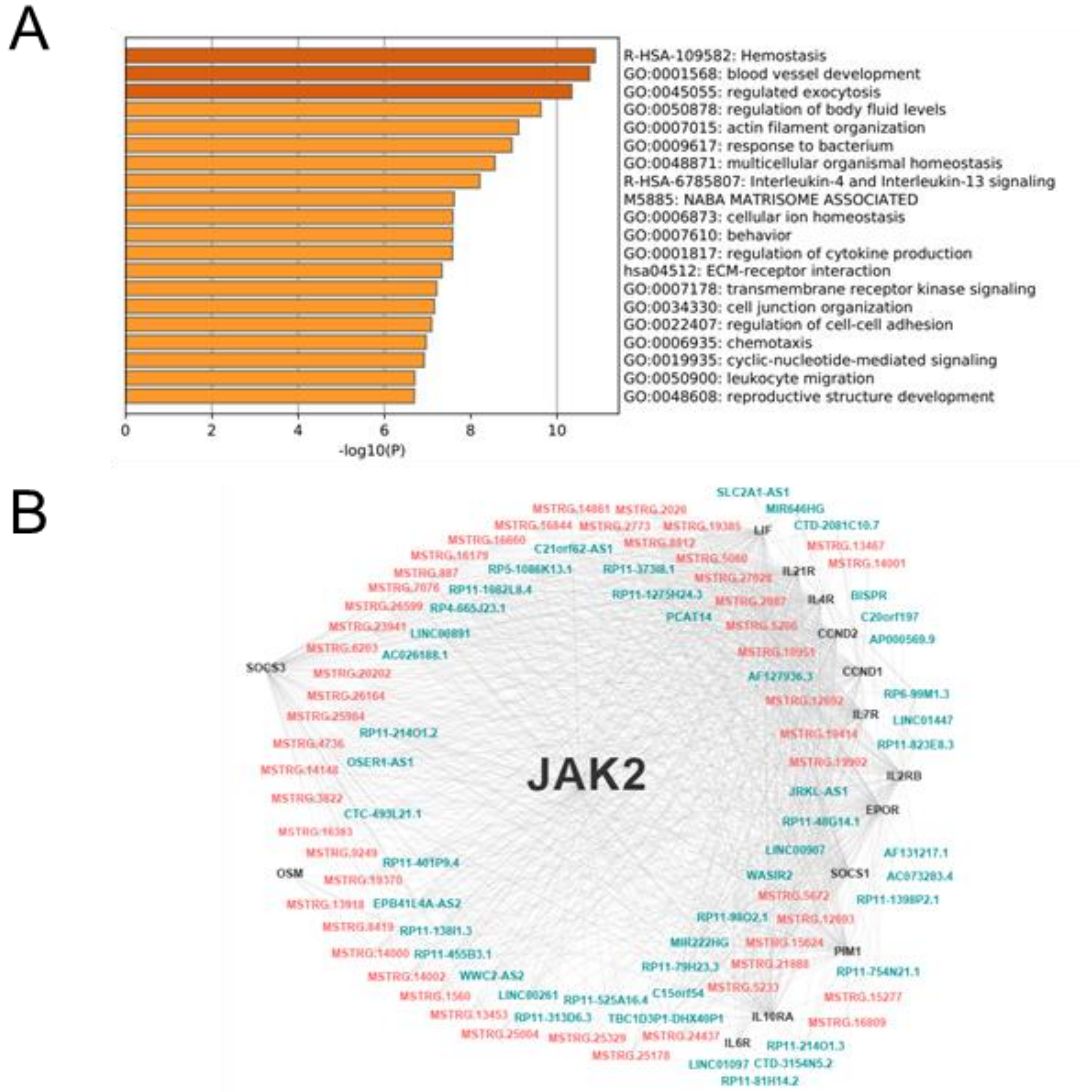


Figure 3.2.2 Gene ontology enrichment and gene co-expression analyses.

(A) The bar graph showed top 20 enriched GO terms determined by Metascape across DEGs and is colored according to *p*-values. (B) The gene co-expression network visualized the associations among different DEGs in RNA-seq, including JAK2-pathway genes (black), known lncRNAs (blue-green) and novel lncRNAs (orange-pink).

3.2.3 Putative ceRNA networks between JAK/STAT pathway and lncRNA

Many recent studies demonstrated lncRNA form regulatory networks with miRNA and mRNA by acting as competing endogenous RNAs (ceRNAs) and the dysregulation of such networks or axes may result in pathological conditions. In my study, I also aimed to explore any ceRNA networks present in the *JAK2*-V617F signaling and I started with the investigation of the lncRNA BANCR. The potential interaction between BANCR and targeted miRNAs associated with JAK2 signaling was predicted by STarMir (Rennie et al., 2014). These miRNA targets were selected from a previous research study conducted by Pagano et al., which studied the association of miRNA expression with *JAK2*-V617F activity by JAK2 inactivation in HEL cells (Pagano et al., 2018). Five miRNAs (miR-1244, miR-1246, miR-1248, miR-3609, and miR-3654) exhibit the greatest upregulated fold change with statistically significant difference were chosen for the prediction analysis, and results showed that BANCR contains putative miRNA binding region with perfectly complementary matched seed sequence (AGUGAAA) for miR-3609 only (Figure 3.2.3 and Table 3.2.3.1).

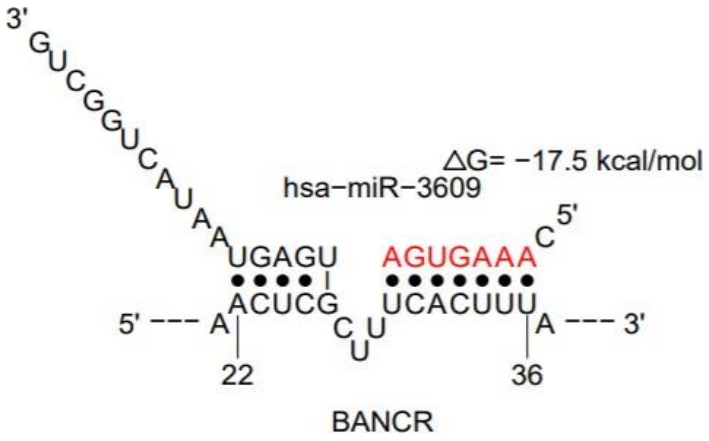


Figure 3.2.3 Putative MRE prediction analysis for BANCR.

The potential interaction between BANCR and miR-3609 was predicted by *in silico* tool STarMir and the result revealed a putative binding region matched to the seed sequence AAAGUGA on miR-3609.

Table 3.2.3.1 Potential binding regions of miR-3609 on BANCR and novel lncRNAs.

lncRNA	Transcript variant	Binding region	Seed sequence position	ΔG* (kcal/mol)
BANCR		22-36	30-36	-17.5
MSTRG.5060		1664-1688	1681-1687	-21.9
MSTRG.5206	Variant 1	1639-1679	1673-1678	-26.5
		90-107	102-107	-16.1
	Variant 3	1541-1581	1575-1580	-26.5
		90-107	102-107	-16.1
MSTRG.12693		1674-1702	1696-1702	-17.2
		292-303	297-302	-14.1
MSTRG.13918		8-28	22-27	-22.1
MSTRG.14000	Variant 1	3-43	37-42	-26.0
	Variant 2	3-43	37-42	-26.0
	Variant 3	3-43	37-42	-26.0
	Variant 4	3-43	37-42	-26.0
MSTRG.14001		1-35	29-34	-23.8
MSTRG.14002		1-35	29-34	-23.8
MSTRG.14148		1443-1483	1477-1483	-16.3
		771-795	789-794	-18.0
		2488-2508	2502-2507	-25.3

*ΔG = A measure of stability for miRNA:target hybrid as computed by RNAhybrid. The more negative the value, the more stable the binding.

Similar analysis showed that miR-3609 can potentially interact with some genes in JAK/STAT pathway as they also possess the same putative binding sites for miR-3609 (Table 3.2.3.2), hence these genes and BANCR could act as ceRNAs in the same network and influence each other's expression by competing for miR-3609. Furthermore, prediction analysis was also performed for the top 20 downregulated novel lncRNAs from the RNA-seq data, and results revealed 8 of them contain potential binding sites for miR-3609 (Table 3.2.3.1). Among them, MSTRG.5206 showed the greatest downregulation in expression with a $\log_2$FC of -5.27 (Table 3.2.1).

Table 3.2.3.2 Potential binding regions of miR-3609 on JAK/STAT-related genes.

Genes	Transcript variant	Binding region	Seed sequence position	ΔG* (kcal/mol)
IL6ST	variant 1	728-762	756-762	-14.5
		4792-4809	4804-4809	-14.8
		4020-4046	4041-4046	-15.4
		2090-2106	2100-2106	-16.7
		4603-4651	4646-4651	-14
IL12A	variant 1	1199-1220	1214-1220	-21.8
CNTF		1356-1376	1370-1376	-21.3
		1695-1711	1705-1711	-20.7
CCND1		1592-1634	1628-1634	-21.5
		2008-2066	2061-2066	-13.2
SPRY3	variant 1	3430-3452	3446-3452	-16.7
		8065-8090	8084-8090	-15.3
STAM2		4454-4474	4468-4474	-14.5
		2200-2220	2215-2220	-19.3
		2289-2315	2309-2315	-20
		4815-4855	4849-4854	-27
SOCS4	variant 1	1842-1854	1848-1853	-13.8
		1144-1168	1162-1167	-15
		6528-6565	6559-6565	-15.6
		6773-6794	6789-6794	-14.1
JAK2	variant 1	2115-2134	2129-2134	-18.4
		6269-6296	6290-6296	-21.6
		5105-5121	5116-5121	-20.1
		6041-6072	6066-6072	-17.8
		2679-2696	2690-2695	-15.1
PIAS1	variant 1	3545-3570	3565-3570	-17.3
LIFR	variant 1	1419-1444	1439-1444	-19.2
		10006-10031	10026-10031	-13.6

*ΔG = A measure of stability for miRNA:target hybrid as computed by RNAhybrid. The more negative the value, the more stable the binding.

3.2.4 Investigation of CNVs in MPNs and their potential correlation with lncRNAs

Besides studying the expression changes that directly reflects the potential significance of newly identified lncRNAs, their locations in the genome were also examined and mapped with copy number variation (CNV) regions identified in MPN patients using our own whole-genome sequencing data. In total, the genomic loci of 7 differentially expressed novel lncRNAs from the RNA-seq data were found to be located within the regions of MPN-related CNVs (Table 3.2.4). Among these, it is noteworthy that MSTRG.1558 is located within the genomic region chr1: 124804374–124817081 that possesses a reduced copy number of MPN-related CNV, and it is one of the top 20 upregulated novel lncRNAs in HEL upon ruxolitinib treatment (Table 3.2.1). This finding suggests that low copy number or expression of MSTRG.1558 may be associated with MPN progression and vice versa. Contrarily, MSTRG.2026 is one of the top 20 downregulated novel lncRNAs (Table 3.2.1) and is located within the genomic region chr1: 161450633–161458545, which contains an MPN-related CNV with increased copies (Table 3.2.4). This finding showed a positive correlation between copy number or expression of MSTRG.2026 and MPNs.

Table 3.2.4 Location of MPN-related CNV and differentially expressed lncRNAs.

Genomic coordinate	lncRNA	lncRNA type	MPN sample	Copy number
chrX:45745211-45770274	MIR222HG	Known	MPN012	1
chrX:45745211-45770274	MIR222HG	Known	MPN027	1
chrX:45745211-45770274	MIR222HG	Known	MPN126	1
chrX:45764772-45765299	RP6-99M1.3	Known	MPN012	1
chrX:45764772-45765299	RP6-99M1.3	Known	MPN027	1
chrX:45764772-45765299	RP6-99M1.3	Known	MPN126	1
chrX:71697196-71706455	LINC00891	Known	MPN012	0
chrX:71697196-71706455	LINC00891	Known	MPN027	1
chrX:71697196-71706455	LINC00891	Known	MPN126	1
chrX:71771506-71784726	MSTRG.26599	Novel	MPN012	0
chrX:71771506-71784726	MSTRG.26599	Novel	MPN027	1
chrX:71771506-71784726	MSTRG.26599	Novel	MPN126	1
chr1:124804374-124817081	MSTRG.1558	Novel	MPN014	1
chr1:124804374-124817081	MSTRG.1558	Novel	MPN126	1
chr1:124804374-124817081	MSTRG.1558	Novel	MPN144	1
chr1:124908709-124912327	MSTRG.1560	Novel	MPN014	1
chr1:124908709-124912327	MSTRG.1560	Novel	MPN126	1
chr1:124908709-124912327	MSTRG.1560	Novel	MPN144	1
chr1:161450633-161458545	MSTRG.2026	Novel	MPN118	4
chr1:167863281-167890957	MSTRG.2087	Novel	MPN118	4
chr13:50108816-50116015	MSTRG.7076	Novel	MPN126	1
chr20:34098756-34101495	MSTRG.16383	Novel	MPN125	1
chr20:44210960-44226027	OSER1-AS1	Known	MPN125	1

3.3 Investigation of ncRNA pathway in *BCR-ABL1*-positive CML cells

3.3.1 Expression of H19 and miR-675 in K562 cells upon imatinib treatment

RT-qPCR analysis was performed to detect the expression change of H19 and its derived miR-675 in K562 cells after imatinib treatment. Results revealed a significant downregulation of H19 expression upon 10 μM of imatinib treatment in K562. Both miR-675-5p and miR-675-3p showed a trend of reduced expression responding to imatinib treatment but the reduction of miR-675-3p was not statistically significant (Figure 3.3.1). As miR-675-3p is not that responsive, miR-675-5p was focused in subsequent investigation of imatinib resistance in K562.

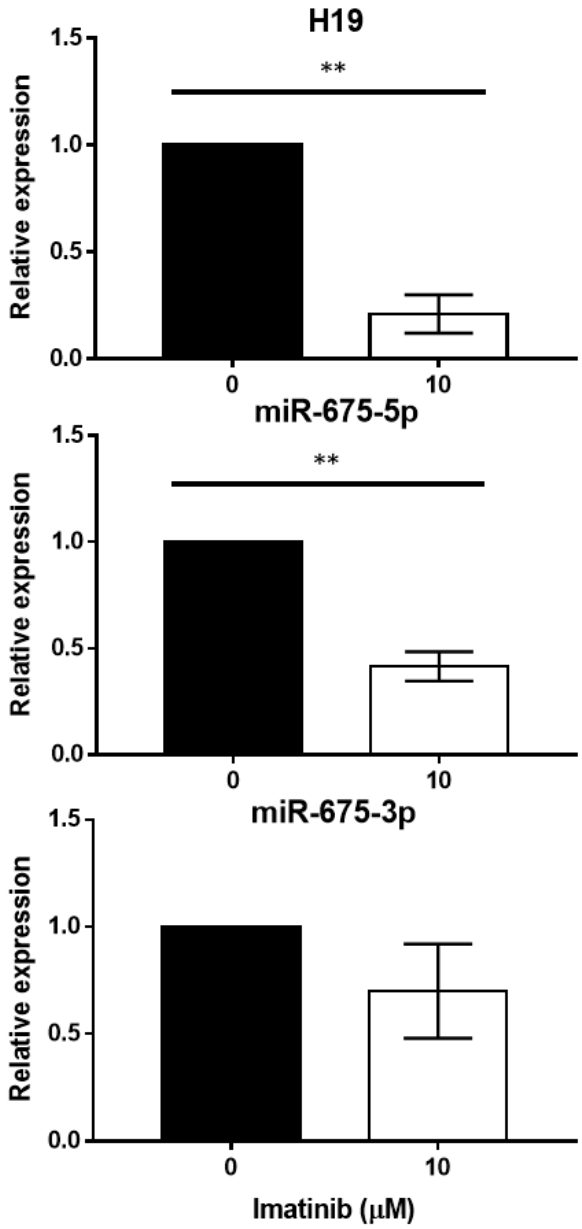


Figure 3.3.1 H19/miR-675 showed downregulation in K562-IMS upon imatinib treatment.

RT-qPCR analysis of H19, miR-675-5p and miR-675-3p expression in K562 after 10 μM imatinib treatment for 72 hours (n=3). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.3.2 Cell viability and MDR1 expression of IMR CML cells

To investigate the involvement of H19/miR-675 activity in CML drug resistance, imatinib-resistant (IMR) K562 cells (K562-IMR) were generated by drug selection under long-term imatinib treatment. Cell viability testing by trypan blue exclusion showed >88.1% viability of K562-IMR upon 10 μM imatinib treatment for 72 hours, while imatinib-sensitive K562 cells (K562-IMS) had a viability of <28.7% (Figure 3.3.2A). In addition, the expression level of MDR-1, a common marker for drug resistance, was also assessed. RT-qPCR results revealed a significant increase of MDR-1 expression by 35.6-fold, suggesting a therapeutic resistance property was gained by K562-IMR cells (Figure 3.3.2B).

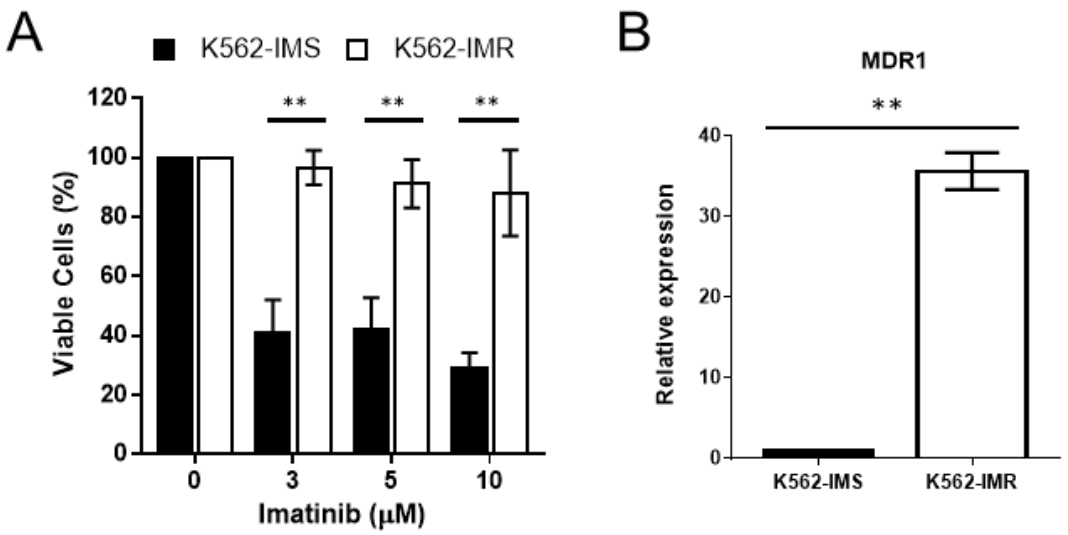


Figure 3.3.2 Cell viability and MDR1 expression of IMR CML cells.

(A) Cell viability of K562-IMS and K562-IMR under different dosage of imatinib treatment was examined by trypan blue exclusion test. (B) A significantly increased expression of MDR1 in K562-IMR compared to K562-IMS was revealed by RT-qPCR assays (n=6). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.3.3 Expression of H19 and miR-675 in IMR CML cells

To investigate the involvement of H19 in imatinib resistance, the expression patterns of H19 in K562-IMR and K562-IMS was assessed by RT-qPCR assays. A significant upregulation of H19 level by 9.2-fold was detected in K562-IMR relative to K562-IMS (Figure 3.3.3A upper panel). This investigation was also performed on another CML cell line LAMA84 and its imatinib-resistant counterpart. RT-qPCR results revealed the same trend of H19 expression in LAMA84-IMR cells with an upregulation by 15.1-fold (Figure 3.3.3A lower panel). Then, the expression of H19-derived miR-675-5p was also examined in these two CML cell lines. In accordance with H19 expression, the expression of miR-675-5p was increased in both K562-IMR (2.3-fold) and LAMA84-IMR (6.1-fold) compared to their IMS counterparts (Figure 3.3.3B).

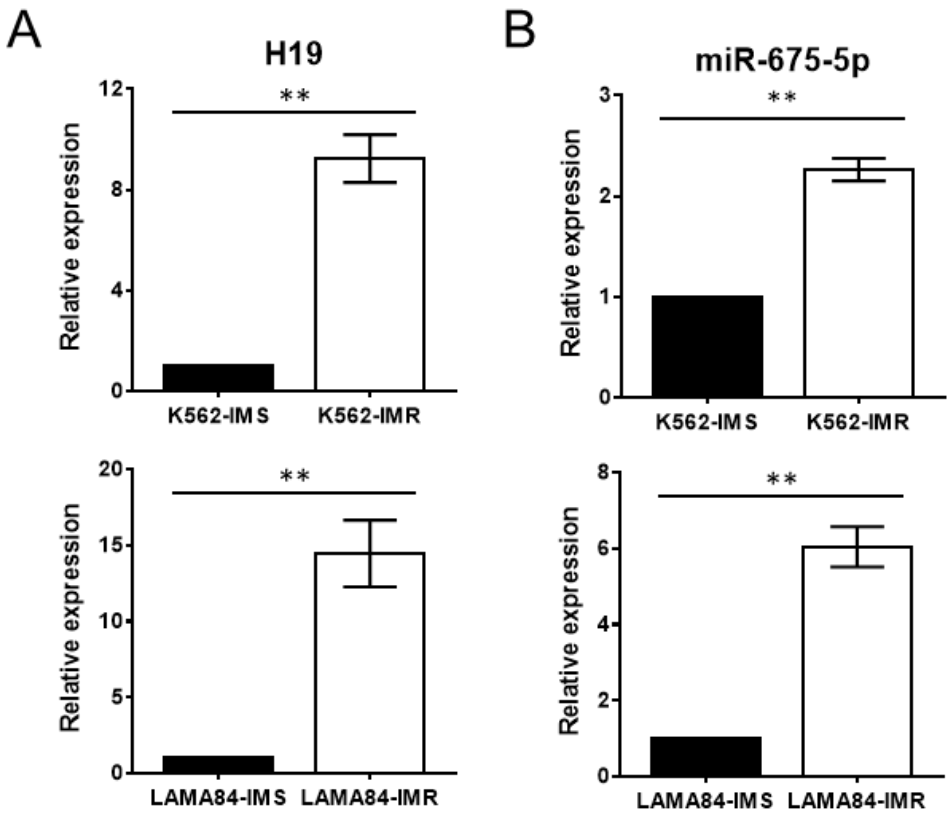


Figure 3.3.3 Expression of H19 and miR-675-5p in IMR CML cells.

(A) RT-qPCR revealed the relative expression changes of H19 in K562-IMR and LAMA84-IMR (n=4). (B) RT-qPCR revealed the relative expressions of miR-675-5p in K562-IMR and LAMA84-IMR (n=4). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.4 Identification of differentially expressed novel lncRNAs in IMR CML cells

3.4.1 Gene differential expression pattern and functional annotation of DEGs

In order to identify novel regulators that may involve in the regulation of imatinib resistance, poly(A)-enriched RNA-seq was conducted with K562-IMR and K562-IMS cells to target differentially expressed lncRNAs, including novel transcripts. Gene differential expression analysis was performed using DEseq2 with the following screening criteria: adjusted *p*-value < 0.05 and $\log_2$ fold change > 1 or < -1. A total of 144 lncRNAs were identified as significantly differentially expressed in RNA-seq, among which 122 were known lncRNAs, and 22 were unannotated novel lncRNAs. To visualize the lncRNA expression pattern, hierarchical clustering analysis based on FPKM (fragments per kilobase of transcript per million mapped read) was performed, and $\log_{10}$(FPKM+1) was used for the clustering (Figure 3.4.1A). In the heatmap, only lncRNAs with statistically significant difference between two groups are shown, and red represents genes with higher expression while blue represents those with reduced expression.

Gene Ontology and KEGG pathway analyses were performed on all the differentially expressed genes to determine the relevant biological function and pathway enrichment. Highly enriched GO terms included cytokine-mediated signaling pathway (GO: 0019221), cell morphogenesis (GO: 0000904) and the tyrosine kinase signaling (GO: 0007169) (Figure 3.4.1B).

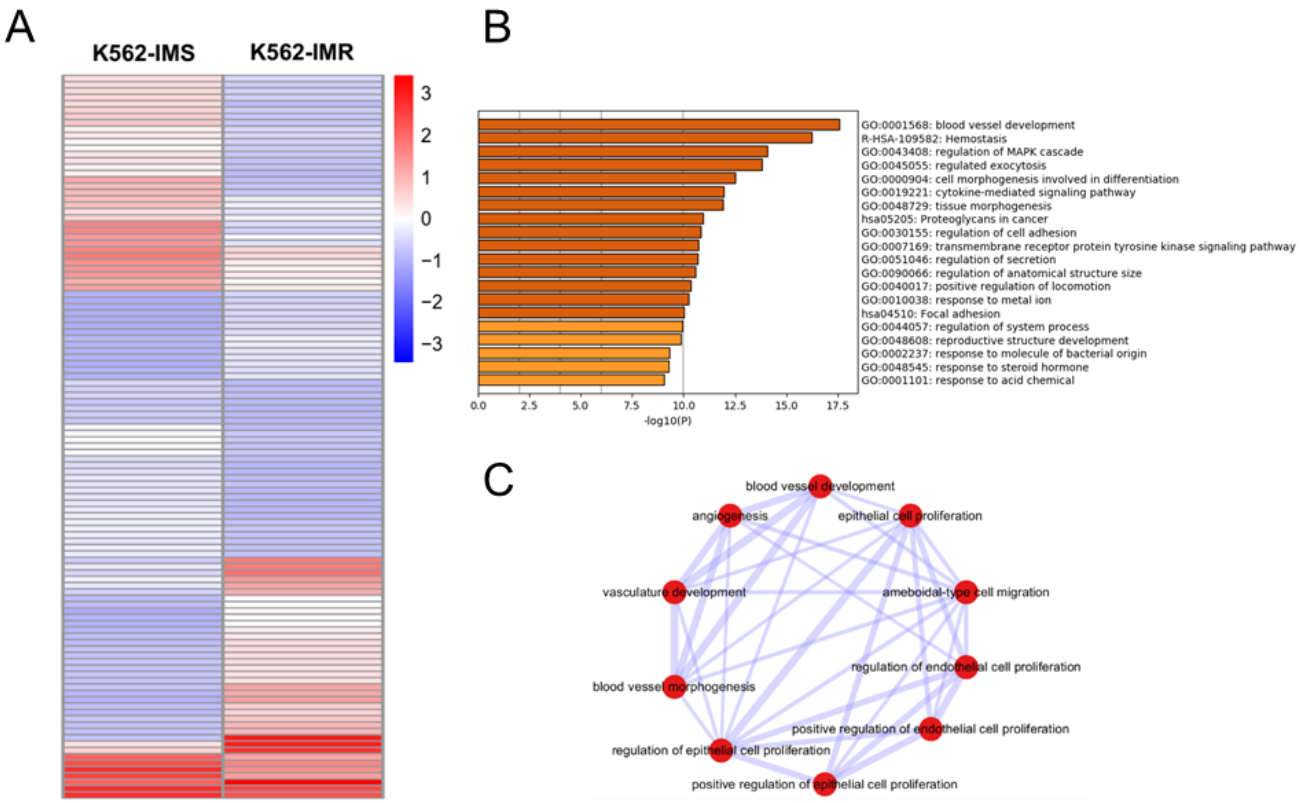


Figure 3.4.1 Differential gene expression pattern and functional annotation of DEGs in K562-IMR.

(A) The heat map showed 144 differentially expressed lncRNAs identified in RNA-seq. (B) The bar graph showed top enriched GO terms determined by Metascape across differentially expressed genes and is colored according to *p*-values. (C) Gene sharing network of sub-terms under the top enriched blood vessel development term. The thickness of lines positively correlated to amount of sharing genes.

3.4.2 Identification of novel lncRNA with significantly differential expression in K562-IMR

From the RNA-seq data analysis, differentially expressed genes including lncRNAs were obtained by comparison between K562-IMR and K562-IMS cells (Figure 3.4.2A). After correction for batch effects among biological replicates, the $\log_2$ fold changes ($\log_2$FC) were calculated. Among the unannotated novel lncRNAs, the most significantly downregulated lncRNA in K562-IMR named LNC000093 ($\log_2$FC = -7.77155, adjusted *p*-value = 4.31E-13) was chosen as the target for subsequent investigation. Then, RT-qPCR was individually performed to validate the expression change, and results confirmed a significant downregulation of LNC000093 in K562-IMR relative to K562-IMS (Figure 3.4.2B), which matched with the data from RNA-seq analysis.

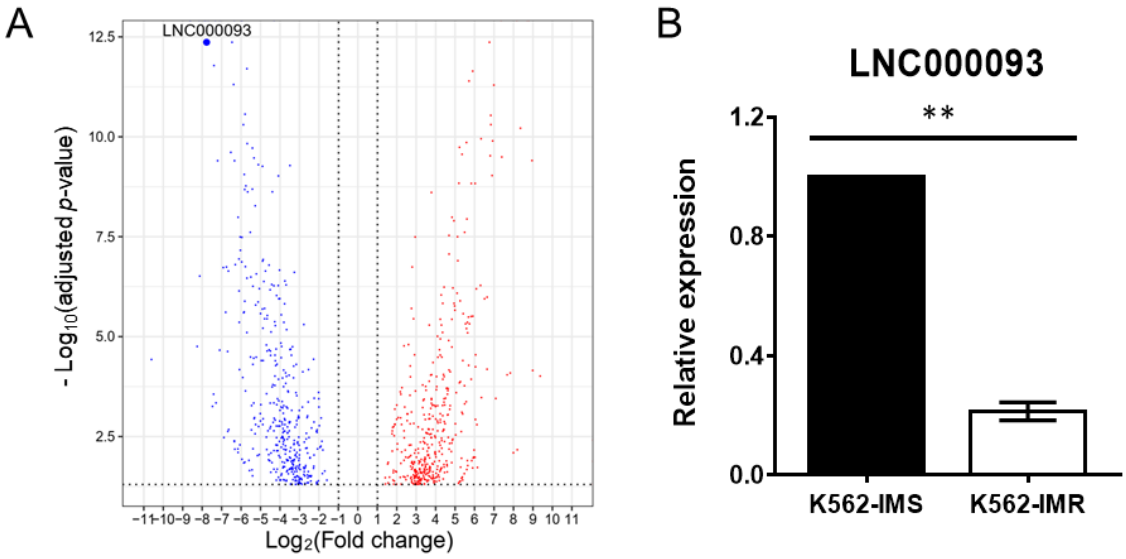


Figure 3.4.2 Novel lncRNA LNC000093 was identified and found to be significantly downregulated in K562-IMR.

(A) The volcano plot showed all DEGs of K562-IMR compared to K562-IMS (adjusted *p*-value<0.05 and |fold change|≥2; the threshold is represented by dotted lines). The red dots and blue dots represent upregulated and downregulated genes, respectively. The most significantly downregulated novel lncRNA LNC000093 was highlighted. (B) RT-qPCR analysis validated the significant downregulation of LNC000093 in K562-IMR cells relative to K562-IMS cells (n=3). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.4.3 Validation of LNC000093 sequence by cycle sequencing

From the alignment analysis of RNA-seq results, LNC000093 is a novel transcript with 1418 bp in length and is located on chromosome 14q23.1 with two exons, where exon 1 located at 59078887 to 59080020 and exon 2 at 59081603 to 59081888 (GRCh38.p12) (Figure 3.4.3A). The whole sequence of LNC000093 transcript was confirmed by cycle sequencing. The sequence analysis was performed in two CML cell lines (K562 and LAMA84), human iPSCs and human peripheral blood cells. Total RNA of the above-mentioned cells was extracted and converted into cDNA by reverse transcription. The first-strand cDNA was then subjected to cycle sequencing to read through the sequence of LNC000093.

The sequencing results showed that in addition to the known reported SNPs according to dbSNP from NCBI, a new polymorphism with six continuous nucleotides was identified in exon 2 of LNC000093 (Figures 3.4.3B, C). The sequencing analysis was also performed using gDNA as input template to examine this newly found polymorphism at genome level. Surprisingly, sequencing results showed that the two different alleles were observed only when using cDNA as input, but not gDNA (Fig 3.4.3B). In order to distinguish the two variants, the whole length of LNC000093 was amplified by PCR and cloned into a vector. After selecting single clones for bacterial culture, plasmids were isolated and subjected to cycle sequencing analysis. Sequencing results showed two distinct variants TGCACC and GTACAA (Figure 3.4.3C).

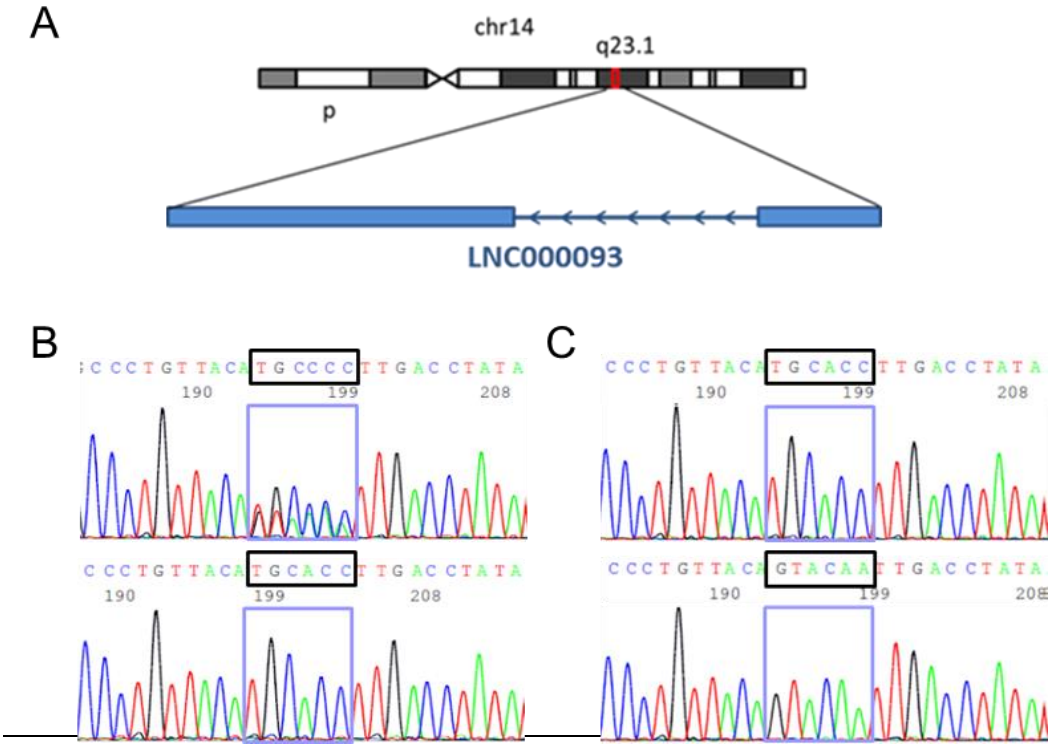


Figure 3.4.3 Validation of LNC000093 sequence.

(A) LNC000093 is located on chromosome 14q23.1 and is transcribed in anti-sense direction with two exons. (B) Cycle sequencing analysis using cDNA as input template demonstrated two different polymorphism variants with 6 continuous nucleotides (overlapping peaks) were existed in exon 2 of LNC000093 (upper panel). However, only one variant was shown in sequencing results when gDNA was used as template (lower panel). (C) After cloning PCR amplicons from cDNA template into empty vectors, two distinct variants (TGCACC/GTACAA) were isolated and examined by cycle sequencing.

3.4.4 Subcellular localization of LNC000093

As subcellular localization of lncRNAs could provide crucial clues for their molecular functions, I also investigate such information for LNC000093. Cytoplasmic and nuclear fractionation was performed with K562 cells and RNA was extracted from each fraction for RT-ddPCR analysis. Results of ddPCR assay showed that LNC000093 was detected in both the cytoplasm and nucleus but was majorly expressed in cytoplasmic fraction with 80.75% relative abundance (Figure 3.4.4). Two well-known nucleus-enriched lncRNAs NEAT1 and MALAT1 were also examined in order to ensure the nuclear fractions were correctly extracted, and ddPCR results revealed their enrichment in nucleus with 76.94% and 86.74% respectively. The results concluded that LNC000093 is mainly expressed in cytoplasm, which implied the primary functional role of LNC000093 should take place in cytoplasm such as regulating mRNA stability, translational regulation, and serving as miRNA sponge.

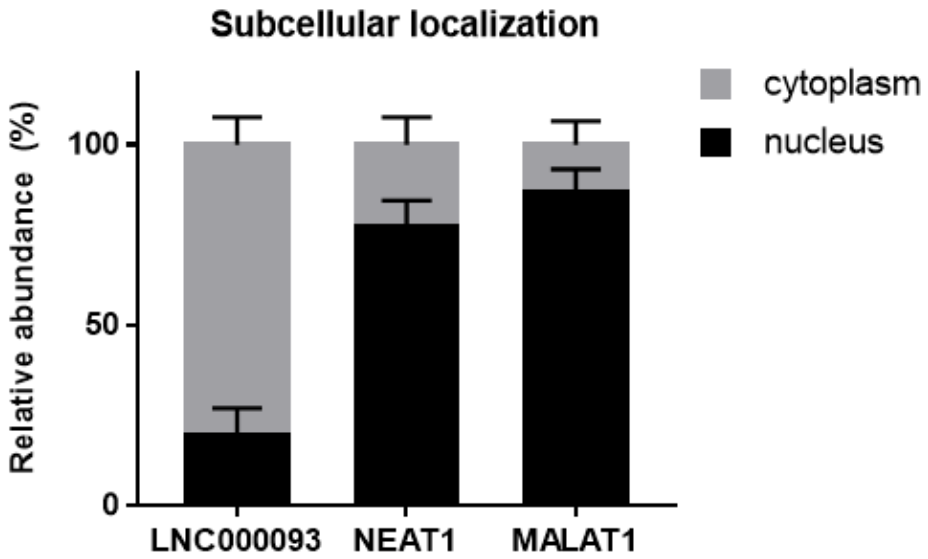


Figure 3.4.4 Subcellular localization of LNC000093.

RT-ddPCR analysis revealed the expression of LNC000093 in K562 is majorly enriched in cytoplasm (n=4). Two nuclear lncRNAs NEAT1 and MALAT1 were served as positive control to validate the nuclear fraction (n=4). Data are displayed as mean ± SD.

3.4.5 Expression of LNC000093 in TKI-sensitive CML patients

As LNC000093 was identified to be downregulated in IMR CML cells, its expression in TKI-sensitive CML patients was also investigated. Total RNA was extracted from the whole blood of ten CML patients or healthy donors, and was used for RT-qPCR to detect and compare the LNC000093 expression. RT-qPCR results revealed a higher LNC000093 expression in TKI-sensitive CML patients compared to the healthy controls, displaying a 3.05-fold increase in average (Figure 3.4.5).

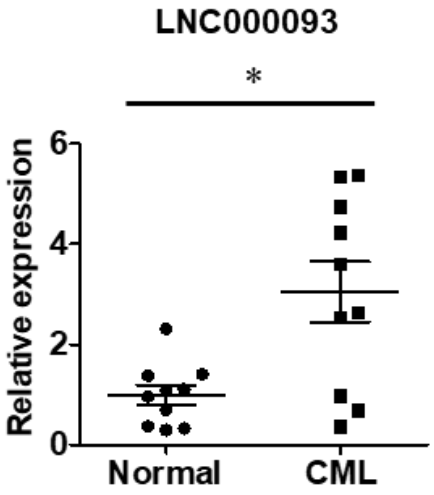


Figure 3.4.5 Expression of LNC000093 in TKI-sensitive CML patient.

RT-qPCR analysis revealed an upregulation of LNC000093 level in white cells from TKI-sensitive CML patients (n=10) compared to healthy individuals (n=10). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.5 Examination of interaction between LNC000093 and H19/miR-675

3.5.1 Prediction of miR-675 binding sites on LNC000093

The prediction of potential binding regions for miRNA on RNA transcript was performed by the prediction tool STarMir and results revealed that LNC000093 contained three putative miRNA response elements (MREs) for miR-675-5p, which is the most among the top 20 significantly downregulated novel lncRNAs in IMR cells. The three putative binding sites were located close to the 3' end of LNC000093 transcripts, designated MRE1 [971-1,008], MRE2 [1,237-1,268], and MRE3 [1,302-1,322], with complementary pairing to the seed sequence (GGUGCG) of miR-675-5p (Figure 3.5.1).

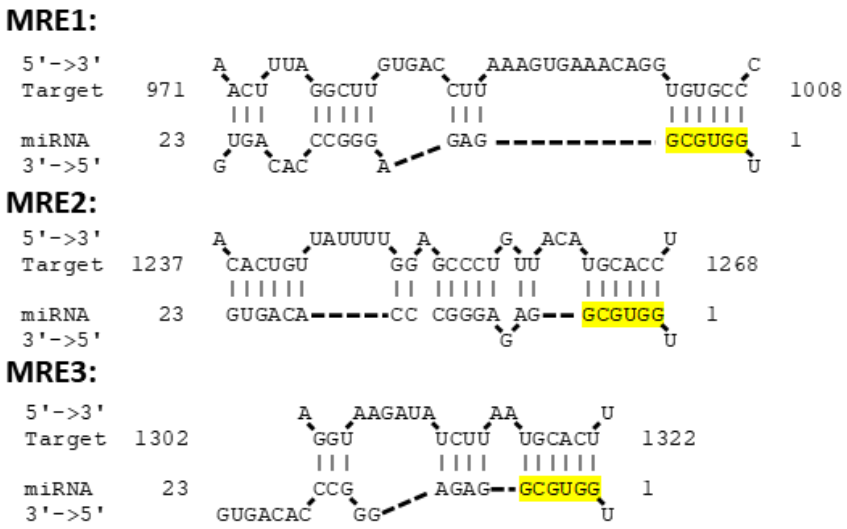


Figure 3.5.1 Prediction of miR-675 binding sites on LNC000093.

The potential interaction between LNC000093 and miR-675-5p was predicted by *in silico* tool STarMir and results revealed three putative binding regions matched to the seed sequence GGUGCG on miR-675-5p.

3.5.2 Expression of LNC000093 upon H19/miR-675 perturbation

With the use of H19-small interfering RNA (H19-siRNA) and CRISPR-Cas9-mediated deletion, loss-of-function studies were conducted with K562 cells to examine the regulatory effects of H19 on LNC000093. Following knockdown of H19 by siRNA, RT-qPCR analysis showed a downregulation of H19 expression with 0.13-fold relative to control. Moreover, an increased expression of LNC000093 by 2.6-fold was detected in H19-siRNA transfected K562 cells (Figure 3.5.2A).

To further investigate whether the regulatory effect is ascribed to miR-675-3p or miR-675-5p derived from H19, specific deletion of miR-675 was achieved by CRISPR-Cas9-mediated cleavage on exon 1 of H19 in K562 cells (Figure 3.5.2B). Subsequent RT-qPCR analysis confirmed both miR-675-3p and miR-675-5p was downregulated after CRISPR-deletion (Figure 3.5.2C). Then, transfection of miR-675 mimics was performed to investigate the effect on LNC000093 in K562 cells with or without H19/miR-675-deletion. H19/miR-675-deleted K562-IMR cells showed an increased LNC000093 expression by 4.1-fold relative to control K562-IMR, and such upregulation was reversed by the ectopic expression of miR-675-5p mimics, but not miR-675-3p (Figure 3.5.2D). In addition, transfection of miR-675-5p mimics into K562-IMS cells resulted in a downregulation of LNC000093 by 2.1-fold (Figure 3.5.2D).

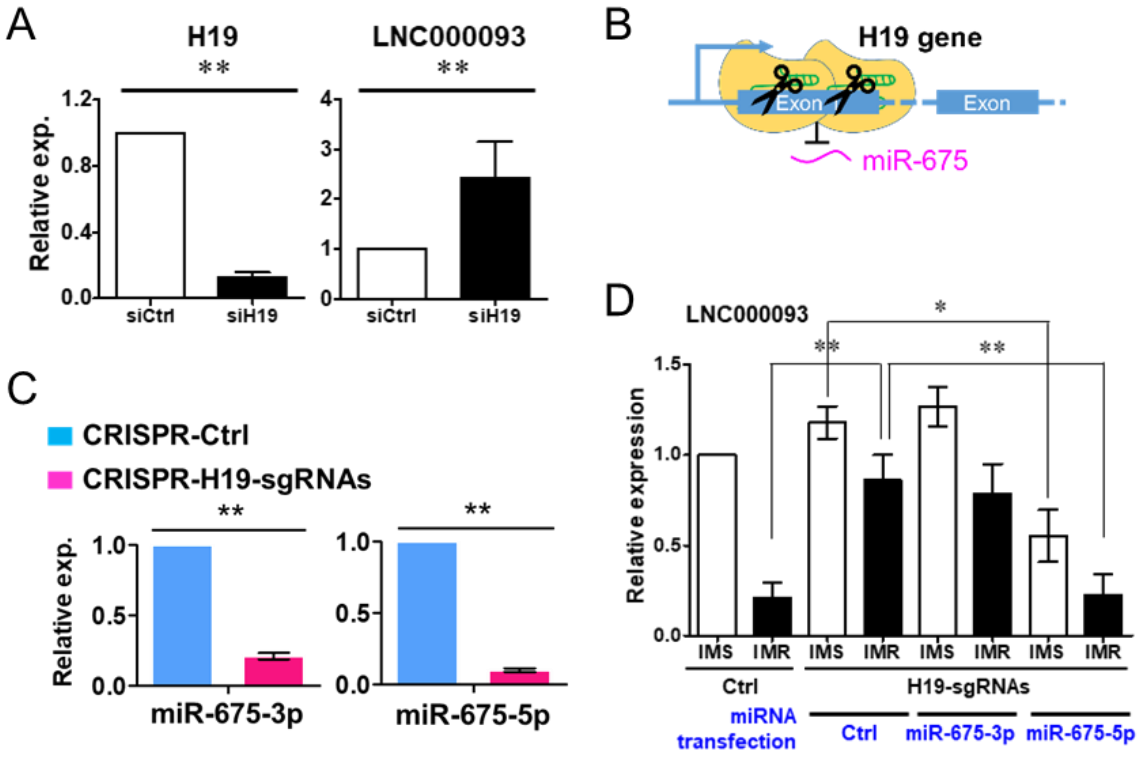


Figure 3.5.2 Expression of LNC000093 upon H19/miR-675 perturbation.

(A) RT-qPCR analysis revealed an increased LNC000093 expression in K562 cells transfected with H19-siRNA (n=3). (B) A schematic diagram showing the CRISPR-Cas9-mediated deletion of H19 exon one in order to suppress miR-675 expression. (C) RT-qPCR analysis showed significant downregulation of both miR-675-3p and miR-675-5p after CRISPR-deletion of H19/miR-675 (n=3). (D) RT-qPCR analysis revealed an upregulation of LNC000093 in K562-IMR cells after H19/miR-675-deletion (n=3). The increased level of LNC000093 in H19/miR-675-deleted K562-IMR was reversed by the transfection of miR-675-5p mimics. Overexpression of miR-675-5p also caused a downregulation of LNC000093 in K562-IMS after H19/miR-675-deletion (n=3). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.5.3 Validation of binding between LNC000093 and miR-675-5p by luciferase reporter assay

Luciferase reporter assays were performed to determine whether miR-675-5p exerts its effects by directly interacting with LNC000093. Luciferase reporter vectors were produced by cloning the LNC000093 constructs with wild type (WT) or mutated (MT) binding region for miR-675-5p to the downstream of the luciferase gene (Figure 3.5.3A). The vectors were then co-transfected with either miR-675-5p or miR-675-3p mimics into K562 cells. Firefly luciferase activity was measured and normalized to Renilla luciferase. Results showed that miR-675-5p bound the LNC000093 fragment containing wild type binding site and resulted in a reduction of luciferase activity by 48% compared to the control group transfected with miRNA negative control mimic (Figure 3.5.3B). The suppressed luciferase activity was abrogated when mutant LNC000093 luciferase vector was used or when miR-675-3p instead of miR-675-5p was co-transfected (Figure 3.5.3B). The combined results suggested the occurrence of direct binding event between LNC000093 and miR-675-5p.

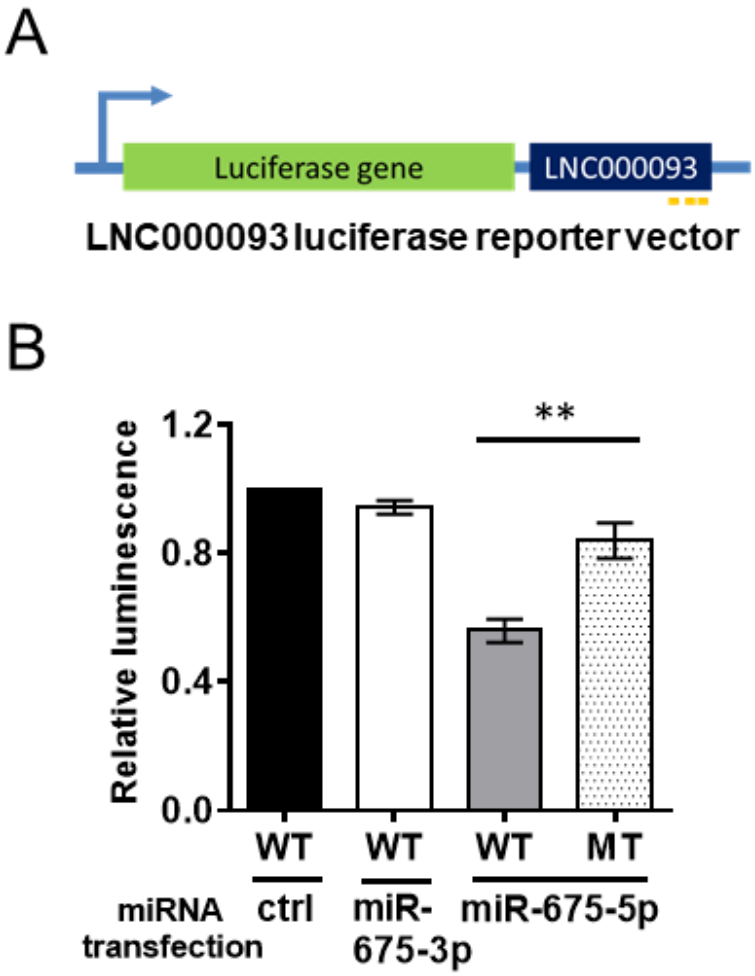


Figure 3.5.3 Luciferase reporter assay validated the direct binding between LNC000093 and miR-675-5p.

(A) LNC000093-luciferase reporter vectors containing either wild-type (WT) or mutant (MT) binding sequence for miR-675-5p were generated by insertion of the full-length LNC000093 cDNA sequence into the 3' end of luciferase gene. (B) The luciferase activity was significantly reduced only in cells co-transfected with miR-675-5p and WT-luciferase vector (n=3). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.6 Investigation of LNC000093-miR-675-RUNX1 axis in IMR CML cells

3.6.1 Cell viability of IMR CML cells upon H19/miR-675 perturbation

To examine the potential role of H19/miR-675 in IMR CML cells, loss-of-function study was performed on K562 cells with siRNA knockdown or genetic manipulation. Inhibition of H19/miR-675 using H19-siRNA, which increased K562-IMR cell death by 3.5-fold and LAMA84-IMR cell death by 2.8-fold relative to their corresponding IMS counterpart (Figure 3.6.1A). By inhibition of miR-675 via CRISPR-Cas9-mediated deletion of H19, the cell death percentage of K562-IMR was elevated by 3.6-fold upon 10 μM imatinib treatment for 72 hours (Figure 3.6.1B). Overexpression of miR-675-5p mimics reversed such increase in cell death, while miR-675-3p overexpression could not (Figure 3.6.1B).

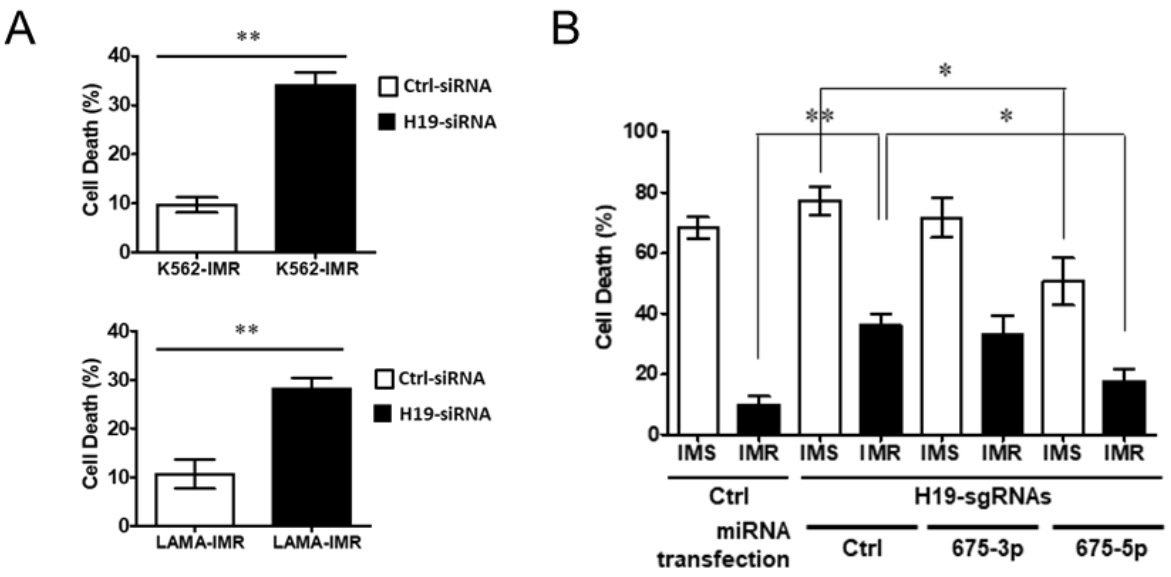


Figure 3.6.1 Cell viability of IMR CML cells upon H19/miR-675 perturbation.

(A) H19 inhibition by siRNA enhanced the cell death of K562-IMR and LAMA84-IMR under 10 μM imatinib treatment for 72 hours (n=3). (B) CRISPR-Cas9-mediated H19/miR-675 deletion enhanced the cell death of K562-IMR under imatinib treatment. Such increase in cell death was reversed by ectopic expression of miR-675-5p (n=3). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.6.2 Cell viability of K562-IMR upon imatinib treatment after overexpression of LNC000093

Transfection experiments of LNC000093 and miR-675-5p or negative control mimics were done in K562-IMR. The transfected cells were subsequently treated with 10 μM imatinib for 72 hours and then subjected to trypan blue assay. Results showed that forced expression of LNC000093 led to increased cell death of K562-IMR in response to imatinib treatment. In the control group co-transfected with miRNA control mimic, the cell death percentage increased from 11.1% to 40.0%. Similarly, the cell death percentage in the miR-675-5p transfected cells increased from 8.7% to 36.6% (Figure 3.6.2).

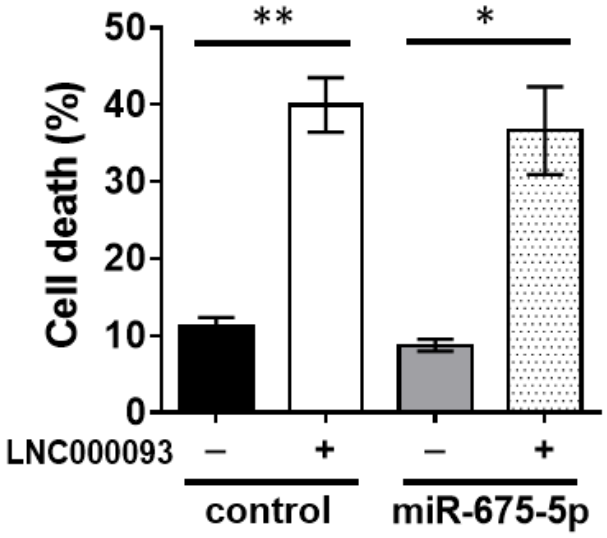


Figure 3.6.2 Cell viability of K562-IMR upon imatinib treatment after overexpression of LNC000093.

Ectopic expression of LNC000093 significantly increased K562-IMR cell death upon imatinib treatment with or without co-expression of miR-675-5p mimics (n=3). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.6.3 Expression change of RUNX1 protein in IMR CML cells

From the prediction analysis results performed using STarMir, *RUNX1* mRNA possesses three putative MREs for miR-675-5p and hence its expression is potentially regulated by miR-675-5p (Figure 3.6.3A). Western blotting revealed a reduced RUNX1 protein expression in both K562-IMR and LAMA84-IMR cells compared to their corresponding IMS cell counterparts (Figure 3.6.3B). Semi-quantification of the western blot results demonstrated the RUNX1 expression difference between IMS and IMR cells is statistically significant in both CML cell lines. To further validate whether RUNX1 expression can be regulated by miR-675-5p, synthetic miR-675 mimics were transfected into K562 and cells were collected after 48 hours for protein assay. Western blotting showed a significant reduction of RUNX1 expression in K562 cells transfected with miR-675-5p, but no significant change was found when transfected with miR-675-3p mimic (Figure 3.6.3C).

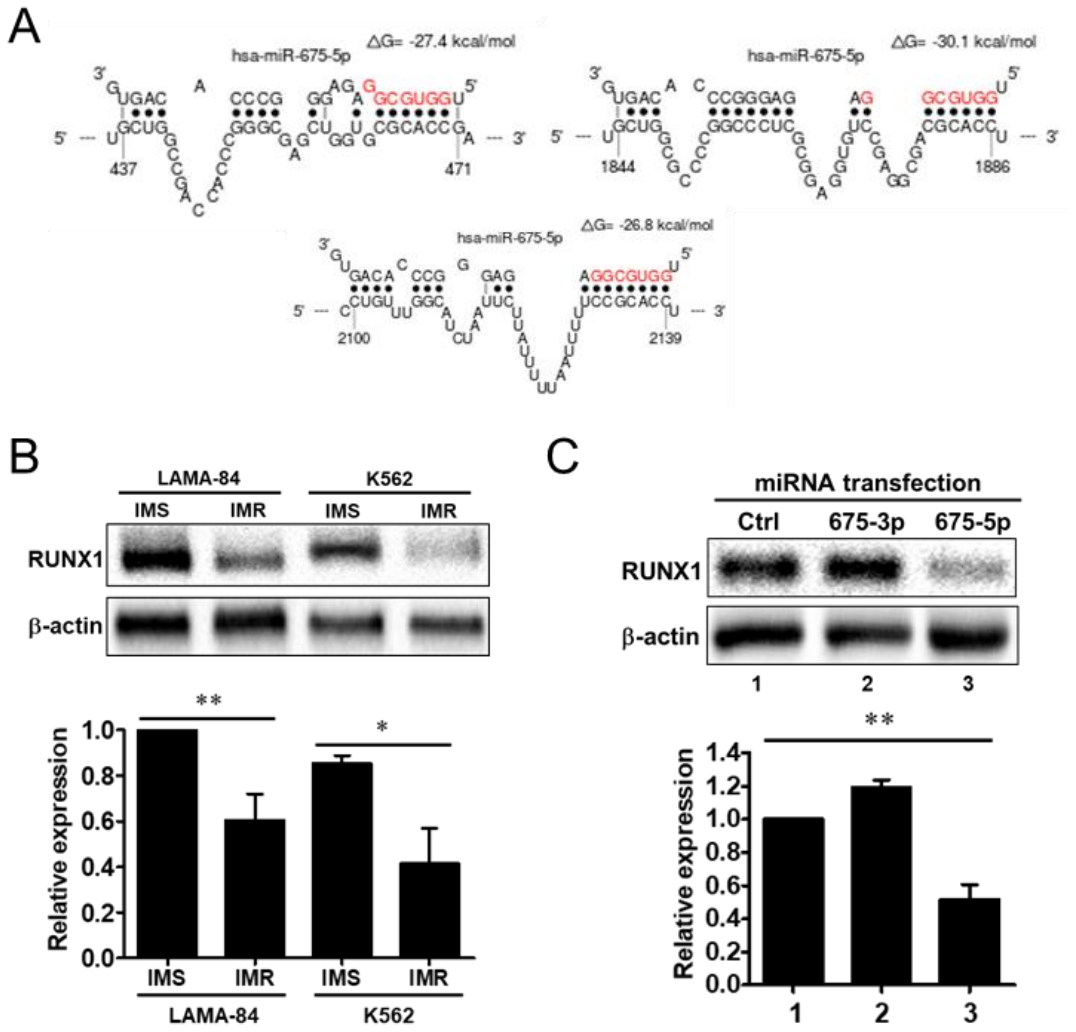


Figure 3.6.3 Expression change of RUNX1 protein in IMR CML cells.

(A) The potential interaction between RUNX1 and miR-675-5p was predicted by *in silico* tool STarMir and result revealed three putative binding regions matched to the seed sequence GGUGCG on miR-675-5p. (B) Western blots revealed a downregulation of RUNX1 protein in both K562-IMR and LAMA84-IMR (n=3). (C) Transfection of miR-675-5p but not miR-675-3p resulted in a downregulation of RUNX1 in K562 cells as shown by western blot results (n=3). Semi-quantification data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.6.4 Expression change of RUNX1 protein in K562 upon co-expression of LNC000093 and miR-675-5p

Co-transfection experiments of LNC000093 overexpression vector and synthetic miRNA mimics were performed to further investigate the potential ceRNA role of LNC000093. Different transfection conditions were done in K562 cells for 48 hours and then total protein was extracted for subsequent assays. Western blot results showed an enhanced RUNX1 expression (4.6-fold) in cells co-transfected with LNC000093 expression vector and miR-675-5p mimic, indicating the miR-675-5p-mediated suppression of RUNX1 is rescued by overexpression of LNC000093 (Figure 3.6.4). In addition, the RUNX1 expression in the cells co-transfected with miRNA control and LNC000093-expressing vector was also increased by 3.1-fold (Figure 3.6.4).

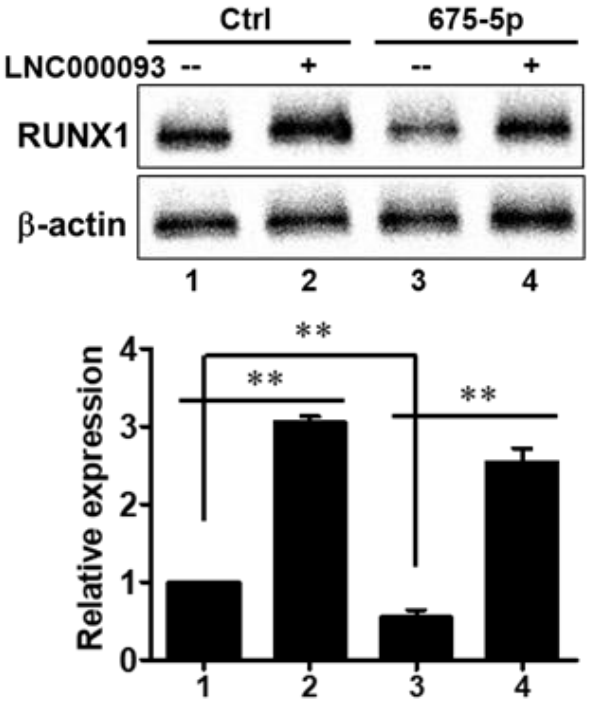


Figure 3.6.4 Expression change of RUNX1 protein in K562 upon co-expression of LNC000093 and miR-675-5p.

An upregulation of RUNX1 protein in cells overexpressing LNC000093 was shown by western blots. The miR-675-5p-mediated RUNX1 repression was rescued by ectopic expression of LNC000093. Semi-quantification data are displayed as mean ± SD (n=3). Student's t-test was used for comparisons between two groups.

3.7 Investigation of LNC000093 in differentiation of iPSCs

3.7.1 Morphology of iPSCs upon spontaneous and directed hematopoietic differentiation

To further investigate the potential functional role of LNC000093 in hematopoietic differentiation, human iPSCs were utilized as an *in vitro* model for differentiation. Spontaneous and hematopoietic differentiation were conducted with iPSCs through embryoid body formation (Figure 2.2). The morphology of undifferentiated iPSC and after different schedule of differentiation was shown in Figure 3.7.1. Undifferentiated iPSCs is grown in form of flat monolayer colonies with tight cellular packing showing prominent nucleoli (Figure 3.7.1A). Embryoid bodies on day 7 appear as dense mass of cell clumps with a regular spherical shape suspended in the culture medium (Figure 3.7.1B). For cells directed to hematopoietic differentiation with specified cocktail medium, the embryoid body-like multilayer cell clumps attached onto the culture plate surface and start to spread out on day 7 (Figure 3.7.1C). Further extensive spreading to larger area and enlarged mass of cellular clump with irregular shape were found on day 14 of differentiation (Figure 3.7.1D).

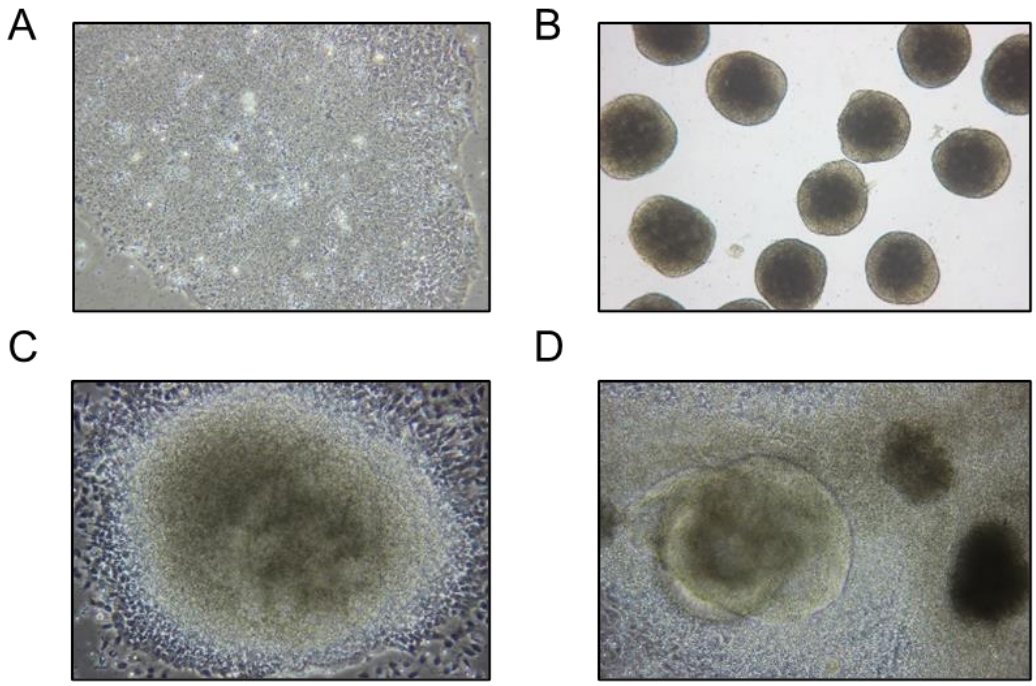


Figure 3.7.1 Morphology of iPSCs upon spontaneous and directed hematopoietic differentiation.

(A-D) Representative bright field images showing the morphologies of undifferentiated iPSCs (A), embryoid bodies on day 7 (B), and cells undergoing directed hematopoietic differentiation for 7 days (C) and 14 days (D). 40X magnification for (B); 100X magnification for (A, C & D)

3.7.2 Expression of LNC000093 during spontaneous differentiation

The expression level of LNC000093 and three selected early hematopoietic lineage markers (CD34, CXCR4 and GATA2) were detected using RT-qPCR in differentiated embryoid bodies (EBs) and compared to undifferentiated iPSCs on day 0. After 7 days of spontaneous differentiation, LNC000093 expression was upregulated by 9.82-fold, and CD34, CXCR4 and GATA2 were also upregulated by 5.06-fold, 45.44-fold, and 11.02-fold, respectively (Figure 3.7.2).

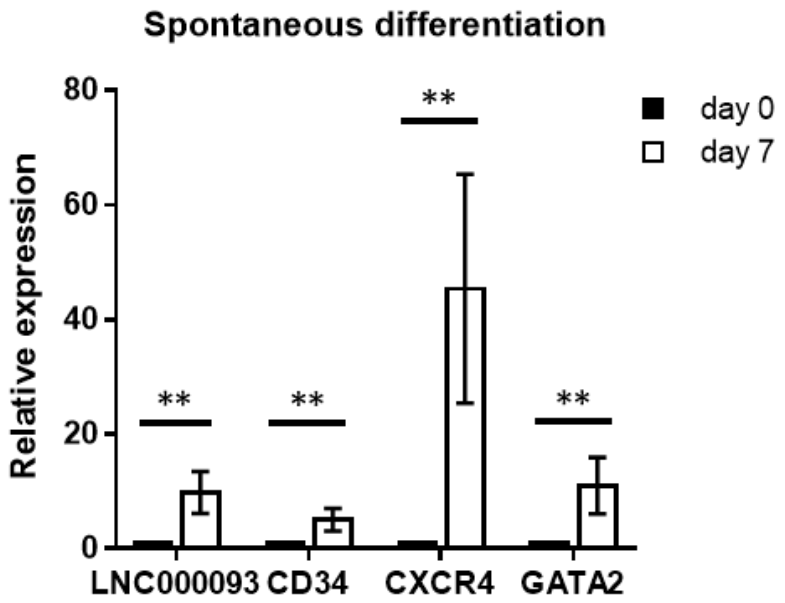


Figure 3.7.2 Expression of LNC000093 during spontaneous differentiation.

After 7 days of spontaneous differentiation, an increased expression of LNC000093, CD34, CXCR4 and GATA2 in EBs was revealed by RT-qPCR analysis (n=4). Data are displayed as mean ± SD. Student's t-test was used for comparisons between two groups.

3.7.3 Expression of LNC000093 during directed hematopoietic differentiation

For directed hematopoietic differentiation under treatment of cytokine cocktails for 7 or 14 days, a similar expression trend for the target genes was obtained (Figure 2.1). After 7 days of hematopoietic differentiation, LNC000093 expression was upregulated by 5.98-fold relative to undifferentiated iPSCs. The expression levels of CD34, CXCR4 and GATA2 were also increased by 11.21-fold, 25.82-fold and 56.63-fold, respectively (Figure 3.7.3). On day 14 of hematopoietic differentiation, all the gene expression levels were further highly upregulated relative to iPSCs on day 0. LNC000093 expression was increased by 33.93-fold, while CD34, CXCR4 and GATA2 were upregulated by 41.94-fold, 57.29-fold, and 229.4-fold, respectively (Figure 3.7.3).

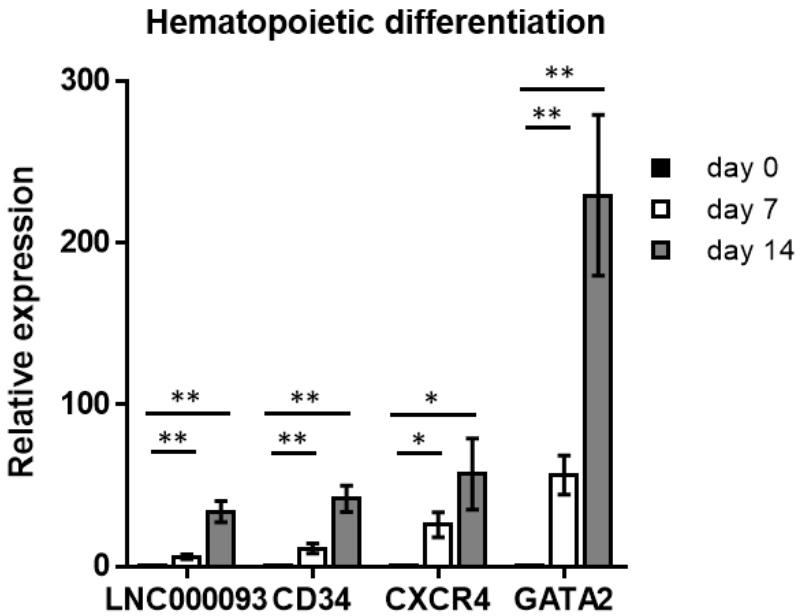


Figure 3.7.3 Expression of LNC000093 during directed hematopoietic differentiation.

An increased expression of LNC000093, CD34, CXCR4 and GATA2 after 7 or 14 days of hematopoietic differentiation was revealed by RT-qPCR analysis (n=6). Data are displayed as mean ± SEM. Student's t-test was used for comparisons between two groups.

3.7.4 Investigation of LNC000093-deletion in short-term spontaneous differentiation

To investigate the involvement of LNC000093 in differentiation process, loss-of-function study was conducted in iPSCs. A pair of sgRNAs was designed and CRISPR-Cas9-mediated deletion were performed to cut off the genomic region with full-length LNC000093 (Figure 3.7.4A). The deletion effect was confirmed by conventional PCR with gDNA, and gel electrophoresis results showed an extra band in LNC000093-CRISPR sample, indicating a successful CRISPR-mediated deletion of LNC000093 in iPSCs (Figure 3.7.4B). RT-qPCR analysis also revealed a reduction of LNC000093 expression at transcript level after CRISPR-deletion, giving 0.44-fold change relative to the control iPSCs (Figure 3.7.4C).

Then, short-term spontaneous differentiation (3 or 7 days) was conducted using the LNC000093-deleted iPSCs. RT-qPCR results revealed that the expression of all the three differentiation markers (CD34, CXCR4 and GATA2) demonstrated a reduced increment (ranged from 0.56-fold to 0.72-fold after normalization to control) in LNC000093-CRISPR-deleted cells compared to the control cells after 3 or 7 days of differentiation (Figure 3.7.4D).

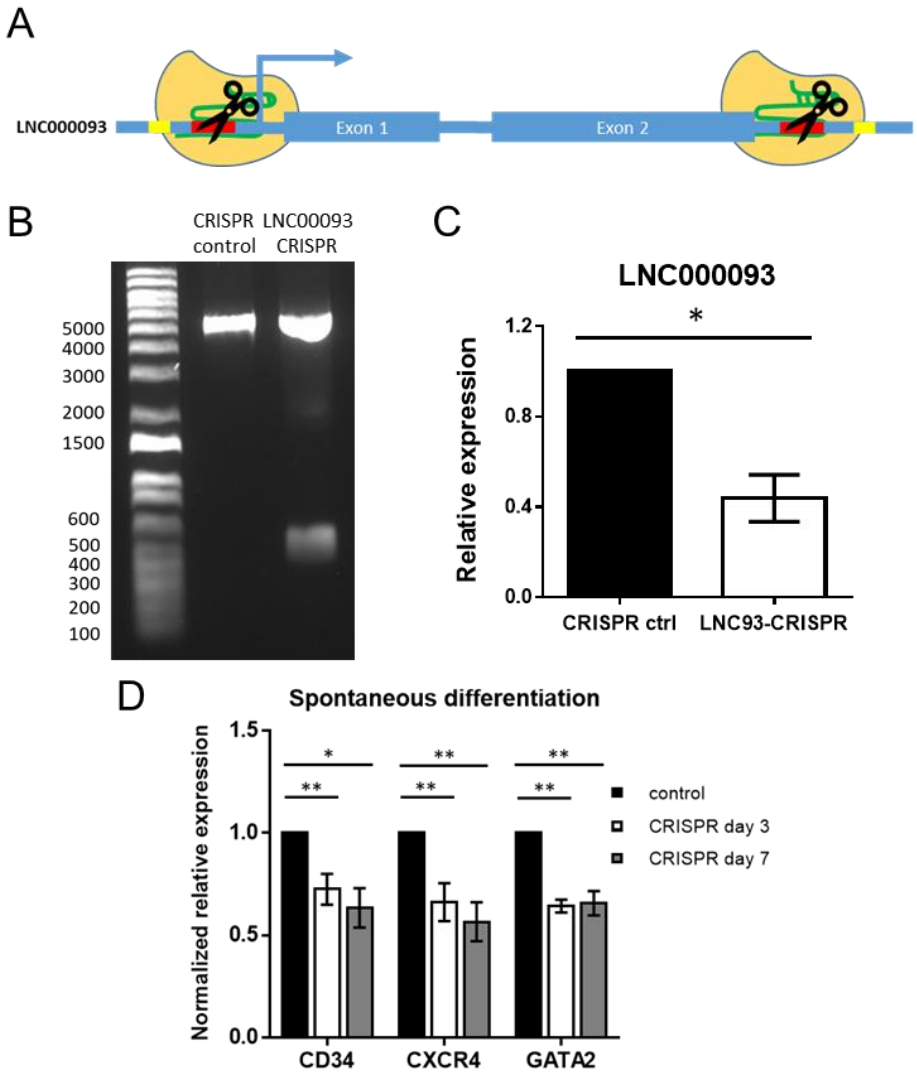


Figure 3.7.4 Investigation of LNC000093-deletion in short-term spontaneous differentiation.

(A) A schematic diagram showing Cas9 enzymes are guided by sgRNAs to their target regions (red) and generate cleavage to delete the full-length of LNC000093 genomic region. A pair of primers flanking the LNC000093 (yellow) are designed to confirm the deletion by PCR. (B) Gel electrophoresis showed an extra band for the LNC000093-CRISPR sample, indicating a successful deletion. The parental amplicon size is around 5 kb and LNC000093-deletion would lead to a shorter amplicon size with 556 bp. (C) RT-qPCR analysis revealed a reduced LNC000093 expression after CRISPR-deletion in iPSCs (n=3). (D) RT-qPCR analysis showed a reduction in fold change of CD34, CXCR4 and GATA2 in LNC000093-deleted EBs after normalized to control EBs (n=6). Data are displayed as mean ± SEM. Student's t-test was used for comparisons between two groups.

3.8 Investigation of potential regulation of chromatin accessibility by LNC000093

3.8.1 Peak calling analysis

During embryonic development, great changes of epigenetic profile occur in cells and the epigenetic regulation of chromatin structure is essential to the repression or activation of genes. Basically, chromatins that are "open" and "closed" represent the active and repressed transcriptional states of certain genes, respectively. To further examine the possible role of LNC000093 in epigenetic regulation during iPSC differentiation, scATAC-seq was performed to assess the change of chromatin accessibility with control iPSCs (day 0) and EBs (day 3) as well as their LNC000093-CRISPR-deleted counterparts.

After pre-processing of scATAC-seq data, the next major step is to identify chromatin accessible regions that also refer to as “peaks”, which is the basis for subsequent advanced analyses. Initial peak calling analysis revealed a lower number of peaks in both day 0 and day 3 LNC000093-CRISPR-deleted samples compared to the corresponding control samples (Figure 3.8.1). This reflected a global reduction of chromatin accessibility after LNC000093 is knockdown by CRISPR-Cas9 system.

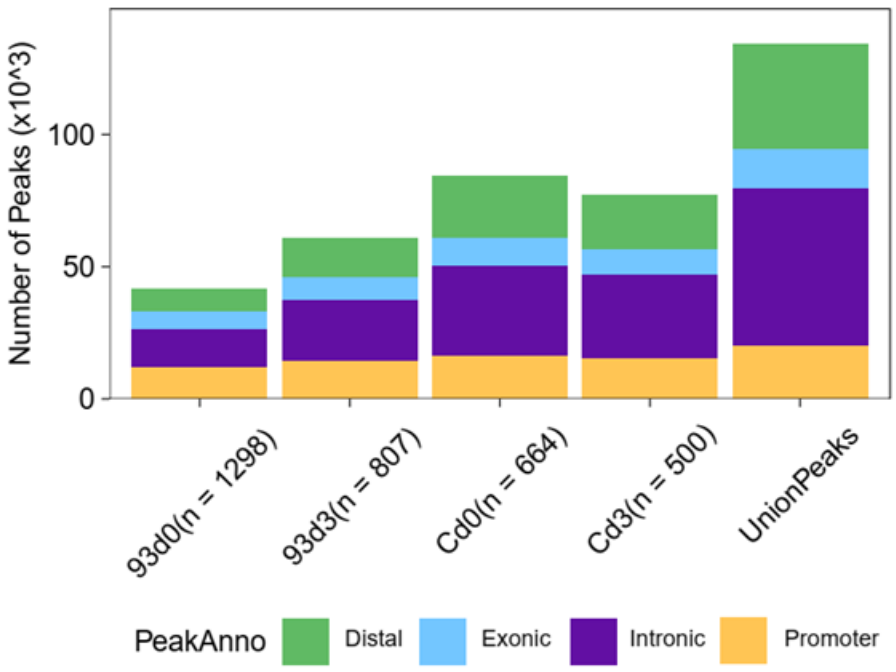


Figure 3.8.1 Peak calling analysis of scATAC-seq data.

Initial peak calling analysis showed a lower number of peaks in LNC000093-CRISPR-deleted iPSCs or EBs compared to the corresponding control samples. Union peaks represent the total combined number of peaks identified in all samples. The colors indicate the properties of the genome location of peaks. The number in brackets represents the cell number of that sample. Abbreviations: C = control; 93 = LNC000093-deleted; d0 = day 0; d3 = day 3

3.8.2 Dimensionality reduction and clustering analysis

The identification of cellular subsets by dimensionality reduction and clustering is one of the major applications of single-cell analysis. Briefly, the cells are grouped into different clusters based on their proximity of chromatin accessibility profile and then visualized on a two-dimensional UMAP plot (Figure 3.8.2A-D). The UMAP plot demonstrates how spatially separable classes are in relation to a selected set of features, chromatin accessibility in this case, on the basis of a method known as "Uniform Manifold Approximation and Projection". Therefore, the distance between different cells or clusters in the UMAP plot could reflect their similarity in chromatin accessibility profile.

In the first UMAP plot (Figure 3.8.2A), the cell distribution of all four samples is shown and this clearly revealed a difference between control (green) and LNC000093-deleted iPSCs (red), as well as the difference after iPSCs were differentiated into EBs (purple and blue). Clustering analysis showed that eight clusters could be identified among the four samples based on their global profiles of chromatin accessibility (Figure 3.8.2B) and the corresponding source of each sample in the clusters is directly reflected in Figure 3.8.2A with the same location. The relative abundance of each sample in each cluster is shown in Figure 3.8.2C. Control iPSCs and LNC000093-deleted iPSCs are diversely distributed among cluster 1, 6, 7 and 8, whereas both EB samples are quite evenly distributed along cluster 2 to 5, except cluster 4 is mainly enriched with LNC000093-deleted EBs.

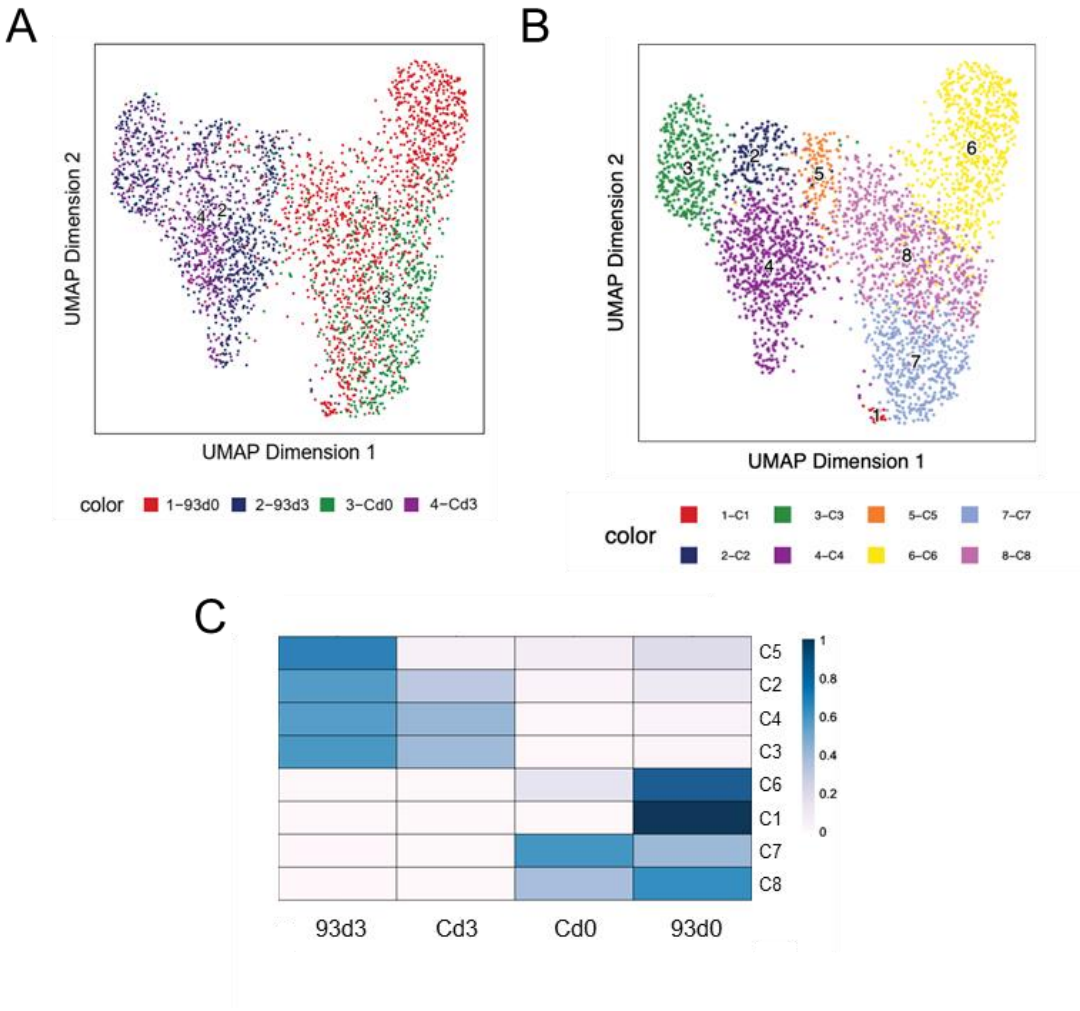


Figure 3.8.2 Dimensionality reduction and clustering analysis.

(A) The UMAP plot showed the cell distribution of all four samples according to their chromatin accessibility, each dot represents a single cell. A closer distance between cells means a higher similarity in chromatin accessibility profile. (B) Clustering analysis further identified eight clusters in the pool four samples. (C) The relative abundance of each iPSC or EB sample in each cluster is shown. The darker color reflects more cells of this sample compose the cluster. Abbreviations: C = control; 93 = LNC000093-deleted; d0 = day 0; d3 = day 3

3.8.3 Chromatin accessibility of LNC000093 loci after CRISPR-mediated deletion

The chromatin accessibility profile of LNC000093 loci was assessed and results revealed an absence of peaks on LNC000093 loci in LNC000093-CRISPR-deleted iPSCs (Figure 3.8.3A) and EBs (Figure 3.8.3B), suggesting the transcriptional activity of LNC000093 is silenced upon CRISPR-mediated-deletion as reflected by the reduction of chromatin accessibility.

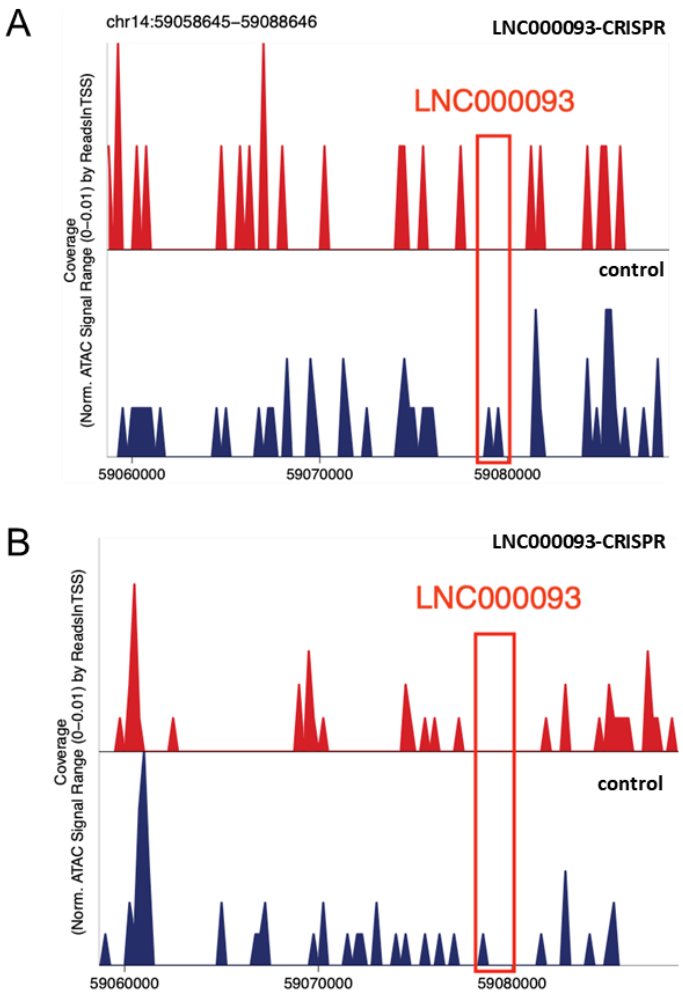


Figure 3.8.3 Chromatin accessibility of LNC000093 loci after CRISPR-mediated deletion.

The examination of chromatin accessibility profile of LNC000093 loci revealed an absence of peaks in LNC000093-CRISPR-deleted iPSCs (A) and EBs (B).

3.8.4 Global differential accessible peak analysis

The global chromatin accessibility profile of all samples was assessed, and results showed 6839 downregulated and 195 upregulated peaks in LNC000093-CRISPR-deleted iPSCs compared to control iPSCs, while there were only 50 downregulated and 12 upregulated peaks when LNC000093-CRISPR EBs compared to control EBs (Figure 3.8.4A). The differential accessible peak analysis is also shown in volcano plot format (Figure 3.8.4B & C) which reflected the high statistical significance ($-\log_{10}$FDR) of the differential chromatin accessibility, particularly the downregulated peaks in LNC000093-CRISPR iPSCs (Figure 3.8.4B). The collective results showed a global reduction of chromatin accessibility in LNC000093-deleted cells relative to the control, especially in undifferentiated iPSCs.

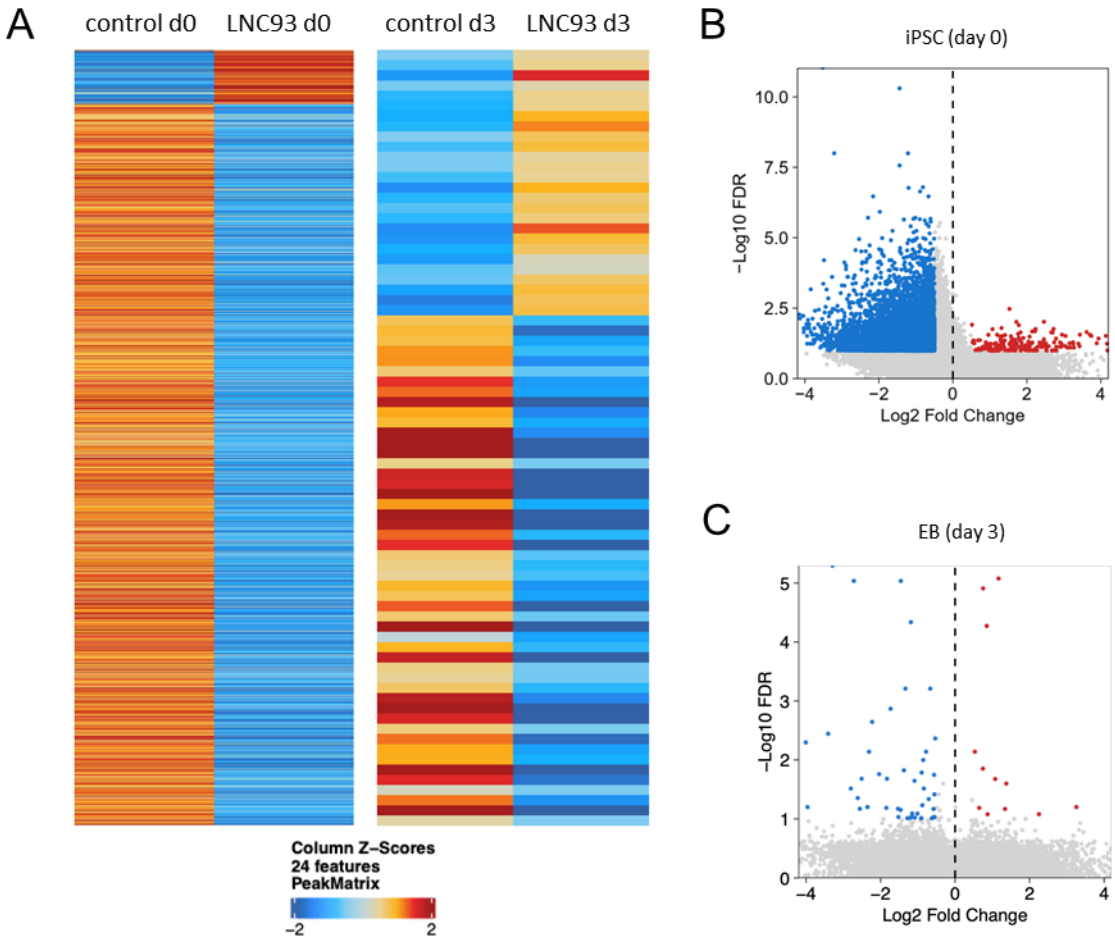


Figure 3.8.4 Global differential accessible peak analysis.

(A) Heat maps showed the global differential chromatin accessibility of all samples, given that LNC000093-CRISPR iPSCs compared to control iPSCs and LNC000093-CRISPR EBs compared to control EBs. (B) Volcano plot showed the differential accessible peaks in iPSCs. (C) Volcano plot showed the differential accessible peaks in EBs. Red and blue color represents upregulated and downregulated peaks, respectively.

3.8.5 Functional annotation of differential accessible regions

After identification of the differential accessible peaks, the next step is to search for their related genes and functions. The closest genes to each peak was determined by BEDtools and then functional enrichment analysis of these genes was performed using g:Profiler. The analysis focused on the downregulated peaks in LNC000093-deleted iPSCs in order to identify the functions disrupted by the silencing of LNC000093. For day 3 EB samples, the number of differential accessible peaks are too small to perform such analysis. From the biological process enrichment analysis, the top 20 enriched clusters were identified (Figure 3.8.5). The results revealed the differential accessible peak-related genes may involve in biological functions associated with various developmental processes (GO: 0032502, GO: 0007399, GO: 0048856, GO: 0048731, GO: 0007275, GO: 0048869, GO:0048468), cell differentiation (GO: 0030154, GO: 0030182) and morphogenesis processes (GO: 0000902, GO: 0009653, GO:0032989, GO:0048812). In other words, these biological processes may be impeded after LNC000093-CRISPR-deletion in iPSCs.

GO:BP		stats	
Term name	Term ID	padj	−log10(padj) 0 ≤16
anatomical structure morphogenesis	GO:0009653	1.135×10^{-19}	
nervous system development	GO:0007399	2.129×10^{-19}	
developmental process	GO:0032502	3.790×10^{-19}	
anatomical structure development	GO:0048856	6.234×10^{-19}	
system development	GO:0048731	7.055×10^{-17}	
generation of neurons	GO:0048699	2.765×10^{-16}	
neurogenesis	GO:0022008	3.554×10^{-16}	
multicellular organism development	GO:0007275	6.135×10^{-16}	
cell differentiation	GO:0030154	4.734×10^{-15}	
cellular developmental process	GO:0048869	5.196×10^{-15}	
neuron differentiation	GO:0030182	5.899×10^{-15}	
cellular component organization	GO:0016043	5.921×10^{-14}	
cell morphogenesis	GO:0000902	1.018×10^{-13}	
cell development	GO:0048468	4.899×10^{-13}	
cellular component morphogenesis	GO:0032989	8.308×10^{-13}	
actin cytoskeleton organization	GO:0030036	1.239×10^{-12}	
actin filament-based process	GO:0030029	2.246×10^{-12}	
positive regulation of cellular process	GO:0048522	3.719×10^{-12}	
cellular component organization or biogenesis	GO:0071840	2.344×10^{-11}	
neuron projection morphogenesis	GO:0048812	3.451×10^{-11}	

Figure 3.8.5 Functional annotation of differential accessible peaks.

The bar graph showed top 20 enriched GO terms determined by g:Profiler across differential accessible peak-related genes and is colored according to adjusted *p*-values.

3.8.6 Gene score analysis

Gene expression could be estimated by chromatin accessibility data by the analysis of gene scores. Basically, gene scores are the prediction of gene expression levels based on the accessibility of regulatory elements adjacent to the gene. Results of gene score analysis showed a total of 747 differentially expressed genes between two undifferentiated iPSC samples, in which 666 genes were downregulated in LNC000093-deleted iPSC compared to the control (Figure 3.8.6A). For the comparison of EB samples, 89 genes were showing differential expression and 68 of these are downregulated in the LNC000093-deleted EBs (Figure 3.8.6B). Notably, two genes showing highly significant downregulation in both LNC000093-CRISPR iPSCs and EBs were observed, which are HIST1H3C and HIST1H2BB.

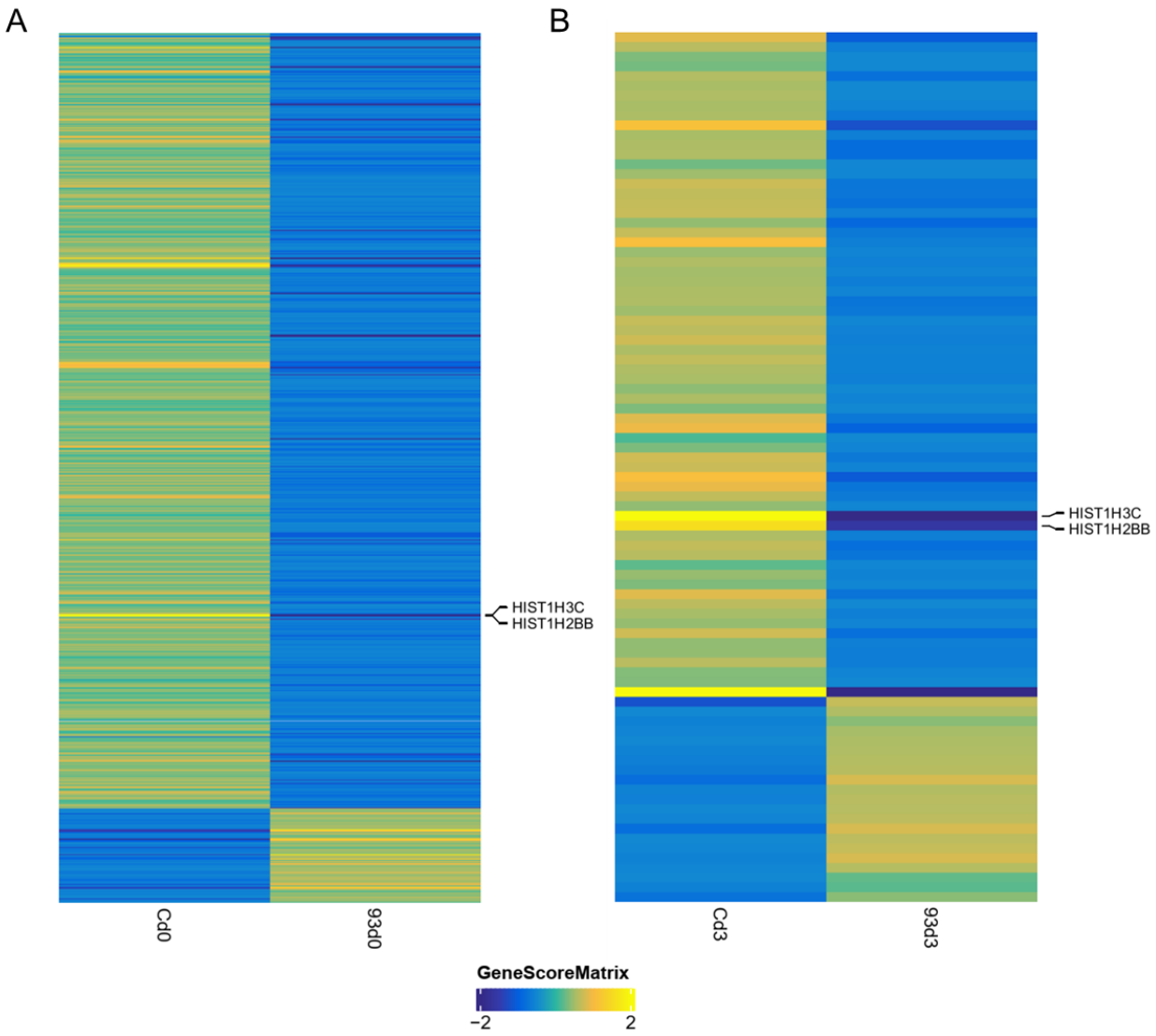


Figure 3.8.6 Gene score analysis.

(A) The heap map showed the differentially expressed genes between LNC000093-CRISPR iPSC and control iPSC determined by gene score analysis. (B) The heap map showed the differentially expressed genes between LNC000093-CRISPR EBs and control EBs determined by gene score analysis. The intensity of color reflects the level of expression change. Two genes showing significant differential expression across two set of comparisons is highlighted, which are HIST1H3C and HIST1H2BB.

Chapter 4 – Discussion

4.1 Identification of lncRNAs involved in *JAK2*-V617F signaling

The present study investigated the lncRNAs participating in both Ph-positive and Ph-negative MPNs, where signaling pathways are dysregulated by aberrant tyrosine kinases due to the presence of molecular abnormalities like somatic mutation and chromosomal rearrangement. As mentioned in Chapter 1.7, there have been various lncRNAs being implicated in Ph^+ MPN (i.e. CML), but none has been reported in Ph-negative classical MPNs. Hence, with the aid of cell line possessing *JAK2*-V617F mutation, which is the main driver mutation of classical MPNs, my first part of study aimed to explore any lncRNAs being involved in this mutant JAK2 signaling. It is anticipated that my findings could confer more insights into the molecular basis of MPN pathogenesis and progression.

In order to search for lncRNAs involved in the *JAK2*-V617F signaling, which is dysregulated by the over-stimulation of JAK2 kinase and its downstream pathway, HEL cell line was used as a model as it possesses homozygosity for the *JAK2*-V617F mutation (Martin and Papayannopoulou, 1982; Quentmeier et al., 2006). By direct interruption of the JAK2 pathway using the chemical drug ruxolitinib, which is the first FDA-approved drug for treating MPNs (Raedler, 2015), differentially expressed lncRNAs has been revealed. It is expected that the responsive lncRNAs are either the regulators involved in the signaling pathway or their expression is regulated by the pathway.

The initial screening was done by qPCR array with a profile of 84 tumorigenesis related lncRNAs, and the results showed that most targets appeared to be non-responsive to the

in vitro JAK2 inhibition in HEL cells (Figure 3.1.1A). Most lncRNAs in the panel showed a small magnitude of expression changes (< 2-fold changes), and this observation may imply that they are not directly involved in JAK/STAT pathway, but rather play a role in other closely related pathways such as PI3K/AKT and MAPK/ERK pathways (Levine et al., 2007). In addition, such slight expression change may imply the mode of their regulatory role since *cis*-acting lncRNAs are generally expressed at lower level and more transiently than *trans*-acting lncRNAs that act on distant targets (Geisler and Coller, 2013).

Still, a total of six lncRNA candidates showed significantly differential expression upon JAK2 inhibition in HEL cells (Table 3.1.1), and they were all subjected to individual RT-qPCR validation with manually re-designed primers. Basically, all the lncRNAs showed analogous expression changes as that in qPCR array data but only four of them were statistically significant, which are BANCR, CBR3-AS1, LINC00261 and LINC00887 (Figure 3.1.1B).

4.2 Investigation of BANCR in *JAK2*-V617F signaling

4.2.1 BANCR showed specific and positive correlation with the activity of *JAK2*-V617F signaling

Among the four selected and validated lncRNAs mentioned in Chapter 4.1, BANCR showed the greatest fold change and hence it was chosen for further investigation. Different dosage of ruxolitinib treatment in HEL cells revealed a dose-dependent downregulation effect on BANCR expression (Figure 3.1.2A). Opposing to inhibiting JAK2, the effect on BANCR expression after overexpression of mutant JAK2 was also investigated. For this purpose, HEK293T cell line was used since it is an excellent cell model for transient transfection experiments with high efficiency and without known disease (Thomas and Smart, 2005). The upregulation of BANCR upon ectopic expression of *JAK2*-V617F in HEK293T demonstrated the overactivation of JAK/STAT signaling could regulate BANCR in general but not specific to blood or leukemic cells (Figure 3.1.3B).

The positive correlation between the activity of JAK/STAT pathway and BANCR expression implicated that BANCR is likely a downstream target of JAK/STAT pathway (i.e. the expression of BANCR is directly regulated by JAK/STAT signaling). However, ruxolitinib treatment did not induce any expression change of BANCR in K562 cell line, which possesses a wild-type JAK2 (Figure 3.1.2B). This finding strongly objected to the above-mentioned idea that JAK/STAT pathway could directly regulate BANCR expression. Therefore, it is suggested that BANCR should be regulated specifically by mutant *JAK2*-V617F signaling via certain mechanisms, or it is regulated through JAK/STAT pathway but in an indirect way.

4.2.2 Potential ceRNA networks could be established by BANCR-miR-3609-JAK/STAT gene axis

Accumulating evidence has shown that lncRNAs could act as miRNA sponges and are involved in competitive regulatory interactions by serving as ceRNAs, which sequester miRNAs and subsequently diminish their repressive effects on other transcripts (Chan and Tay, 2018b; Sun et al., 2020). To be concise, ceRNAs are any transcripts that could regulate each other through competition for one or more shared miRNA, so both mRNA and lncRNA are possible to work in such post-transcriptional regulation and act as ceRNAs. Thus, I started to investigate whether BANCR can also be involved in JAK/STAT signaling through a ceRNA network.

miRNAs act through a seed region composed of the nucleotides 2-8 from 5' end that is crucial for recognition and interaction with their target mRNAs or ncRNAs which possess corresponding miRNA response elements (MREs) that are predominantly located on the 3'UTR of the transcript (Lewis et al., 2005). Given that transcripts harboring MREs for specific miRNA could serve as ceRNAs (Qi et al., 2015), I performed *in silico* analysis to identify the putative miRNA and mRNA candidates that may participate in the ceRNA network with BANCR. Several miRNA targets were selected from a published research study by Pagano et al. as mentioned in Chapter 3.2.3. In a ceRNA interplay model, upregulation of one transcript will result in more shared miRNA to be sequestered and hence another transcript will be de-repressed, and vice versa (Qi et al., 2015). In that sense, as BANCR was downregulated upon JAK2 inhibition in HEL cells, I need to search for miRNA showing upregulation since fewer competition can be occurred, and the shared miRNA would become more freely available. Hence, the top five significantly upregulated miRNA upon JAK2 inhibition

in HEL cells were selected from Pagano's data. After bioinformatics prediction analysis, miR-3609 has been selected and targeted from the five miRNAs since it is the only one that BANCR possesses potential MREs for it (Figure 3.2.3).

Then, I aimed to search for putative ceRNAs for BANCR from JAK/STAT pathway. The gene set for JAK/STAT pathway annotated by KEGG database was found, and their expression levels from my RNA-seq data were checked. Only those genes showing downregulated expression upon JAK2 inhibition were chosen for the prediction analysis, and results showed some of them harbor the same putative binding sites for miR-3609 (Table 3.2.3.2), indicating they could potentially act as ceRNAs of BANCR and regulate the expression of each other by competing for miR-3609.

In such scenario, overactivation of the JAK/STAT pathway due to *JAK2*-V617F mutation could lead to a low level of miR-3609 and in turn increase BANCR expression. Accordingly, JAK2 inhibition in HEL cells would restore the level of miR-3609 and indirectly repress BANCR. However, in normal JAK/STAT signaling, miR-3609 expression is not suppressed and JAK2 inhibition might not affect the miR-3609 level so much, hence BANCR expression is also not affected. This could provide a rational direction to explain why BANCR expression is not responsive to JAK2 inhibition in K562, but greatly decreased in HEL and SET2 cells, which possess aberrant JAK2 signaling. Indeed, JAK/STAT pathway is also activated by BCR-ABL1 oncoprotein in K562 cells, but genomic amplifications of JAK2 occur in both HEL and SET-2 cell lines, and lead to several fold of mutant JAK2 activity which far outweighs that in K562 (Quentmeier et al., 2006), hence the whole situation was mentioned in a simplified way.

At the moment, no further experiment has been done to validate this scheme, thus additional investigation is warranted to elucidate how BANCR is involved in the *JAK2*-V617F pathway. Nonetheless, my findings demonstrated the specificity of lncRNA regulation mechanisms in distinct oncogenic signaling pathways, as shown by Ph-positive and Ph-negative MPN cells.

4.3 Identification of unannotated lncRNAs involved in *JAK2*-V617F signaling

Besides known lncRNAs that may be involved in *JAK2*-V617F pathway, the present study also aimed to identify novel lncRNAs and explore new regulatory networks under the aegis of rapid advancement in next-generation sequencing and bioinformatics tools in order to broaden our knowledge. In total, 51 unannotated lncRNAs showed significant differential expression in HEL cells upon JAK2 inhibition by ruxolitinib treatment. Gene ontology (GO) enrichment analysis revealed the known differentially expressed genes are mainly involved in cytokine regulation and kinase signaling (Figure 3.2.2A), which is expected, since JAK/STAT signaling was disrupted by the JAK2 inhibitor. Other enriched biological functions and pathways include blood vessel development, exocytosis regulation, cellular ion homeostasis, chemotaxis, etc., which are so broad and with high functional diversity. Therefore, it is hard to imply the potential functions that the novel lncRNAs may take place based on the GO enrichment analysis, and hence the JAK/STAT signaling pathway was again focused. A gene co-expression network analysis was performed, and it was found that considerable numbers of lncRNAs, including known and novel ones, showed high co-expression correlation with each other as well as certain genes in the JAK/STAT pathway, especially *JAK2* (Figure 3.2.2B). This reflected a complicated network among these JAK/STAT-associated lncRNAs that deserves further investigation.

Similar to the investigation of BANCR mentioned in Chapter 4.2, *in silico* MRE prediction analysis was performed for the novel lncRNAs. My results showed that eight novel lncRNAs possess shared MREs for miR-3609, which were also found in BANCR and some JAK/STAT pathway genes (Chapter 3.2.3). Consequently, they could form a putative inter-regulatory network, in which these JAK/STAT-related genes and lncRNAs could act as ceRNAs and influence each others' expression by competing for miR-3609. Therefore, the current findings provide a prospect of exploring novel regulators in the *JAK2*-V617F signaling by examining the interplay of JAK/STAT pathway, miRNAs and lncRNAs, including the brand-new ones.

In addition, the correlation between MPN-related CNVs and the genomic loci of differentially expressed novel lncRNAs was investigated. Both MSTRG.2026 and MSTRG.1558 are differentially expressed novel lncRNAs in the RNA-seq data (Table 3.2.1), and are located within two different MPN-related CNV regions with gain and loss copy numbers, respectively (Table 3.2.4). It is expected that alteration of copy numbers could lead to aberrant expression levels of lncRNAs located within the CNV regions and hence result in deregulatory effects that ultimately facilitate the initiation or progression of disease. As an example, a lncRNA named SNHG6 has been shown to have a high expression in tissues and cells of colorectal cancer as a consequence of DNA copy number gains caused by amplification of this genomic region, which is frequently occurred in colorectal cancer (Xu et al., 2019). Further experimental works have revealed that SNHG6 promotes growth and metastasis of colorectal cancer cells, and upregulated level of SNHG6 is associated with tumor progression and poor prognosis of colorectal cancer. Therefore, it is clearly shown that CNV analysis could help to provide insights into clinical relevance of disease diagnosis and treatment. In

fact, all the Ph-negative MPN patient samples used for CNV analysis are from the PMF subtype (Table 2.2), which is a more severe form of MPNs, and has the highest risk of leukemic transformation among different MPN subtypes (Yogarajah and Tefferi, 2017). Hence, the identification of MPN-related CNVs and investigating their impacts on lncRNAs could confer certain clinical significance for understanding MPN progression, which is deserving of further investigation.

4.4 H19/miR-675 plays a role in drug resistance of CML cells

Accumulating studies have shown that the dysregulation of the lncRNA H19 is associated with diverse human diseases especially cancers, and the dual roles of H19 acting either as a tumor suppressor or an oncogene have been suggested, depending on disease type or cellular content (Yoshimura et al., 2018; Ghafouri-Fard et al., 2020). This lncRNA encodes a 2.3 kb functional transcript and it also serves as a precursor of miR-675, which is encoded by the first exon of H19. Basically, H19 could act by itself through binding to other proteins or RNAs and regulate their function, or through its derived miR-675 to repress other genes (Raveh et al., 2015; Ghafouri-Fard et al., 2020).

In CML tumorigenesis, the involvement of H19 has also been reported. H19 expression is BCR-ABL1 kinase-dependent. It is highly expressed in cell lines with transformed BCR-ABL1 or primary cells from CML patients (Guo et al., 2014). It has also been reported that H19 overexpression in CML patients is mediated by hypomethylation and associated with higher risk of disease progression (Zhou et al., 2018). Another study also indicated the regulation of H19 expression by DDX43 could facilitate CML pathogenesis *in vitro* and *in vivo* (Lin et al., 2018). Yet, the potential linkage of

H19/miR-675 and CML drug response has not been clearly revealed in existing research. Therefore, H19 along with its derived miR-675 became my first targets for the investigation into CML therapeutic resistance.

From my results, both H19 and miR-675-5p showed a significant downregulation upon imatinib treatment in K562 cells. This matches previous findings that H19 expression is BCR-ABL1 kinase-dependent (Guo et al., 2014), and H19 could function as primary miRNA transcript of miR-675 (Keniry et al., 2012; Zou et al., 2016). For miR-675-3p, it also showed a slight downregulation after imatinib treatment, but it is not that responsive when compared to miR-675-5p, hence subsequent investigation mainly focused on miR-675-5p.

To determine whether H19/miR-675 play a potential role of in CML drug resistance, imatinib-resistant K562 cells (K562-IMR) were generated by long-term drug treatment. It has been found that both H19 and miR-675-5p showed a higher relative expression in K562-IMR compared to K562-IMS (Figure 3.3.3). Such upregulated levels of H19 and miR-675-5p in IMR cells were also validated with another CML cell line LAMA84 and its IMR counterpart (Figure 3.3.3). My results, for the first time, showed the association between H19/miR-675 and imatinib resistance in CML cells. It was hypothesized that H19 or miR-675-5p could regulate the expression of certain genes, that finally results in a lower sensitivity of CML cells to imatinib. This conception led to the subsequent findings that is discussed in Chapter 4.5.

4.5 Identification of novel lncRNAs involved in CML therapeutic resistance

4.5.1 A downregulated novel lncRNA LNC000093 is identified in K562-IMR

With high throughput gene expression profiling by RNA-seq, differentially expressed lncRNAs, including unannotated ones, were targeted to identify novel regulators of imatinib resistance. As per previously mentioned in Chapter 1.6.3, one common interplay mechanism between lncRNA and miRNA is acting as ceRNAs (Chan and Tay, 2018b; Sun et al., 2020). In this study, I aimed to search for any novel lncRNA that may work together with miR-675-5p and form an axis to regulate the drug resistance in CML cells.

From the RNA-seq data, a significantly downregulated novel lncRNA was identified in K562-IMR and was named as LNC000093 (Figure 3.4.2A). It is located on chromosome 14q23.1 and is transcribed in anti-sense direction with two exons. Further RT-qPCR analysis validated its downregulation in K562-IMR compared to K562-IMS (Figure 3.4.2B), whereas an upregulation of LNC000093 was observed in TKI-sensitive CML patients (Figure 3.4.5). These collective findings indicated LNC000093 expression is positively associated with the sensitivity to imatinib in CML cases.

From the investigation of subcellular localization of LNC000093, my results showed the majority of LNC000093 is expressed in the cytoplasm and only small portion in the nucleus of K562 (Figure 3.4.4). This subcellular distribution pattern provides important clues for the molecular function of LNC000093 since lncRNAs play diverse roles in different subcellular compartments as mentioned in Chapter 1.6.2 & 1.6.3. In the nucleus, lncRNAs may function to regulate transcriptional process through chromatin

interactions and remodeling, or help establishing spatial organization of the nuclear compartment. In the cytoplasm, lncRNAs may function to mediate the transduction of signaling pathways, translational process, and posttranscriptional control of gene expression. For example, lncRNAs could sequester miRNAs and proteins to regulate their activity and levels, or mediate mRNA stability and translation, or influence the posttranslational modification of proteins. Therefore, the cytoplasmic localization of LNC000093 may imply its potential role in posttranscriptional or translational regulation.

4.5.2 Expression of LNC000093 is negatively regulated by H19/miR-675

Both H19 and its derived miR-675-5p were upregulated in imatinib-resistant (IMR) CML cells (Figures 3.3.3), while LNC000093 expression was downregulated (Figure 3.4.2). From the results of prediction analysis by StarMir, LNC000093 possesses 3 putative miRNA response elements (MREs) for miR-675-5p, which are all located close to the 3' end of its transcript (Figure 3.5.1). Subsequent luciferase reporter assay validated LNC000093 and miR-675-5p could interact with each other by direct binding through the seed sequence GGUGCG on miR-675-5p (Figure 3.5.3). This indicated that the expression of LNC000093 may have a high chance to be downregulated by miR-675-5p through RNA-induced silencing complex (RISC)-mediated degradation since the reduced luciferase activity should be due to the elimination of luciferase transcripts containing the LNC000093 sequence through miR-675-5p-mediated degradation process.

Silencing of H19 by siRNA and CRISPR-Cas9-mediated deletion of H19/miR-675 resulted in an upregulation of LNC000093 especially in IMR cells, while transfection of miR-675-5p mimic into H19/miR-675-deleted K562-IMR reversed the increase of LNC000093 expression (Figure 3.5.2D). This provided additional support that miR-675-5p could negatively regulate the expression of LNC000093. It is also noteworthy that the knockdown of miR-675-5p via H19/miR-675-deletion led to a higher increase of LNC000093 expression level in K562-IMR compared to K562-IMS, and this phenomenon matched with my hypothesis that LNC000093 could function as a ceRNA to compete for miR-675-5p since the abundance of ceRNA and miRNA in a ceRNA netwok could influence the interplay (Qi et al., 2015). In IMS cells, there is a higher expression of LNC000093 relative to IMR cells and miR-675-5p is then less abundant with reduced availability to mediate active ceRNA interplay, thereby a further reduction of miR-675-5p level would cause fewer effect on their target ceRNAs including LNC000093 when compared to IMR cells, which have relatively higher abundance of miR-675-5p.

4.5.3 LNC000093 acts as a ceRNA of RUNX1 to compete for miR-675-5p and regulate imatinib resistance

From the investigation of H19/miR-675-mediated cell death, H19 silencing by siRNA or CRISPR-Cas9-deletion reduced the cell viability of IMR CML cells upon imatinib treatment, and the effect was reversed by the ectopic expression of miR-675-5p mimic (Figure 3.6.1). Moreover, ectopic expression of LNC000093 increased K562-IMR cell death upon imatinib treatment in the presence or absence of miR-675-5p co-transfection (Figure 3.6.2). This indicated the expression change of miR-675-5p and LNC000093 could actively modulate the responsiveness of CML cells to imatinib rather than being a consequence of developing resistance, giving miR-675-5p promotes imatinib resistance whereas LNC000093 acts in opposite by interacting with miR-675-5p.

RUNX1 mutations and translocation commonly occur in different hematological disorders, including myelodysplastic syndrome, AML, ALL and CML, and some of these genetic abnormalities in *RUNX1* have been implicated in chemotherapeutic resistance (Sood et al., 2017). Some research studies also reported RUNX1 alterations may contribute to CML progression and persistence to imatinib treatment (Miething et al., 2007; Roche-Lestienne et al., 2008). A recent study also revealed that RUNX1 could play a direct role in cytotoxic drug response and resulting apoptotic activities (Speidel et al., 2017). In view of the above, RUNX1 was targeted for the examination of its potential to be involved in the interaction with H19/miR-675 and LNC000093.

In my study, western blot results showed a lower expression of RUNX1 protein in both K562-IMR and LAMA84-IMR cells compared to their IMS counterparts (Figure

3.6.3B), which implied RUNX1 may play a negative regulatory role in the imatinib resistance in CML so that IMR cells express less RUNX1. *In silico* prediction analysis revealed that *RUNX1* mRNA transcript possesses three putative MREs for miR-675-5p (Figure 3.6.3A), thus it can plausibly be a ceRNA partner of LNC000093 and compete for miR-675-5p. Subsequent experiments demonstrated RUNX1 expression in K562 cells could be significantly downregulated by transfection of miR-675-5p mimic (Figure 3.6.3C), and such repression of RUNX1 could be rescued by overexpression of LNC000093 (Figure 3.6.4). These collective findings demonstrated that in normal cells, LNC000093 serves as a ceRNA partner of *RUNX1* mRNA to consume the available pool of miR-675-5p and de-repress the RUNX1 expression mediated by H19/miR-675-5p, leading to a sensitive response to imatinib treatment. Accordingly, the low LNC000093 expression in IMR CML cells could result in more miR-675-5p to be released, which lead to non-responsiveness to imatinib via RUNX1 repression. To conclude, my results demonstrated the association between the LNC000093-H19/miR-675-RUNX1 axis and imatinib resistance.

4.6 Investigation of LNC000093 in differentiation of iPSCs

4.6.1 LNC000093 is actively involved in differentiation process

Apart from the potential regulatory role in therapeutic response, RUNX1 principally functions as an important transcription factor of hematopoietic lineage (Imperato et al., 2015). Since LNC000093 was identified in leukemic cell K562, which is hematopoietic lineage, and overexpression of exogenous LNC000093 was found to indirectly regulate the expression of RUNX1 via miR-675-5p by acting as a ceRNA, its potential regulatory role in hematopoietic differentiation was also investigated.

To achieve this, I made use of iPSCs to establish a differentiation model for the study as iPSCs are pluripotent in nature and can be induced to any differentiation lineages under specific culture condition. After induction of spontaneous differentiation as embryoid bodies (EBs) or directed to hematopoietic differentiation, the LNC000093 expression level was detected by RT-qPCR. Three marker genes that are known to be actively involved in early differentiation process towards hematopoietic lineage were also detected to monitor the differentiation status, which are CD34, GATA2 and CXCR4.

During differentiation through embryonic stage, the EBs develop into three primary germ layers including endoderm, mesoderm and ectoderm, where the mesoderm would further differentiate into hemangioblast and HSCs. CD34 is one of the important markers representing both endothelial and hematopoietic progenitor commitment (Sidney et al., 2014). GATA2 is a transcription factor regulating many genes that are critical for embryonic development and self-renewal especially those involved in

hematopoiesis. At mesodermal stage, GATA2 promotes the production of hemogenic endothelial progenitors and subsequent differentiation to hematopoietic progenitors by suppression of cardiac differentiation (Castano et al., 2019). CXCR4 is a chemokine receptor specific for stromal-derived-factor-1 (SDF-1) and the CXCR4/SDF-1 axis is important in chemotaxis and homing in immune and hematopoietic cells. It also plays a role during the differentiation at mesodermal stage to attenuate the endothelial potential and promote the induction of hematopoietic progenitor cells (Ahmed et al., 2016; Kawaguchi et al., 2019).

From my results, RT-qPCR analysis showed an upregulation for all the three marker genes in spontaneously differentiated EBs (Figure 3.7.2), indicating some portion of cells differentiated towards the early embryonic stage like mesoderm or hemangioblast. The expression of these genes was increased in a greater extent during directed hematopoietic differentiation for 7 days, and a further increment was observed in cells differentiated for 14 days, notably > 200-fold change in GATA2 expression (Figure 3.7.3). This reflected successful hematopoietic differentiation was induced from iPSCs under the specified culture medium.

For LNC000093, its expression was significantly upregulated during the process of iPSC differentiation (Figure 3.7.2), especially when the differentiation was directed to hematopoietic lineage with a longer period (Figure 3.7.3). LNC000093 expression increased further when the duration of differentiation was longer, thus, it means LNC000093 was upregulated along the course of differentiation process. As mentioned in Chapter 1.6.5, many lncRNA examples that are involved in cell differentiation would

be specifically expressed in corresponding cell type or lineage. My results implied that LNC000093 may be directly involved in the regulation of differentiation progress (e.g. regulating transcription of differentiation-related genes) or through indirect regulation on other cellular functions, such as self-renewal ability and proliferative capacity, which will be reduced during stem cell differentiation process (Seita and Weissman, 2010).

To ascertain the involvement of LNC000093 in general cellular differentiation, CRISPR-Cas9-mediated deletion of LNC000093 was done in iPSCs and the following impact on EB formation was examined. As directed differentiation of iPSC involves successive stages of development, and the great changes in gene expression profile or signaling pathways are complicated that may confound our preliminary findings, I started the functional study of LNC000093 with spontaneous differentiation model. Moreover, LNC000093 showed a greater expression change in EBs than that directed towards hematopoietic lineage on day 7 (Figure 3.7.2 & 3.7.3), hence it is expected to see a greater effect of LNC000093 silencing on EBs in a relatively short time course. A shorter-term day 3 model was also included to further reduce the variation caused by the random spontaneous differentiation as well as to determine the role of LNC000093 from the early stage of cell differentiation. My results showed the suppression of LNC000093 by CRISPR-Cas9-deletion led to a reduced extent of differentiation at early stage, as shown by the expression levels of CD34, GATA2 and CXCR4 (Figure 3.7.4) on day 3 and day 7 relative to the undifferentiated iPSC. This supported that LNC000093 plays a role in the regulation of general cell differentiation process at least in early stage, and a lack of LNC000093 could hamper the differentiation progress.

4.6.2 LNC000093 regulates differentiation potential of iPSC by alteration of chromatin accessibility

The homeostasis of stem cells is maintained through epigenetic mechanisms that regulate both the chromatin structure and specific gene transcription programs. Given that transcription factors play an essential role in determining cell fate during embryogenesis, a global epigenome remodeling occurs during early stem cell development in mammals, requiring commitment of cells to be restricted to a specific lineage (Stadhouders et al., 2019). Several epigenetic mechanisms, such as histone modifications, DNA methylation, and ATP-dependent chromatin remodeling, have been implicated in maintaining stem cell lineage commitment (Zhou et al., 2011). Basically, these epigenetic modifications would result in the change of DNA accessibility in order to allow specific transcription factors to bind on their target gene promotor regions adjacent to transcription start sites or any enhancer elements, leading to particular gene activation. Hence, the regulation of epigenome status through modification and accessibility of chromatin greatly influences early stem cell development (Xu and Xie, 2018).

During pluripotent cell differentiation, reconfiguration of chromatin structure occurs, such that euchromatic regions correlated to pluripotency genes become heterochromatin (“close” state), whereas those regions correlated to differentiation genes are transformed to euchromatic (“open” state). Hence, chromatin modifying factors that regulate the transition between euchromatin and heterochromatin play a crucial role in self-renewal and cell fate determination of stem cells (Meshorer and Misteli, 2006; Mattout and Meshorer, 2010; Xu and Xie, 2018). As my results demonstrated LNC000093 is generally involved in cell differentiation process, it was

hypothesized that the regulation is through the epigenetic changes and hence the transcriptional changes, which are important mechanisms that actively take place during stem cell differentiation as mentioned above.

The assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) is an advanced assay that could map the global chromatin accessibility at the genome level and allow the investigation of transcriptional regulation of gene expression (Buenrostro et al., 2015). In order to examine the potential role of LNC000093 in cell differentiation via regulating chromatin accessibility, scATAC-seq was performed with control and LNC000093-CRISPR-deleted iPSCs and EBs. My results revealed a reduction of chromatin peaks, which is the readout reflecting all the chromatin accessible regions, after LNC000093 was silenced by CRISPR-deletion in iPSCs (Figure 3.8.1). Further differential accessible peak analysis clearly showed a global change of chromatin accessibility with a large number of downregulated peaks (i.e. more inaccessible chromatin regions) in LNC000093-deleted cells relative to the control cells (Figure 3.8.4).

More importantly, by searching for the genes associated with these differential accessible peaks and examining their functional annotation, it was found that the genes related to the downregulated peaks in LNC000093-deleted cells are enriched in biological functions including cell differentiation, developmental and morphogenesis processes (Figure 3.8.5). This outcome totally matched with the previous observation that LNC000093 is responsible for differentiation process of iPSCs, and this time I showed the removal of LNC000093 would lead to the suppression of genes regulating

cell development and differentiation through transcriptional repression, as reflected by the chromatin accessibility profile. Gene score analysis of the scATAC-seq data estimated the differentially expressed genes with another algorithm that takes the accessibility of regulatory elements into account. Again, considerable number of genes were found to be downregulated after LNC000093-deletion in iPSCs.

Surprisingly, two genes, HIST1H3C and HIST1H2BB (also known as H3C3 and H2BC3, respectively) were significantly downregulated, and the downregulation was consistent in both iPSCs and EBs (Figure 3.8.6). Both HIST1H3C and HIST1H2BB encode variants of core histone proteins that are the basic nuclear proteins responsible for the construction of nucleosomes, the fundamental unit of chromatin. In eukaryotes, nucleosomes are composed of approximately 146 bp of DNA wrapped around a histone octamer that consists of a pair of each of the 4 core histones (H2A, H2B, H3, and H4). To create higher order chromatin structures, the linker histones H1 interact with DNA between nucleosomes to compact the chromatin fiber (Luger et al., 1997). My results showed that the expression of two genes responsible for specific histone variants could be altered by the interference of LNC000093 via transcriptional regulation as reflected by accessibility of regulatory elements.

The two histone encoding genes are located on chromosome 6 while LNC000093 is located on chromosome 14, thereby it is not likely that the regulation is through *cis*-acting mechanisms. As described in Chapter 1.6.2, a predominant regulatory mechanism of nucleus enriched lncRNAs is to interact with different chromatin modifying complexes or chromatin remodelers in order to enhance the modification of chromatin or histone, and alter the chromatin state that in turn modulate other gene

expression at epigenetic and transcriptional levels. Since the expression and regulation of lncRNA is temporal, spatial and cell- or tissue-type-specific (Engreitz et al., 2016), it is possible that a portion of LNC000093 expression become localized in nucleus during differentiation process of iPSCs and exert its regulatory function there. A good example is lncRNA lncMyoD, which has been found to regulate differentiation towards myogenic lineage by promoting the myogenic gene expression via modulation of MyoD accessibility to the chromatin. In their study, it has been demonstrated that lncMyoD is expressed in both cytoplasm and nucleus. However, the relative abundance in nuclear fraction increased when myoblasts were differentiating to myotubes, suggesting the nuclear regulatory role of lncMyoD during myogenic differentiation (Dong et al., 2020). Nevertheless, my previous results showed that LNC000093 could function as a miRNA sponge and a ceRNA, hence it is possible that LNC000093 is actually acting on the transcript of specific chromatin modifiers before their translation in cytoplasm, thereby indirectly influence the accessibility of regulatory region of the histone genes. Further experimental investigation is warranted to elucidate the details of such molecular mechanisms.

Indeed, previous research has demonstrated a significant difference of histone content between pluripotent and differentiated cells. The study revealed that spontaneous differentiation of embryonic stem cells (ESCs) to EBs entails an increase in histone content and the low histone content in ESCs would contribute to the maintenance of ESC totipotency and stemness (Karnavas et al., 2014). Therefore, my results showed the correlation between LNC000093 and cell differentiation is due to the regulation of cell pluripotency via controlling the expression of certain histone variants indirectly.

4.7 Translational perspectives of lncRNA research

LncRNAs have been emerging as important regulators in different pathological conditions and accumulating research evidence showed dysregulation of their expression is highly associated with the development of human disorders (Cipolla et al., 2018; Hu et al., 2018; Sanchez Calle et al., 2018). Such association conferred crucial clinical implications and provided great help with exploring novel options for diagnosis and therapy. Since lncRNAs regulate gene expression at an extra level in different physiological conditions, deciphering their molecular roles is essential for improving the understanding of human diseases and their treatment, which may significantly enhance the medical and clinical practice.

4.7.1 LncRNAs act as potential diagnostic and prognostic biomarkers

The expression patterns of lncRNAs are more cell/tissue type-specific or disease type-specific than protein-coding genes, which may make them become promising diagnostic or prognostic biological markers in clinical applications (Li and Chen, 2013). The high specificity of lncRNAs enables them to be a good predictor or indicator of disease stage, which is indicated by their abnormal expression compared to healthy individuals as reference. Currently, screening of biomarker in extracellular fluids from patients is one of the most favourable approaches for early diagnosis due to the non-invasive nature. It has been demonstrated that aberrant lncRNA expression correlated to tumorigenesis could also be detected in body fluids such as urine and blood plasma that are readily collected from patients, thereby makes lncRNA an ideal molecular biomarker (Bolha et al., 2017).

The lncRNA PCA3 (Prostate cancer antigen 3) is a well-known lncRNA example being

applied in clinical diagnostic tests as it showed superior performance than the traditional detection of prostate-specific antigen (PSA) serum level. The expression of PCA3 is highly prostate-specific and its overexpression is found in more than 90% of primary prostate tumors comparing with benign tissues, but is undetectable in other tumor types (Bussemakers et al., 1999; Hessels et al., 2003). The U.S. Food and Drug Administration (FDA) has approved the urine-based molecular diagnostic test to detect PCA3 in prostate cancer, which is now widely used (Groskopf et al., 2006; Lee et al., 2011).

In fact, lncRNAs could also be potential diagnostic and prognostic markers with regards to other human disorders. A subset of lncRNAs called BALR (B-ALL-associated long RNAs) has been revealed in clinical research by microarray analysis. The study has demonstrated that BALR levels allow the prediction of the cytogenetic subtypes of B-cell acute lymphoblastic leukemia (B-ALL) among the most frequently occurred abnormalities, including E2A-PBX1 translocation, mixed lineage leukemia (MLL) rearrangement, and TEL-AML1 fusion. Additional clinicopathologic data analysis demonstrated that high BALR-2 expression level was associated with poor survival of B-ALL patients and lower responses to prednisolone therapy (Fernando et al., 2015). This finding clearly shows the strong potential of lncRNA in sub-classification of disease and prediction of therapeutic response.

The lncRNA PRAL is another good example, which has downregulated expression in primary cells of multiple myeloma (MM), in particular, the MM cells with del(17p). Survival curve analysis showed a shorter overall survival in MM patients with low expression of PRAL (Xiao et al., 2018). The same study also showed PRAL is

potentially associated with bortezomib sensitivity through the interaction with miR-210. The collective findings implicated that PRAL expression levels could be used as a prognostic marker to predict the disease progression of MM patients and the effectiveness of bortezomib therapy.

In summary, the high expression specificity and relative ease of sampling make lncRNAs effective molecular biomarkers for diagnosis and prognosis of human diseases. Moreover, their expression levels could be readily detected by commonly used equipment and techniques such as real-time qPCR, microarray assay, and RNA-seq. It is anticipated that the combination of detecting disease-related lncRNAs and conventional biomarkers could greatly enhance the clinical judgment from medical professionals in future. Nevertheless, PCA3 is currently the only lncRNA that has been approved by the U.S. FDA and recommended as molecular marker at the moment. LncRNAs are still in their infancy as clinical biomarkers, and the applicability of many known disease-associated lncRNAs is uncertain yet. Hence, in order to build up the clinical use of lncRNAs, additional research should focus on how lncRNAs influence disease pathology, as well as standardizing the detection approach.

4.7.2 Development of therapeutic strategies by targeting lncRNA

Besides diagnostic and prognostic values, the unique specificity of lncRNAs also enables them to be conceivable therapeutic targets with potentially lower off-target effects. With the advancement in oligonucleotide-based therapeutic approached, any RNA transcripts that could not be targeted by antibody-drugs or small molecules have become possible to be reached by nucleic acid-based drugs (Nero et al., 2014; Beck et al., 2017). Particularly, antisense oligonucleotides (ASOs) have been tested in clinical

trials, and certain ASOs have already gained approval by the U.S. FDA (Gaudet et al., 2014; Noveck et al., 2014; Buller et al., 2015). This lncRNA-targeting approach is eventually emerging as a practicable way of therapy, which could provide more potential therapeutic options owing to their functional diversity.

A proof-of-concept study with animal models has demonstrated the therapeutic efficacy of lncRNA MALAT1 *in vivo*. In a MMTV-PyMT mouse model of breast cancer, increased cystic differentiation and cell adhesion, as well as reduced cell migration were shown after knockdown of MALAT1 using ASOs (Arun et al., 2016). The lncRNA Ube3a-ats is another example that has been tested in mouse model by the administration of ASOs, which provides a putative therapeutic target for Angelman syndrome (Meng et al., 2015). Even though this approach seems promising, clinical application is still quite challenging since efficient delivery and long-term effectiveness must be verified in human, thus, more preclinical and clinical studies in human subjects are needed. Nonetheless, given the rapid evolution of nucleic acid-targeting therapeutics in terms of improved pharmacokinetics and toxicity, it might be feasible to translate the identification and evaluation of disease-related lncRNAs into clinical applications very soon.

4.8 Limitation and suggestion for future work

In this study, the investigation of lncRNAs and related molecular networks in both Ph^+ MPNs and Ph-negative MPNs are virtually all based on *in vitro* experiments with the use of cell lines due to the inaccessibility of patient samples. A major limitation of such approach is that although cell lines possess the representative genetic abnormalities of the corresponding disease (i.e. *BCR-ABL1* in CML and *JAK2*-V617F mutation in classical MPNs), it may not represent the real situation occurring in patients, which could be much more complicated. In particular, HEL cells are even not derived from a MPN patient but rather a patient with erythroleukemia (Martin and Papayannopoulou, 1982). However, it has been chosen for the present study because I focused on investigating the disease-driven molecular pathway in order to gain insights for further exploration, and HEL cells showed a strong activation of JAK/STAT signaling due to *JAK2*-V617F, which is the most prevalent mutation across all classical MPNs. It would be better to include clinical samples in future studies to consolidate the findings and give more promising outcomes to show the value of lncRNA as a potential biomarker. Besides, all the cell lines used in the study were purchased from ATCC or DSMZ, which already checked for no mycoplasma contamination. However, there is a lack of routine checking of mycoplasma in our laboratory and therefore the effect of mycoplasma contamination on the cells may be overlooked if there is any. Another limitation in MPN study is that the establishment of potential ceRNA networks between BANCR or novel lncRNA with miR-3609 and JAK/STAT pathway genes were done by *in silico* analysis but not yet validated by hands-on experiments, hence additional studies are prospected in the future.

For the study of LNC000093, a brand-new observation of six-nucleotide polymorphisms existed in the RNA transcript (Figure 3.4.3) but no further investigation has been performed at the moment. It may be noteworthy to explore how the other variant transcript could be produced and what is the potential functional role of it. Moreover, the BCR-ABL1 mutation status of the in-house generated K562-IMR cells has not been checked. This raised a question that whether the observed LNC000093 action is totally independent of BCR-ABL1 abnormality, which merits additional follow-up investigation. Regarding the scATAC-seq experiment for LNC000093-CRISPR-deleted iPSCs, single-cell approach has been applied but technical issues were faced during data analysis, which is the incapacity to define and compare the different clusters with the current data, thus the cell population as a whole in each sample was used for comparison. It is suggested to perform further experiments together with single-cell RNA-seq (scRNA-seq), and the integration of scATAC-seq and scRNA-seq data could allow more comprehensive analysis in order to capture further information (Stuart et al., 2019; Forcato et al., 2021). Additional follow-up experiments to elucidate the regulatory mechanism of LNC000093 via interaction with chromatin remodeling factors is also prospected. Furthermore, apart from early differentiation stage, the potential regulatory role of LNC000093 in long-term differentiation especially in hematopoietic specific lineage could also be examined in the future since a further increase of LNC000093 expression was demonstrated when differentiation is directed towards hematopoietic lineage for 14 days (Figure 3.7.2). In addition, since the data are mostly generated using cell lines, it would be better to expand the clinical significance of the work by examining the expression levels of our target genes from public database to further support the correlation.

4.9 Concluding remarks

This thesis extensively explored the non-coding area, especially lncRNAs, in MPNs including *BCR-ABL1*-positive CML. The high-throughput profiling of *JAK2*-V617F MPN-like cells identified a number of novel and known lncRNA targets that are strongly associated with the mutant JAK2 signaling pathway, which warrant further examination. For the investigation of CML drug resistance, it has been shown that a higher level of H19/miR-675-5p promotes imatinib resistance in CML cells by regulating the expression of RUNX1, while the novel lncRNA LNC000093 acts as a ceRNA to compete for miR-675-5p. Further study has also demonstrated another regulatory role of LNC000093 in the pluripotency of iPSCs via modulation of chromatin accessibility.

In conclusion, my study examined MPNs in different perspectives regarding the molecular basis through *in vitro* models that possess distinct genetic features of Ph^+ or Ph-negative MPNs. It is anticipated that more clinically relevant translational research will be conducted in the near future, which could lead to the establishment of potential lncRNA biomarkers for MPNs. In addition, this study contributed to a greater understanding of lncRNA-miRNA-mRNA axes in hematological malignancies by highlighting the comprehensive crosstalks between lncRNAs and their diverse molecular targets.

References

Aboul-Soud, M.A.M., Alzahrani, A.J., and Mahmoud, A. (2021). Induced Pluripotent Stem Cells (iPSCs)-Roles in Regenerative Therapies, Disease Modelling and Drug Screening. *Cells* 10(9)**,** 2319. doi: 10.3390/cells10092319.

Ahmed, T., Tsuji-Tamura, K., and Ogawa, M. (2016). CXCR4 Signaling Negatively Modulates the Bipotential State of Hemogenic Endothelial Cells Derived from Embryonic Stem Cells by Attenuating the Endothelial Potential. *Stem Cells* 34(12)**,** 2814-2824. doi: 10.1002/stem.2441.

Al-Rugeebah, A., Alanazi, M., and Parine, N.R. (2019). MEG3: an Oncogenic Long Non-coding RNA in Different Cancers. *Pathol Oncol Res* 25(3)**,** 859-874. doi: 10.1007/s12253-019-00614-3.

Alexandrova, E.M., and Moll, U.M. (2012). Role of p53 family members p73 and p63 in human hematological malignancies. *Leukemia & lymphoma* 53(11)**,** 2116-2129.

Alvarez-Dominguez, J.R., Hu, W., Gromatzky, A.A., and Lodish, H.F. (2014). Long noncoding RNAs during normal and malignant hematopoiesis. *International journal of hematology* 99(5)**,** 531-541.

Ambros, V. (2001). microRNAs: tiny regulators with great potential. *Cell* 107(7)**,** 823-826. doi: 10.1016/s0092-8674(01)00616-x.

An, X., Tiwari, A.K., Sun, Y., Ding, P.-R., Ashby Jr, C.R., and Chen, Z.-S. (2010). BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. *Leukemia research* 34(10)**,** 1255-1268.

Anderson, L.A., and McMullin, M.F. (2014). Epidemiology of MPN: what do we know? *Curr Hematol Malig Rep* 9(4)**,** 340-349. doi: 10.1007/s11899-014-0228-z.

Andrews, S. (2010). "FastQC: a quality control tool for high throughput sequence data.".

Araki, M., Yang, Y., Masubuchi, N., Hironaka, Y., Takei, H., Morishita, S., et al. (2016). Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. *Blood* 127(10)**,** 1307-1316. doi: 10.1182/blood-2015-09-671172.

Arun, G., Diermeier, S., Akerman, M., Chang, K.C., Wilkinson, J.E., Hearn, S., et al. (2016). Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev* 30(1)**,** 34-51. doi: 10.1101/gad.270959.115.

Balligand, T., Achouri, Y., Pecquet, C., Chachoua, I., Nivarthi, H., Marty, C., et al. (2016). Pathologic activation of thrombopoietin receptor and JAK2-STAT5 pathway by frameshift mutants of mouse calreticulin. *Leukemia* 30(8)**,** 1775-1778. doi: 10.1038/leu.2016.47.

Barbui, T., Barosi, G., Birgegard, G., Cervantes, F., Finazzi, G., Griesshammer, M., et al. (2011). Philadelphia-negative classical myeloproliferative neoplasms: critical concepts and management recommendations from European LeukemiaNet. *J Clin Oncol* 29(6)**,** 761-770. doi: 10.1200/JCO.2010.31.8436.

Barbui, T., Thiele, J., Gisslinger, H., Kvasnicka, H.M., Vannucchi, A.M., Guglielmelli, P., et al. (2018). The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. *Blood Cancer J* 8(2)**,** 15. doi: 10.1038/s41408-018-0054-y.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2)**,** 281-297. doi: 10.1016/s0092-8674(04)00045-5.

Batista, P.J., and Chang, H.Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. *Cell* 152(6)**,** 1298-1307.

Beck, A., Goetsch, L., Dumontet, C., and Corvaia, N. (2017). Strategies and challenges for the next generation of antibody-drug conjugates. *Nat Rev Drug Discov* 16(5)**,** 315-337. doi: 10.1038/nrd.2016.268.

Bewry, N.N., Nair, R.R., Emmons, M.F., Boulware, D., Pinilla-Ibarz, J., and Hazlehurst, L.A. (2008). Stat3 contributes to resistance toward BCR-ABL inhibitors in a bone marrow

microenvironment model of drug resistance. *Mol Cancer Ther* 7(10)**,** 3169-3175. doi: 10.1158/1535-7163.MCT-08-0314.

Bhat, A.A., Younes, S.N., Raza, S.S., Zarif, L., Nisar, S., Ahmed, I., et al. (2020). Role of non-coding RNA networks in leukemia progression, metastasis and drug resistance. *Mol Cancer* 19(1)**,** 57. doi: 10.1186/s12943-020-01175-9.

Bixby, D., and Talpaz, M. (2009). Mechanisms of resistance to tyrosine kinase inhibitors in chronic myeloid leukemia and recent therapeutic strategies to overcome resistance. *Hematology Am Soc Hematol Educ Program* 2009(1)**,** 461-476. doi: 10.1182/asheducation-2009.1.461.

Bolha, L., Ravnik-Glavac, M., and Glavac, D. (2017). Long Noncoding RNAs as Biomarkers in Cancer. *Dis Markers* 2017**,** 7243968. doi: 10.1155/2017/7243968.

Bousard, A., Raposo, A.C., Zylicz, J.J., Picard, C., Pires, V.B., Qi, Y., et al. (2019). The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep* 20(10)**,** e48019. doi: 10.15252/embr.201948019.

Bozkurt, S., Ozkan, T., Ozmen, F., Baran, Y., Sunguroglu, A., and Kansu, E. (2013). The roles of epigenetic modifications of proapoptotic BID and BIM genes in imatinib-resistant chronic myeloid leukemia cells. *Hematology* 18(4)**,** 217-223. doi: 10.1179/1607845412Y.0000000056.

Braconi, C., Kogure, T., Valeri, N., Huang, N., Nuovo, G., Costinean, S., et al. (2011). microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene* 30(47)**,** 4750-4756.

Braganca, J., Lopes, J.A., Mendes-Silva, L., and Almeida Santos, J.M. (2019). Induced pluripotent stem cells, a giant leap for mankind therapeutic applications. *World J Stem Cells* 11(7)**,** 421-430. doi: 10.4252/wjsc.v11.i7.421.

Branford, S., Wang, P., Yeung, D.T., Thomson, D., Purins, A., Wadham, C., et al. (2018). Integrative genomic analysis reveals cancer-associated mutations at diagnosis of CML in patients with high-risk disease. *Blood* 132(9)**,** 948-961. doi: 10.1182/blood-2018-02-832253.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. *PLoS Biol* 3(3)**,** e85. doi: 10.1371/journal.pbio.0030085.

Buchdunger, E., Zimmermann, J., Mett, H., Meyer, T., Muller, M., Druker, B.J., et al. (1996). Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative. *Cancer Res* 56(1)**,** 100-104.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* 109**,** 21 29 21-21 29 29. doi: 10.1002/0471142727.mb2129s109.

Buller, H.R., Bethune, C., Bhanot, S., Gailani, D., Monia, B.P., Raskob, G.E., et al. (2015). Factor XI antisense oligonucleotide for prevention of venous thrombosis. *N Engl J Med* 372(3)**,** 232-240. doi: 10.1056/NEJMoa1405760.

Burchert, A., Wang, Y., Cai, D., von Bubnoff, N., Paschka, P., Muller-Brusselbach, S., et al. (2005). Compensatory PI3-kinase/Akt/mTor activation regulates imatinib resistance development. *Leukemia* 19(10)**,** 1774-1782. doi: 10.1038/sj.leu.2403898.

Bussemakers, M.J., van Bokhoven, A., Verhaegh, G.W., Smit, F.P., Karthaus, H.F., Schalken, J.A., et al. (1999). DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res* 59(23)**,** 5975-5979.

Byun, J.M., Kim, Y.J., Youk, T., Yang, J.J., Yoo, J., and Park, T.S. (2017). Real world epidemiology of myeloproliferative neoplasms: a population based study in Korea 2004-2013. *Ann Hematol* 96(3)**,** 373-381. doi: 10.1007/s00277-016-2902-9.

Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., et al. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* 16(1)**,** 20. doi: 10.1186/s13059-015-0586-4.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global

properties and specific subclasses. *Genes Dev* 25(18)**,** 1915-1927. doi: 10.1101/gad.17446611.

Cai, X., and Cullen, B.R. (2007). The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* 13(3)**,** 313-316. doi: 10.1261/rna.351707.

Cancer Genome Atlas Research, N., Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., et al. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368(22)**,** 2059-2074. doi: 10.1056/NEJMoa1301689.

Cao, L., Zhang, P., Li, J., and Wu, M. (2017). LAST, a c-Myc-inducible long noncoding RNA, cooperates with CNBP to promote CCND1 mRNA stability in human cells. *Elife* 6**,** e30433. doi: 10.7554/eLife.30433.

Carette, J.E., Pruszak, J., Varadarajan, M., Blomen, V.A., Gokhale, S., Camargo, F.D., et al. (2010). Generation of iPSCs from cultured human malignant cells. *Blood* 115(20)**,** 4039-4042. doi: 10.1182/blood-2009-07-231845.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491(7424)**,** 454-457. doi: 10.1038/nature11508.

Castano, J., Aranda, S., Bueno, C., Calero-Nieto, F.J., Mejia-Ramirez, E., Mosquera, J.L., et al. (2019). GATA2 Promotes Hematopoietic Development and Represses Cardiac Differentiation of Human Mesoderm. *Stem Cell Reports* 13(3)**,** 515-529. doi: 10.1016/j.stemcr.2019.07.009.

Cervantes, F., Dupriez, B., Passamonti, F., Vannucchi, A.M., Morra, E., Reilly, J.T., et al. (2012). Improving survival trends in primary myelofibrosis: an international study. *J Clin Oncol* 30(24)**,** 2981-2987. doi: 10.1200/JCO.2012.42.0240.

Cervantes, F., Dupriez, B., Pereira, A., Passamonti, F., Reilly, J.T., Morra, E., et al. (2009). New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood* 113(13)**,** 2895-2901. doi: 10.1182/blood-2008-07-170449.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., et al. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147(2)**,** 358-369. doi: 10.1016/j.cell.2011.09.028.

Chachoua, I., Pecquet, C., El-Khoury, M., Nivarthi, H., Albu, R.I., Marty, C., et al. (2016). Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood* 127(10)**,** 1325-1335. doi: 10.1182/blood-2015-11-681932.

Chan, J.J., and Tay, Y. (2018a). Noncoding RNA:RNA Regulatory Networks in Cancer. *Int J Mol Sci* 19(5)**,** 1310. doi: 10.3390/ijms19051310.

Chan, J.J., and Tay, Y. (2018b). Noncoding RNA:RNA Regulatory Networks in Cancer. *Int J Mol Sci* 19(5). doi: 10.3390/ijms19051310.

Chen, J., Wang, Y., Wang, C., Hu, J.F., and Li, W. (2020). LncRNA Functions as a New Emerging Epigenetic Factor in Determining the Fate of Stem Cells. *Front Genet* 11(277)**,** 277. doi: 10.3389/fgene.2020.00277.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17)**,** i884-i890. doi: 10.1093/bioinformatics/bty560.

Cho, Y.J., Cunnick, J.M., Yi, S.J., Kaartinen, V., Groffen, J., and Heisterkamp, N. (2007). Abr and Bcr, two homologous Rac GTPase-activating proteins, control multiple cellular functions of murine macrophages. *Mol Cell Biol* 27(3)**,** 899-911. doi: 10.1128/MCB.00756-06.

Ciarlo, E., Massone, S., Penna, I., Nizzari, M., Gigoni, A., Dieci, G., et al. (2013). An intronic ncRNA-dependent regulation of SORL1 expression affecting Abeta formation is upregulated in post-mortem Alzheimer's disease brain samples. *Dis Model Mech* 6(2)**,** 424-433. doi: 10.1242/dmm.009761.

Cilloni, D., and Saglio, G. (2012). Molecular pathways: BCR-ABL. *Clin Cancer Res* 18(4)**,** 930-937. doi: 10.1158/1078-0432.CCR-10-1613.

Cipolla, G., de Oliveira, J., Salviano-Silva, A., Lobo-Alves, S., Lemos, D., Oliveira, L., et al. (2018). Long non-coding RNAs in multifactorial diseases: another layer of complexity. *Noncoding RNA* 4(2)**,** 13.

Clark, S.S., McLaughlin, J., Crist, W.M., Champlin, R., and Witte, O.N. (1987). Unique forms of the abl tyrosine kinase distinguish Ph1-positive CML from Ph1-positive ALL. *Science* 235(4784)**,** 85-88. doi: 10.1126/science.3541203.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., et al. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 33(6)**,** 717-726. doi: 10.1016/j.molcel.2009.01.026.

Cohen, G.B., Ren, R., and Baltimore, D. (1995). Modular binding domains in signal transduction proteins. *Cell* 80(2)**,** 237-248. doi: 10.1016/0092-8674(95)90406-9.

Cohen, M.H., Williams, G., Johnson, J.R., Duan, J., Gobburu, J., Rahman, A., et al. (2002). Approval summary for imatinib mesylate capsules in the treatment of chronic myelogenous leukemia. *Clin Cancer Res* 8(5)**,** 935-942.

Colognori, D., Sunwoo, H., Kriz, A.J., Wang, C.Y., and Lee, J.T. (2019). Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Mol Cell* 74(1)**,** 101-117 e110. doi: 10.1016/j.molcel.2019.01.015.

Cortez, D., Reuther, G., and Pendergast, A.M. (1997). The Bcr-Abl tyrosine kinase activates mitogenic signaling pathways and stimulates G1-to-S phase transition in hematopoietic cells. *Oncogene* 15(19)**,** 2333-2342. doi: 10.1038/sj.onc.1201400.

Crea, F., Paolicchi, E., Marquez, V.E., and Danesi, R. (2012). Polycomb genes and cancer: time for clinical application? *Crit Rev Oncol Hematol* 83(2)**,** 184-193. doi: 10.1016/j.critrevonc.2011.10.007.

Crisa, E., Venturino, E., Passera, R., Prina, M., Schinco, P., Borchiellini, A., et al. (2010). A retrospective study on 226 polycythemia vera patients: impact of median hematocrit value on clinical outcomes and survival improvement with anti-thrombotic prophylaxis and non-alkylating drugs. *Ann Hematol* 89(7)**,** 691-699. doi: 10.1007/s00277-009-0899-z.

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., et al. (2019). Ensembl 2019. *Nucleic Acids Res* 47(D1)**,** D745-D751. doi: 10.1093/nar/gky1113.

De Klein, A., Hagemeijer, A., Bartram, C.R., Houwen, R., Hoefsloot, L., Carbonell, F., et al. (1986). bcr rearrangement and translocation of the c-abl oncogene in Philadelphia positive acute lymphoblastic leukemia. *Blood* 68(6)**,** 1369-1375.

Defour, J.P., Chachoua, I., Pecquet, C., and Constantinescu, S.N. (2016). Oncogenic activation of MPL/thrombopoietin receptor by 17 mutations at W515: implications for myeloproliferative neoplasms. *Leukemia* 30(5)**,** 1214-1216. doi: 10.1038/leu.2015.271.

Deininger, M.W., Goldman, J.M., and Melo, J.V. (2000). The molecular biology of chronic myeloid leukemia. *Blood* 96(10)**,** 3343-3356.

Denhardt, D.T. (1996). Signal-transducing protein phosphorylation cascades mediated by Ras/Rho proteins in the mammalian cell: the potential for multiplex signalling. *Biochem J* 318 (Pt 3)**,** 729-747. doi: 10.1042/bj3180729.

Desnuelle, C., Dib, M., Garrel, C., and Favier, A. (2001). A double-blind, placebo-controlled randomized clinical trial of alpha-tocopherol (vitamin E) in the treatment of amyotrophic lateral sclerosis. ALS riluzole-tocopherol Study Group. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2(1)**,** 9-18. doi: 10.1080/146608201300079364.

Diekmann, D., Brill, S., Garrett, M.D., Totty, N., Hsuan, J., Monfries, C., et al. (1991). Bcr encodes a GTPase-activating protein for p21rac. *Nature* 351(6325)**,** 400-402. doi: 10.1038/351400a0.

Ding, J., Komatsu, H., Wakita, A., Kato-Uranishi, M., Ito, M., Satoh, A., et al. (2004). Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin. *Blood* 103(11)**,** 4198-4200. doi: 10.1182/blood-2003-10-3471.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. *Genes Dev* 18(5)**,** 504-511. doi: 10.1101/gad.1184404.

Dong, A., Preusch, C.B., So, W.K., Lin, K., Luan, S., Yi, R., et al. (2020). A long noncoding RNA, LncMyoD, modulates chromatin accessibility to regulate muscle stem cell myogenic lineage progression. *Proc Natl Acad Sci U S A* 117(51)**,** 32464-32475. doi: 10.1073/pnas.2005868117.

Druker, B.J., and Lydon, N.B. (2000). Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest* 105(1)**,** 3-7. doi: 10.1172/JCI9083.

Dubash, A.D., Koetsier, J.L., Amargo, E.V., Najor, N.A., Harmon, R.M., and Green, K.J. (2013). The GEF Bcr activates RhoA/MAL signaling to promote keratinocyte differentiation via desmoglein-1. *J Cell Biol* 202(4)**,** 653-666. doi: 10.1083/jcb.201304133.

Economides, M.P., Verstovsek, S., and Pemmaraju, N. (2019). Novel Therapies in Myeloproliferative Neoplasms (MPN): Beyond JAK Inhibitors. *Curr Hematol Malig Rep* 14(5)**,** 460-468. doi: 10.1007/s11899-019-00538-4.

Eiring, A.M., Page, B.D.G., Kraft, I.L., Mason, C.C., Vellore, N.A., Resetca, D., et al. (2015). Combined STAT3 and BCR-ABL1 inhibition induces synthetic lethality in therapy-resistant chronic myeloid leukemia. *Leukemia* 29(3)**,** 586-597. doi: 10.1038/leu.2014.245.

Elf, S., Abdelfattah, N.S., Chen, E., Perales-Paton, J., Rosen, E.A., Ko, A., et al. (2016). Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation. *Cancer Discov* 6(4)**,** 368-381. doi: 10.1158/2159-8290.CD-15-1434.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* 17(12)**,** 756-770. doi: 10.1038/nrm.2016.126.

Fallah, P., Amirizadeh, N., Poopak, B., Toogeh, G., Arefian, E., Kohram, F., et al. (2015). Expression pattern of key microRNAs in patients with newly diagnosed chronic myeloid leukemia in chronic phase. *Int J Lab Hematol* 37(4)**,** 560-568. doi: 10.1111/ijlh.12351.

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., et al. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res* 46(D1)**,** D308-D314. doi: 10.1093/nar/gkx1107.

Fatica, A., and Bozzoni, I. (2014a). Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15(1)**,** 7-21. doi: 10.1038/nrg3606.

Fatica, A., and Bozzoni, I. (2014b). Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics* 15(1)**,** 7.

Faust, T., Frankel, A., and D'Orso, I. (2012). Transcription control by long non-coding RNAs. *Transcription* 3(2)**,** 78-86. doi: 10.4161/trns.19349.

Feller, S.M., Knudsen, B., and Hanafusa, H. (1994). c-Abl kinase regulates the protein binding activity of c-Crk. *EMBO J* 13(10)**,** 2341-2351.

Fernandez Tde, S., de Souza Fernandez, C., and Mencalha, A.L. (2013). Human induced pluripotent stem cells from basic research to potential clinical applications in cancer. *Biomed Res Int* 2013**,** 430290. doi: 10.1155/2013/430290.

Fernando, T.R., Rodriguez-Malave, N.I., Waters, E.V., Yan, W., Casero, D., Basso, G., et al. (2015). LncRNA Expression Discriminates Karyotype and Predicts Survival in B-Lymphoblastic Leukemia. *Mol Cancer Res* 13(5)**,** 839-851. doi: 10.1158/1541-7786.MCR-15-0006-T.

Forcato, M., Romano, O., and Bicciato, S. (2021). Computational methods for the integrative analysis of single-cell data. *Brief Bioinform* 22(1)**,** 20-29. doi: 10.1093/bib/bbaa042.

Garding, A., Bhattacharya, N., Claus, R., Ruppel, M., Tschuch, C., Filarsky, K., et al. (2013). Epigenetic upregulation of lncRNAs at 13q14.3 in leukemia is linked to the In Cis downregulation of a gene cluster that targets NF-kB. *PLoS Genet* 9(4)**,** e1003373. doi: 10.1371/journal.pgen.1003373.

Garitano-Trojaola, A., Agirre, X., Prósper, F., and Fortes, P. (2013). Long non-coding RNAs in haematological malignancies. *International journal of molecular sciences* 14(8)**,** 15386-15422.

Gaspar, J.M. (2018). Improved peak-calling with MACS2. *bioRxiv***,** 496521. doi: 10.1101/496521.

Gaudet, D., Brisson, D., Tremblay, K., Alexander, V.J., Singleton, W., Hughes, S.G., et al. (2014). Targeting APOC3 in the familial chylomicronemia syndrome. *N Engl J Med* 371(23)**,** 2200-2206. doi: 10.1056/NEJMoa1400284.

Geisler, S., and Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 14(11)**,** 699-712. doi: 10.1038/nrm3679.

Geyer, H.L., and Mesa, R.A. (2014). Therapy for myeloproliferative neoplasms: when, which agent, and how? *Blood* 124(24)**,** 3529-3537. doi: 10.1182/blood-2014-05-577635.

Geyer, J.T., and Orazi, A. (2016). Myeloproliferative neoplasms (BCR-ABL1 negative) and myelodysplastic/myeloproliferative neoplasms: current diagnostic principles and upcoming updates. *Int J Lab Hematol* 38 Suppl 1**,** 12-19. doi: 10.1111/ijlh.12509.

Ghafouri-Fard, S., Esmaeili, M., and Taheri, M. (2020). H19 lncRNA: Roles in tumorigenesis. *Biomed Pharmacother* 123**,** 109774. doi: 10.1016/j.biopha.2019.109774.

Giotopoulos, G., van der Weyden, L., Osaki, H., Rust, A.G., Gallipoli, P., Meduri, E., et al. (2015). A novel mouse model identifies cooperating mutations and therapeutic targets critical for chronic myeloid leukemia progression. *J Exp Med* 212(10)**,** 1551-1569. doi: 10.1084/jem.20141661.

Girodon, F., Dutrillaux, F., Broseus, J., Mounier, M., Goussot, V., Bardonnaud, P., et al. (2010). Leukocytosis is associated with poor survival but not with increased risk of thrombosis in essential thrombocythemia: a population-based study of 311 patients. *Leukemia* 24(4)**,** 900-903. doi: 10.1038/leu.2010.5.

Glickman, M.H., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* 82(2)**,** 373-428. doi: 10.1152/physrev.00027.2001.

Gong, C., Li, Z., Ramanujan, K., Clay, I., Zhang, Y., Lemire-Brachat, S., et al. (2015). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev Cell* 34(2)**,** 181-191. doi: 10.1016/j.devcel.2015.05.009.

Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470(7333)**,** 284-288. doi: 10.1038/nature09701.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., et al. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* 53(3)**,** 403-411. doi: 10.1038/s41588-021-00790-6.

Grinfeld, J., Nangalia, J., and Green, A.R. (2017). Molecular determinants of pathogenesis and clinical phenotype in myeloproliferative neoplasms. *Haematologica* 102(1)**,** 7-17. doi: 10.3324/haematol.2014.113845.

Groskopf, J., Aubin, S.M., Deras, I.L., Blase, A., Bodrug, S., Clark, C., et al. (2006). APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. *Clin Chem* 52(6)**,** 1089-1095. doi: 10.1373/clinchem.2005.063289.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Wahrisch, S., Beisaw, A., et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* 24(2)**,** 206-214. doi: 10.1016/j.devcel.2012.12.012.

Guo, G., Kang, Q., Chen, Q., Chen, Z., Wang, J., Tan, L., et al. (2014). High expression of long non-coding RNA H19 is required for efficient tumorigenesis induced by Bcr-Abl oncogene. *FEBS Lett* 588(9)**,** 1780-1786. doi: 10.1016/j.febslet.2014.03.038.

Han, P., and Chang, C.P. (2015). Long non-coding RNA and chromatin remodeling. *RNA Biol* 12(10)**,** 1094-1098. doi: 10.1080/15476286.2015.1063770.

Hanfstein, B., Muller, M.C., Hehlmann, R., Erben, P., Lauseker, M., Fabarius, A., et al. (2012). Early molecular and cytogenetic response is predictive for long-term progression-free

and overall survival in chronic myeloid leukemia (CML). *Leukemia* 26(9)**,** 2096-2102. doi: 10.1038/leu.2012.85.

Hanly, D.J., Esteller, M., and Berdasco, M. (2018). Interplay between long non-coding RNAs and epigenetic machinery: emerging targets in cancer? *Philos Trans R Soc Lond B Biol Sci* 373(1748)**,** 20170074. doi: 10.1098/rstb.2017.0074.

Harrison, C., Kiladjian, J.J., Al-Ali, H.K., Gisslinger, H., Waltzman, R., Stalbovskaya, V., et al. (2012). JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N Engl J Med* 366(9)**,** 787-798. doi: 10.1056/NEJMoa1110556.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22(9)**,** 1760-1774. doi: 10.1101/gr.135350.111.

He, B.M., Bai, Y., Kang, W., Zhang, X.P., and Jiang, X.J. (2017). LncRNA SNHG5 regulates imatinib resistance in chronic myeloid leukemia via acting as a CeRNA against MiR-205-5p. *Am J Cancer Res* 7(8)**,** 1704-1713.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5(7)**,** 522-531. doi: 10.1038/nrg1379.

Hellmann, A. (1992). [Molecular biology of chronic myeloid leukemia]. *Acta Haematol Pol* 23(2 Suppl 1)**,** 13-17.

Hessels, D., Klein Gunnewiek, J.M., van Oort, I., Karthaus, H.F., van Leenders, G.J., van Balken, B., et al. (2003). DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur Urol* 44(1)**,** 8-15; discussion 15-16. doi: 10.1016/s0302-2838(03)00201-x.

Ho, R., Chronis, C., and Plath, K. (2011). Mechanistic insights into reprogramming to induced pluripotency. *J Cell Physiol* 226(4)**,** 868-878. doi: 10.1002/jcp.22450.

Hochhaus, A., Kreil, S., Corbin, A.S., La Rosee, P., Muller, M.C., Lahaye, T., et al. (2002). Molecular and chromosomal mechanisms of resistance to imatinib (STI571) therapy. *Leukemia* 16(11)**,** 2190-2196. doi: 10.1038/sj.leu.2402741.

Hochhaus, A., Larson, R.A., Guilhot, F., Radich, J.P., Branford, S., Hughes, T.P., et al. (2017). Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia. *N Engl J Med* 376(10)**,** 917-927. doi: 10.1056/NEJMoa1609324.

Hochhaus, A., O'Brien, S.G., Guilhot, F., Druker, B.J., Branford, S., Foroni, L., et al. (2009). Six-year follow-up of patients receiving imatinib for the first-line treatment of chronic myeloid leukemia. *Leukemia* 23(6)**,** 1054-1061. doi: 10.1038/leu.2009.38.

Hu, G., Lou, Z., and Gupta, M. (2014). The long non-coding RNA GAS5 cooperates with the eukaryotic translation initiation factor 4E to regulate c-Myc translation. *PLoS One* 9(9)**,** e107016. doi: 10.1371/journal.pone.0107016.

Hu, G., Niu, F., Humburg, B.A., Liao, K., Bendi, S., Callen, S., et al. (2018). Molecular mechanisms of long noncoding RNAs and their role in disease pathogenesis. *Oncotarget* 9(26)**,** 18648-18663. doi: 10.18632/oncotarget.24307.

Huang, X., Cortes, J., and Kantarjian, H. (2012). Estimations of the increasing prevalence and plateau prevalence of chronic myeloid leukemia in the era of tyrosine kinase inhibitor therapy. *Cancer* 118(12)**,** 3123-3127. doi: 10.1002/cncr.26679.

Hughes, J.M., Legnini, I., Salvatori, B., Masciarelli, S., Marchioni, M., Fazi, F., et al. (2015). C/EBPalpha-p30 protein induces expression of the oncogenic long non-coding RNA UCA1 in acute myeloid leukemia. *Oncotarget* 6(21)**,** 18534-18544. doi: 10.18632/oncotarget.4069.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 43(7)**,** 621-629. doi: 10.1038/ng.848.

Ilaria, R.L., Jr., and Van Etten, R.A. (1996). P210 and P190(BCR/ABL) induce the tyrosine phosphorylation and DNA binding activity of multiple specific STAT family members. *J Biol Chem* 271(49)**,** 31704-31710. doi: 10.1074/jbc.271.49.31704.

Imai-Sumida, M., Dasgupta, P., Kulkarni, P., Shiina, M., Hashimoto, Y., Shahryari, V., et al. (2020). Genistein Represses HOTAIR/Chromatin Remodeling Pathways to Suppress Kidney Cancer. *Cell Physiol Biochem* 54(1)**,** 53-70. doi: 10.33594/000000205.

Imperato, M.R., Cauchy, P., Obier, N., and Bonifer, C. (2015). The RUNX1-PU.1 axis in the control of hematopoiesis. *Int J Hematol* 101(4)**,** 319-329. doi: 10.1007/s12185-015-1762-8.

Isin, M., and Dalay, N. (2015). LncRNAs and neoplasia. *Clinica Chimica Acta* 444**,** 280-288.

Jabbour, E., and Kantarjian, H. (2020). Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. *Am J Hematol* 95(6)**,** 691-709. doi: 10.1002/ajh.25792.

James, C., Ugo, V., Le Couedic, J.P., Staerk, J., Delhommeau, F., Lacout, C., et al. (2005). A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 434(7037)**,** 1144-1148. doi: 10.1038/nature03546.

Jang, M.A., and Choi, C.W. (2020). Recent insights regarding the molecular basis of myeloproliferative neoplasms. *Korean J Intern Med* 35(1)**,** 1-11. doi: 10.3904/kjim.2019.317.

Javidi-Sharifi, N., and Hobbs, G. (2021). Future Directions in Chronic Phase CML Treatment. *Curr Hematol Malig Rep* 16(6)**,** 500-508. doi: 10.1007/s11899-021-00658-w.

Jia, R., and Kralovics, R. (2020). Progress in elucidation of molecular pathophysiology of myeloproliferative neoplasms and its application to therapeutic decisions. *Int J Hematol* 111(2)**,** 182-191. doi: 10.1007/s12185-019-02778-9.

Kang, Y.J., Yang, D.C., Kong, L., Hou, M., Meng, Y.Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res* 45(W1)**,** W12-W16. doi: 10.1093/nar/gkx428.

Karnavas, T., Pintonello, L., Agresti, A., and Bianchi, M.E. (2014). Histone content increases in differentiating embryonic stem cells. *Front Physiol* 5**,** 330. doi: 10.3389/fphys.2014.00330.

Kawaguchi, N., Zhang, T.T., and Nakanishi, T. (2019). Involvement of CXCR4 in Normal and Abnormal Development. *Cells* 8(2)**,** 185. doi: 10.3390/cells8020185.

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., et al. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol* 14(7)**,** 659-665. doi: 10.1038/ncb2521.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12(4)**,** 357-360. doi: 10.1038/nmeth.3317.

Kim, T., Tyndel, M.S., Zhang, Z., Ahn, J., Choi, S., Szardenings, M., et al. (2017). Exome sequencing reveals DNMT3A and ASXL1 variants associate with progression of chronic myeloid leukemia after tyrosine kinase inhibitor therapy. *Leuk Res* 59**,** 142-148. doi: 10.1016/j.leukres.2017.06.009.

Kim, Y., Rim, Y.A., Yi, H., Park, N., Park, S.H., and Ju, J.H. (2016). The Generation of Human Induced Pluripotent Stem Cells from Blood Cells: An Efficient Protocol Using Serial Plating of Reprogrammed Cells by Centrifugation. *Stem Cells Int* 2016**,** 1329459. doi: 10.1155/2016/1329459.

Kindler, S., Wang, H., Richter, D., and Tiedge, H. (2005). RNA transport and local control of translation. *Annu Rev Cell Dev Biol* 21**,** 223-245. doi: 10.1146/annurev.cellbio.21.122303.120653.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3(107)**,** ra8. doi: 10.1126/scisignal.2000568.

Kipreos, E.T., and Wang, J.Y. (1992). Cell cycle-regulated binding of c-Abl tyrosine kinase to DNA. *Science* 256(5055)**,** 382-385. doi: 10.1126/science.256.5055.382.

Klampfl, T., Gisslinger, H., Harutyunyan, A.S., Nivarthi, H., Rumi, E., Milosevic, J.D., et al. (2013). Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med* 369(25)**,** 2379-2390. doi: 10.1056/NEJMoa1311347.

Kralovics, R., Passamonti, F., Buser, A.S., Teo, S.S., Tiedt, R., Passweg, J.R., et al. (2005). A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 352(17)**,** 1779-1790. doi: 10.1056/NEJMoa051113.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493(7431)**,** 231-235. doi: 10.1038/nature11661.

Laneuville, P. (1995). Abl tyrosine protein kinase. *Semin Immunol* 7(4)**,** 255-266. doi: 10.1006/smim.1995.0030.

Lanz, R.B., Chua, S.S., Barron, N., Soder, B.M., DeMayo, F., and O'Malley, B.W. (2003). Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo. *Mol Cell Biol* 23(20)**,** 7163-7176. doi: 10.1128/MCB.23.20.7163-7176.2003.

Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., et al. (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* 8(1)**,** 106-118. doi: 10.1016/j.stem.2010.12.003.

Lee, G.L., Dobi, A., and Srivastava, S. (2011). Prostate cancer: diagnostic performance of the PCA3 urine test. *Nat Rev Urol* 8(3)**,** 123-124. doi: 10.1038/nrurol.2011.10.

Levine, R.L., Pardanani, A., Tefferi, A., and Gilliland, D.G. (2007). Role of JAK2 in the pathogenesis and therapy of myeloproliferative disorders. *Nat Rev Cancer* 7(9)**,** 673-683. doi: 10.1038/nrc2210.

Levine, R.L., Wadleigh, M., Cools, J., Ebert, B.L., Wernig, G., Huntly, B.J., et al. (2005). Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 7(4)**,** 387-397. doi: 10.1016/j.ccr.2005.03.023.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120(1)**,** 15-20. doi: 10.1016/j.cell.2004.12.035.

Li, C.H., and Chen, Y. (2013). Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol* 45(8)**,** 1895-1910. doi: 10.1016/j.biocel.2013.05.030.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, Q., Huang, Z., Gao, M., Cao, W., Xiao, Q., Luo, H., et al. (2017). Blockade of Y177 and Nuclear Translocation of Bcr-Abl Inhibits Proliferation and Promotes Apoptosis in Chronic Myeloid Leukemia Cells. *Int J Mol Sci* 18(3). doi: 10.3390/ijms18030537.

Li, Z., Yang, L., Liu, X., Nie, Z., and Luo, J. (2018a). Long noncoding RNA MEG3 inhibits proliferation of chronic myeloid leukemia cells by sponging microRNA21. *Biomed Pharmacother* 104**,** 181-192. doi: 10.1016/j.biopha.2018.05.047.

Li, Z.Y., Yang, L., Liu, X.J., Wang, X.Z., Pan, Y.X., and Luo, J.M. (2018b). The Long Noncoding RNA MEG3 and its Target miR-147 Regulate JAK/STAT Pathway in Advanced Chronic Myeloid Leukemia. *EBioMedicine* 34**,** 61-75. doi: 10.1016/j.ebiom.2018.07.013.

Liang, W.C., Fu, W.M., Wang, Y.B., Sun, Y.X., Xu, L.L., Wong, C.W., et al. (2016). H19 activates Wnt signaling and promotes osteoblast differentiation by functioning as a competing endogenous RNA. *Sci Rep* 6(1)**,** 20121. doi: 10.1038/srep20121.

Lim, S.N., Lee, J.H., Lee, J.H., Kim, D.Y., Kim, S.D., Kang, Y.A., et al. (2013). Allogeneic hematopoietic cell transplantation in adult patients with myelodysplastic/myeloproliferative neoplasms. *Blood Res* 48(3)**,** 178-184. doi: 10.5045/br.2013.48.3.178.

Lin, J., Ma, J.C., Yang, J., Yin, J.Y., Chen, X.X., Guo, H., et al. (2018). Arresting of miR-186 and releasing of H19 by DDX43 facilitate tumorigenesis and CML progression. *Oncogene* 37(18)**,** 2432-2443. doi: 10.1038/s41388-018-0146-y.

Lin, N., Chang, K.Y., Li, Z., Gates, K., Rana, Z.A., Dang, J., et al. (2014). An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell* 53(6)**,** 1005-1019. doi: 10.1016/j.molcel.2014.01.021.

Liu, C., Oikonomopoulos, A., Sayed, N., and Wu, J.C. (2018). Modeling human diseases with induced pluripotent stem cells: from 2D to 3D and beyond. *Development* 145(5). doi: 10.1242/dev.156166.

Liu, Y., Liu, X., Lin, C., Jia, X., Zhu, H., Song, J., et al. (2021). Noncoding RNAs regulate alternative splicing in Cancer. *J Exp Clin Cancer Res* 40(1)**,** 11. doi: 10.1186/s13046-020-01798-2.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* 417(1)**,** 15-27. doi: 10.1042/BJ20081501.

Lopotová, T., }â ková, M., Klamová, H., and Moravcová, J. (2011). MicroRNA-451 in chronic myeloid leukemia: miR-451-BCR-ABL regulatory loop? *Leukemia research* 35 7**,** 974-977.

Loscocco, F., Visani, G., Galimberti, S., Curti, A., and Isidori, A. (2019). BCR-ABL Independent Mechanisms of Resistance in Chronic Myeloid Leukemia. *Front Oncol* 9(939)**,** 939. doi: 10.3389/fonc.2019.00939.

Loscocco, F., Visani, G., and Isidori, A. (2018). ENL YEATS domain: targeting the acute myeloid leukemia epigenome. *Biotarget* 2.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12)**,** 550. doi: 10.1186/s13059-014-0550-8.

Lu, K.H., Li, W., Liu, X.H., Sun, M., Zhang, M.L., Wu, W.Q., et al. (2013). Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. *BMC Cancer* 13**,** 461. doi: 10.1186/1471-2407-13-461.

Lu, X., Huang, L.J., and Lodish, H.F. (2008). Dimerization by a cytokine receptor is necessary for constitutive activation of JAK2V617F. *J Biol Chem* 283(9)**,** 5258-5266. doi: 10.1074/jbc.M707125200.

Lu, X., Levine, R., Tong, W., Wernig, G., Pikman, Y., Zarnegar, S., et al. (2005). Expression of a homodimeric type I cytokine receptor is required for JAK2V617F-mediated transformation. *Proc Natl Acad Sci U S A* 102(52)**,** 18962-18967. doi: 10.1073/pnas.0509714102.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389(6648)**,** 251-260. doi: 10.1038/38444.

Lundberg, P., Karow, A., Nienhold, R., Looser, R., Hao-Shen, H., Nissen, I., et al. (2014). Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* 123(14)**,** 2220-2228. doi: 10.1182/blood-2013-11-537167.

Ma, L., Shan, Y., Bai, R., Xue, L., Eide, C.A., Ou, J., et al. (2014). A therapeutically targetable mechanism of BCR-ABL-independent imatinib resistance in chronic myeloid leukemia. *Sci Transl Med* 6(252)**,** 252ra121. doi: 10.1126/scitranslmed.3009073.

Malik, N., and Rao, M.S. (2013). "A Review of the Methods for Human iPSC Derivation," in *Pluripotent Stem Cells: Methods and Protocols*, eds. U. Lakshmipathy & M.C. Vemuri. (Totowa, NJ: Humana Press), 23-33.

Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol* 13(1)**,** 95-101. doi: 10.1038/ncb2140.

Martin, P., and Papayannopoulou, T. (1982). HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. *Science* 216(4551)**,** 1233-1235. doi: 10.1126/science.6177045.

Marty, C., Pecquet, C., Nivarthi, H., El-Khoury, M., Chachoua, I., Tulliez, M., et al. (2016). Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. *Blood* 127(10)**,** 1317-1324. doi: 10.1182/blood-2015-11-679571.

Mathy, N.W., and Chen, X.M. (2017). Long non-coding RNAs (lncRNAs) and their transcriptional control of inflammatory responses. *J Biol Chem* 292(30)**,** 12375-12382. doi: 10.1074/jbc.R116.760884.

Mattout, A., and Meshorer, E. (2010). Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Curr Opin Cell Biol* 22(3)**,** 334-341. doi: 10.1016/j.ceb.2010.02.001.

McWhirter, J.R., Galasso, D.L., and Wang, J.Y. (1993). A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. *Mol Cell Biol* 13(12)**,** 7587-7595. doi: 10.1128/mcb.13.12.7587-7595.1993.

McWhirter, J.R., and Wang, J.Y. (1993). An actin-binding function contributes to transformation by the Bcr-Abl oncoprotein of Philadelphia chromosome-positive human leukemias. *EMBO J* 12(4)**,** 1533-1546.
Melo, J.V. (1996). The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood* 88(7)**,** 2375-2384.
Meng, L., Ward, A.J., Chun, S., Bennett, C.F., Beaudet, A.L., and Rigo, F. (2015). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 518(7539)**,** 409-412. doi: 10.1038/nature13975.
Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* 7(7)**,** 540-546. doi: 10.1038/nrm1938.
Miao, Y., Ajami, N.E., Huang, T.S., Lin, F.M., Lou, C.H., Wang, Y.T., et al. (2018). Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. *Nat Commun* 9(1)**,** 292. doi: 10.1038/s41467-017-02113-y.
Michalak, M., Groenendyk, J., Szabo, E., Gold, L.I., and Opas, M. (2009). Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. *Biochem J* 417(3)**,** 651-666. doi: 10.1042/BJ20081847.
Michlewski, G., and Caceres, J.F. (2019). Post-transcriptional control of miRNA biogenesis. *RNA* 25(1)**,** 1-16. doi: 10.1261/rna.068692.118.
Miething, C., Grundler, R., Mugler, C., Brero, S., Hoepfl, J., Geigl, J., et al. (2007). Retroviral insertional mutagenesis identifies RUNX genes involved in chronic myeloid leukemia disease persistence under imatinib treatment. *Proc Natl Acad Sci U S A* 104(11)**,** 4594-4599. doi: 10.1073/pnas.0604716104.
Mistry, J., Bateman, A., and Finn, R.D. (2007). Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics* 8(1)**,** 298. doi: 10.1186/1471-2105-8-298.
Nair, R.R., Tolentino, J.H., Argilagos, R.F., Zhang, L., Pinilla-Ibarz, J., and Hazlehurst, L.A. (2012). Potentiation of Nilotinib-mediated cell death in the context of the bone marrow microenvironment requires a promiscuous JAK inhibitor in CML. *Leuk Res* 36(6)**,** 756-763. doi: 10.1016/j.leukres.2011.12.002.
Nangalia, J., Massie, C.E., Baxter, E.J., Nice, F.L., Gundem, G., Wedge, D.C., et al. (2013). Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 369(25)**,** 2391-2405. doi: 10.1056/NEJMoa1312542.
Neguembor, M.V., Jothi, M., and Gabellini, D. (2014). Long noncoding RNAs, emerging players in muscle differentiation and disease. *Skelet Muscle* 4(1)**,** 8. doi: 10.1186/2044-5040-4-8.
Nero, T.L., Morton, C.J., Holien, J.K., Wielens, J., and Parker, M.W. (2014). Oncogenic protein interfaces: small molecules, big challenges. *Nat Rev Cancer* 14(4)**,** 248-262. doi: 10.1038/nrc3690.
Nishizawa, M., Chonabayashi, K., Nomura, M., Tanaka, A., Nakamura, M., Inagaki, A., et al. (2016). Epigenetic Variation between Human Induced Pluripotent Stem Cell Lines Is an Indicator of Differentiation Capacity. *Cell Stem Cell* 19(3)**,** 341-354. doi: 10.1016/j.stem.2016.06.019.
Nivarthi, H., Chen, D., Cleary, C., Kubesova, B., Jager, R., Bogner, E., et al. (2016). Thrombopoietin receptor is required for the oncogenic function of CALR mutants. *Leukemia* 30(8)**,** 1759-1763. doi: 10.1038/leu.2016.32.
Noh, J.H., Kim, K.M., McClusky, W.G., Abdelmohsen, K., and Gorospe, M. (2018). Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip Rev RNA* 9(3)**,** e1471. doi: 10.1002/wrna.1471.
Noveck, R., Stroes, E.S., Flaim, J.D., Baker, B.F., Hughes, S., Graham, M.J., et al. (2014). Effects of an antisense oligonucleotide inhibitor of C-reactive protein synthesis on the endotoxin challenge response in healthy human male volunteers. *J Am Heart Assoc* 3(4)**,** e001084. doi: 10.1161/JAHA.114.001084.
Nowell, P.C. (1985). Citation Classic - a Minute Chromosome in Human Chronic Granulocytic-Leukemia. *Current Contents/Life Sciences* (8)**,** 19-19.
O'Brien, S.G., Guilhot, F., Larson, R.A., Gathmann, I., Baccarani, M., Cervantes, F., et al. (2003). Imatinib compared with interferon and low-dose cytarabine for newly

diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med* 348(11)**,** 994-1004. doi: 10.1056/NEJMoa022457.

Pagano, F., Comoglio, F., Grinfeld, J., Li, J., Godfrey, A., Baxter, J., et al. (2018). MicroRNA-101 expression is associated with JAK2V617F activity and regulates JAK2/STAT5 signaling. *Leukemia* 32(8)**,** 1826-1830. doi: 10.1038/s41375-018-0053-9.

Pane, F., Frigeri, F., Sindona, M., Luciano, L., Ferrara, F., Cimino, R., et al. (1996). Neutrophilic-chronic myeloid leukemia: a distinct disease with a specific molecular marker (BCR/ABL with C3/A2 junction). *Blood* 88(7)**,** 2410-2414.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 374(23)**,** 2209-2221. doi: 10.1056/NEJMoa1516192.

Pardanani, A.D., Levine, R.L., Lasho, T., Pikman, Y., Mesa, R.A., Wadleigh, M., et al. (2006). MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood* 108(10)**,** 3472-3476. doi: 10.1182/blood-2006-04-018879.

Pasquier, F., Cabagnols, X., Secardin, L., Plo, I., and Vainchenker, W. (2014). Myeloproliferative neoplasms: JAK2 signaling pathway as a central target for therapy. *Clin Lymphoma Myeloma Leuk* 14 Suppl**,** S23-35. doi: 10.1016/j.clml.2014.06.014.

Passamonti, F., Elena, C., Schnittger, S., Skoda, R.C., Green, A.R., Girodon, F., et al. (2011). Molecular and clinical features of the myeloproliferative neoplasm associated with JAK2 exon 12 mutations. *Blood* 117(10)**,** 2813-2816. doi: 10.1182/blood-2010-11-316810.

Passamonti, F., Rumi, E., Pietra, D., Elena, C., Boveri, E., Arcaini, L., et al. (2010). A prospective study of 338 patients with polycythemia vera: the impact of JAK2 (V617F) allele burden and leukocytosis on fibrotic or leukemic disease transformation and vascular complications. *Leukemia* 24(9)**,** 1574-1579. doi: 10.1038/leu.2010.148.

Pendergast, A.M., Muller, A.J., Havlik, M.H., Maru, Y., and Witte, O.N. (1991). BCR sequences essential for transformation by the BCR-ABL oncogene bind to the ABL SH2 regulatory domain in a non-phosphotyrosine-dependent manner. *Cell* 66(1)**,** 161-171. doi: 10.1016/0092-8674(91)90148-r.

Pendergast, A.M., Quilliam, L.A., Cripe, L.D., Bassing, C.H., Dai, Z., Li, N., et al. (1993). BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. *Cell* 75(1)**,** 175-185.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33(3)**,** 290-295. doi: 10.1038/nbt.3122.

Pietra, D., Rumi, E., Ferretti, V.V., Di Buduo, C.A., Milanesi, C., Cavalloni, C., et al. (2016). Differential clinical effects of different mutation subtypes in CALR-mutant myeloproliferative neoplasms. *Leukemia* 30(2)**,** 431-438. doi: 10.1038/leu.2015.277.

Pikman, Y., Lee, B.H., Mercher, T., McDowell, E., Ebert, B.L., Gozo, M., et al. (2006). MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med* 3(7)**,** e270. doi: 10.1371/journal.pmed.0030270.

Pizzatti, L., Binato, R., Cofre, J., Gomes, B.E., Dobbin, J., Haussmann, M.E., et al. (2010). SUZ12 is a candidate target of the non-canonical WNT pathway in the progression of chronic myeloid leukemia. *Genes Chromosomes Cancer* 49(2)**,** 107-118. doi: 10.1002/gcc.20722.

Qi, X., Zhang, D.H., Wu, N., Xiao, J.H., Wang, X., and Ma, W. (2015). ceRNA in cancer: possible functions and clinical implications. *J Med Genet* 52(10)**,** 710-718. doi: 10.1136/jmedgenet-2015-103334.

Quentmeier, H., MacLeod, R.A., Zaborski, M., and Drexler, H.G. (2006). JAK2 V617F tyrosine kinase mutation in cell lines derived from myeloproliferative disorders. *Leukemia* 20(3)**,** 471-476. doi: 10.1038/sj.leu.2404081.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6)**,** 841-842. doi: 10.1093/bioinformatics/btq033.

Raedler, L.A. (2015). Jakafi (Ruxolitinib): First FDA-Approved Medication for the Treatment of Patients with Polycythemia Vera. *Am Health Drug Benefits* 8(Spec Feature)**,** 75-79.

Rampal, R., Ahn, J., Abdel-Wahab, O., Nahas, M., Wang, K., Lipson, D., et al. (2014). Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. *Proc Natl Acad Sci U S A* 111(50)**,** E5401-5410. doi: 10.1073/pnas.1407792111.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47(W1)**,** W191-W198. doi: 10.1093/nar/gkz369.

Raveh, E., Matouk, I.J., Gilon, M., and Hochberg, A. (2015). The H19 Long non-coding RNA in cancer initiation, progression and metastasis - a proposed unifying theory. *Mol Cancer* 14**,** 184. doi: 10.1186/s12943-015-0458-2.

Rea, D., Mauro, M.J., Boquimpani, C., Minami, Y., Lomaia, E., Voloshin, S., et al. (2021). A Phase 3, Open-Label, Randomized Study of Asciminib, a STAMP Inhibitor, vs Bosutinib in CML After >/=2 Prior TKIs. *Blood*. doi: 10.1182/blood.2020009984.

Rennie, W., Liu, C., Carmack, C.S., Wolenc, A., Kanoria, S., Lu, J., et al. (2014). STarMir: a web server for prediction of microRNA binding sites. *Nucleic Acids Res* 42(Web Server issue)**,** W114-118. doi: 10.1093/nar/gku376.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81**,** 145-166. doi: 10.1146/annurev-biochem-051410-092902.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *cell* 129(7)**,** 1311-1323.

Roche-Lestienne, C., Deluche, L., Corm, S., Tigaud, I., Joha, S., Philippe, N., et al. (2008). RUNX1 DNA-binding mutations and RUNX1-PRDM16 cryptic fusions in BCR-ABL+ leukemias are frequently associated with secondary trisomy 21 and may contribute to clonal evolution and imatinib resistance. *Blood* 111(7)**,** 3735-3741. doi: 10.1182/blood-2007-07-102533.

Rodriguez-Malave, N.I., Fernando, T.R., Patel, P.C., Contreras, J.R., Palanichamy, J.K., Tran, T.M., et al. (2015). BALR-6 regulates cell growth and cell survival in B-lymphoblastic leukemia. *Mol Cancer* 14**,** 214. doi: 10.1186/s12943-015-0485-z.

Rokah, O.H., Granot, G., Ovcharenko, A., Modai, S., Pasmanik-Chor, M., Toren, A., et al. (2012). Downregulation of miR-31, miR-155, and miR-564 in chronic myeloid leukemia cells. *PLoS One* 7(4)**,** e35501. doi: 10.1371/journal.pone.0035501.

Rosti, G., Castagnetti, F., Gugliotta, G., and Baccarani, M. (2017). Tyrosine kinase inhibitors in chronic myeloid leukaemia: which, when, for whom? *Nat Rev Clin Oncol* 14(3)**,** 141-154. doi: 10.1038/nrclinonc.2016.139.

Rowley, J.D. (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 243(5405)**,** 290-293. doi: 10.1038/243290a0.

Rumi, E., and Cazzola, M. (2017). Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms. *Blood* 129(6)**,** 680-692. doi: 10.1182/blood-2016-10-695957.

Rumi, E., Pietra, D., Ferretti, V., Klampfl, T., Harutyunyan, A.S., Milosevic, J.D., et al. (2014a). JAK2 or CALR mutation status defines subtypes of essential thrombocythemia with substantially different clinical course and outcomes. *Blood* 123(10)**,** 1544-1551. doi: 10.1182/blood-2013-11-539098.

Rumi, E., Pietra, D., Pascutto, C., Guglielmelli, P., Martinez-Trillos, A., Casetti, I., et al. (2014b). Clinical effect of driver mutations of JAK2, CALR, or MPL in primary myelofibrosis. *Blood* 124(7)**,** 1062-1069. doi: 10.1182/blood-2014-05-578435.

Saha, K., and Jaenisch, R. (2009). Technical challenges in using human induced pluripotent stem cells to model disease. *Cell Stem Cell* 5(6)**,** 584-595. doi: 10.1016/j.stem.2009.11.009.

Saharinen, P., Takaluoma, K., and Silvennoinen, O. (2000). Regulation of the Jak2 tyrosine kinase by its pseudokinase domain. *Mol Cell Biol* 20(10)**,** 3387-3395. doi: 10.1128/mcb.20.10.3387-3395.2000.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146(3)**,** 353-358. doi: 10.1016/j.cell.2011.07.014.

Sanchez Calle, A., Kawamura, Y., Yamamoto, Y., Takeshita, F., and Ochiya, T. (2018). Emerging roles of long non-coding RNA in cancer. *Cancer Sci* 109(7)**,** 2093-2100. doi: 10.1111/cas.13642.

Sangkhae, V., Etheridge, S.L., Kaushansky, K., and Hitchcock, I.S. (2014). The thrombopoietin receptor, MPL, is critical for development of a JAK2V617F-induced myeloproliferative neoplasm. *Blood* 124(26)**,** 3956-3963. doi: 10.1182/blood-2014-07-587238.

Sattari, A., Siddiqui, H., Moshiri, F., Ngankeu, A., Nakamura, T., Kipps, T.J., et al. (2016). Upregulation of long noncoding RNA MIAT in aggressive form of chronic lymphocytic leukemias. *Oncotarget* 7(34)**,** 54174-54182. doi: 10.18632/oncotarget.11099.

Sattler, M., Mohi, M.G., Pride, Y.B., Quinnan, L.R., Malouf, N.A., Podar, K., et al. (2002). Critical role for Gab2 in transformation by BCR/ABL. *Cancer Cell* 1(5)**,** 479-492. doi: 10.1016/s1535-6108(02)00074-0.

Sawyers, C.L. (1999). Chronic myeloid leukemia. *N Engl J Med* 340(17)**,** 1330-1340. doi: 10.1056/NEJM199904293401706.

Sayed, N., Liu, C., and Wu, J.C. (2016). Translation of Human-Induced Pluripotent Stem Cells: From Clinical Trial in a Dish to Precision Medicine. *J Am Coll Cardiol* 67(18)**,** 2161-2176. doi: 10.1016/j.jacc.2016.01.083.

Sayed, N., and Wu, J.C. (2017). Towards Cardio-Precision medicine. *Eur Heart J* 38(14)**,** 1014-1016. doi: 10.1093/eurheartj/ehx089.

Schaukowitch, K., and Kim, T.-K. (2014). Emerging epigenetic mechanisms of long non-coding RNAs. *Neuroscience* 264**,** 25-38.

Scholl, V., Hassan, R., and Zalcberg, I.R. (2012). miRNA-451: A putative predictor marker of Imatinib therapy response in chronic myeloid leukemia. *Leuk Res* 36(1)**,** 119-121. doi: 10.1016/j.leukres.2011.08.023.

Scott, L.M., Tong, W., Levine, R.L., Scott, M.A., Beer, P.A., Stratton, M.R., et al. (2007). JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med* 356(5)**,** 459-468. doi: 10.1056/NEJMoa065202.

Scott, M.T., Korfi, K., Saffrey, P., Hopcroft, L.E., Kinstrie, R., Pellicano, F., et al. (2016). Epigenetic Reprogramming Sensitizes CML Stem Cells to Combined EZH2 and Tyrosine Kinase Inhibition. *Cancer Discov* 6(11)**,** 1248-1257. doi: 10.1158/2159-8290.CD-16-0263.

Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med* 2(6)**,** 640-653. doi: 10.1002/wsbm.86.

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., et al. (2010). Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* 7(1)**,** 11-14. doi: 10.1016/j.stem.2010.06.003.

Sessarego, M., Pasquali, F., Bianchi Scarra, G.L., and Ajmar, F. (1983). Masked Philadelphia chromosome caused by translocation (9;11;22). *Cancer Genet Cytogenet* 8(4)**,** 319-323. doi: 10.1016/0165-4608(83)90074-2.

Shah, N.P. (2008). Advanced CML: therapeutic options for patients in accelerated and blast phases. *J Natl Compr Canc Netw* 6 Suppl 2**,** S31-S36.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11)**,** 2498-2504. doi: 10.1101/gr.1239303.

Shefner, J.M., Cudkowicz, M.E., Schoenfeld, D., Conrad, T., Taft, J., Chilton, M., et al. (2004). A clinical trial of creatine in ALS. *Neurology* 63(9)**,** 1656-1661. doi: 10.1212/01.wnl.0000142992.81995.f0.

Shi, X., Sun, M., Liu, H., Yao, Y., and Song, Y. (2013). Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer letters* 339(2)**,** 159-166.

Sidney, L.E., Branch, M.J., Dunphy, S.E., Dua, H.S., and Hopkinson, A. (2014). Concise review: evidence for CD34 as a common marker for diverse progenitors. *Stem Cells* 32(6)**,** 1380-1389. doi: 10.1002/stem.1661.

Silver, R.T., Woolf, S.H., Hehlmann, R., Appelbaum, F.R., Anderson, J., Bennett, C., et al. (1999). An evidence-based analysis of the effect of busulfan, hydroxyurea, interferon, and allogeneic bone marrow transplantation in treating the chronic phase of chronic myeloid leukemia: developed for the American Society of Hematology. *Blood* 94(5)**,** 1517-1536.

Singh, V.K., Kalsan, M., Kumar, N., Saini, A., and Chandra, R. (2015a). Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. *Front Cell Dev Biol* 3(2)**,** 2. doi: 10.3389/fcell.2015.00002.

Singh, V.K., Kalsan, M., Kumar, N., Saini, A., and Chandra, R. (2015b). Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. *Front Cell Dev Biol* 3**,** 2. doi: 10.3389/fcell.2015.00002.

Skorski, T., Kanakaraj, P., Nieborowska-Skorska, M., Ratajczak, M.Z., Wen, S.C., Zon, G., et al. (1995). Phosphatidylinositol-3 kinase activity is regulated by BCR/ABL and is required for the growth of Philadelphia chromosome-positive cells. *Blood* 86(2)**,** 726-736.

Soltani, I., Gharbi, H., Hassine, I.B., Bouguerra, G., Douzi, K., Teber, M., et al. (2017). Regulatory network analysis of microRNAs and genes in imatinib-resistant chronic myeloid leukemia. *Funct Integr Genomics* 17(2-3)**,** 263-277. doi: 10.1007/s10142-016-0520-1.

Sood, R., Kamikubo, Y., and Liu, P. (2017). Role of RUNX1 in hematological malignancies. *Blood* 129(15)**,** 2070-2082. doi: 10.1182/blood-2016-10-687830.

Soverini, S., Mancini, M., Bavaro, L., Cavo, M., and Martinelli, G. (2018). Chronic myeloid leukemia: the paradigm of targeting oncogenic tyrosine kinase signaling and counteracting resistance for successful cancer therapy. *Mol Cancer* 17(1)**,** 49. doi: 10.1186/s12943-018-0780-6.

Speidel, D., Wellbrock, J., and Abas, M. (2017). RUNX1 Upregulation by Cytotoxic Drugs Promotes Apoptosis. *Cancer Res* 77(24)**,** 6818-6824. doi: 10.1158/0008-5472.CAN-17-0319.

Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* 569(7756)**,** 345-354. doi: 10.1038/s41586-019-1182-7.

Staerk, J., Lacout, C., Sato, T., Smith, S.O., Vainchenker, W., and Constantinescu, S.N. (2006). An amphipathic motif at the transmembrane-cytoplasmic junction prevents autonomous activation of the thrombopoietin receptor. *Blood* 107(5)**,** 1864-1871. doi: 10.1182/blood-2005-06-2600.

Stam, K., Heisterkamp, N., Grosveld, G., de Klein, A., Verma, R.S., Coleman, M., et al. (1985). Evidence of a new chimeric bcr/c-abl mRNA in patients with chronic myelocytic leukemia and the Philadelphia chromosome. *N Engl J Med* 313(23)**,** 1429-1433. doi: 10.1056/NEJM198512053132301.

Stegelmann, F., Bullinger, L., Griesshammer, M., Holzmann, K., Habdank, M., Kuhn, S., et al. (2010). High-resolution single-nucleotide polymorphism array-profiling in myeloproliferative neoplasms identifies novel genomic aberrations. *Haematologica* 95(4)**,** 666-669. doi: 10.3324/haematol.2009.013623.

Stetka, J., Vyhlidalova, P., Lanikova, L., Koralkova, P., Gursky, J., Hlusi, A., et al. (2019). Addiction to DUSP1 protects JAK2V617F-driven polycythemia vera progenitors against inflammatory stress and DNA damage, allowing chronic proliferation. *Oncogene* 38(28)**,** 5627-5642. doi: 10.1038/s41388-019-0813-7.

Strassler, E.T., Aalto-Setala, K., Kiamehr, M., Landmesser, U., and Krankel, N. (2018). Age Is Relative-Impact of Donor Age on Induced Pluripotent Stem Cell-Derived Cell Functionality. *Front Cardiovasc Med* 5**,** 4. doi: 10.3389/fcvm.2018.00004.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, et al. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177(7)**,** 1888-1902 e1821. doi: 10.1016/j.cell.2019.05.031.

Sun, B., Liu, C., Li, H., Zhang, L., Luo, G., Liang, S., et al. (2020). Research progress on the interactions between long non-coding RNAs and microRNAs in human cancer. *Oncol Lett* 19(1)**,** 595-605. doi: 10.3892/ol.2019.11182.

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* 41(17)**,** e166. doi: 10.1093/nar/gkt646.

Sun, Q., Hao, Q., and Prasanth, K.V. (2018). Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. *Trends Genet* 34(2)**,** 142-157. doi: 10.1016/j.tig.2017.11.005.

Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* 19(3)**,** 347-359. doi: 10.1101/gr.087775.108.

Sweet, K., Hazlehurst, L., Sahakian, E., Powers, J., Nodzon, L., Kayali, F., et al. (2018). A phase I clinical trial of ruxolitinib in combination with nilotinib in chronic myeloid leukemia patients with molecular evidence of disease. *Leuk Res* 74**,** 89-96. doi: 10.1016/j.leukres.2018.10.002.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4)**,** 663-676. doi: 10.1016/j.cell.2006.07.024.

Takahashi, K., and Yamanaka, S. (2013). Induced pluripotent stem cells in medicine and biology. *Development* 140(12)**,** 2457-2461. doi: 10.1242/dev.092551.

Talati, C., and Pinilla-Ibarz, J. (2018). Resistance in chronic myeloid leukemia: definitions and novel therapeutic agents. *Curr Opin Hematol* 25(2)**,** 154-161. doi: 10.1097/MOH.0000000000000403.

Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* 12(4)**,** e1004873. doi: 10.1371/journal.pcbi.1004873.

Tefferi, A., Guglielmelli, P., Larson, D.R., Finke, C., Wassie, E.A., Pieri, L., et al. (2014). Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood* 124(16)**,** 2507-2513; quiz 2615. doi: 10.1182/blood-2014-05-579136.

Tefferi, A., Rumi, E., Finazzi, G., Gisslinger, H., Vannucchi, A.M., Rodeghiero, F., et al. (2013). Survival and prognosis among 1545 patients with contemporary polycythemia vera: an international study. *Leukemia* 27(9)**,** 1874-1881. doi: 10.1038/leu.2013.163.

Tefferi, A., Skoda, R., and Vardiman, J.W. (2009). Myeloproliferative neoplasms: contemporary diagnosis using histology and genetics. *Nat Rev Clin Oncol* 6(11)**,** 627-637. doi: 10.1038/nrclinonc.2009.149.

Thomas, J., Wang, L., Clark, R.E., and Pirmohamed, M. (2004). Active transport of imatinib into and out of cells: implications for drug resistance. *Blood* 104(12)**,** 3739-3745. doi: 10.1182/blood-2003-12-4276.

Thomas, P., and Smart, T.G. (2005). HEK293 cell line: a vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods* 51(3)**,** 187-200. doi: 10.1016/j.vascn.2004.08.014.

Thomson, D.W., Bracken, C.P., and Goodall, G.J. (2011). Experimental strategies for microRNA target identification. *Nucleic Acids Res* 39(16)**,** 6845-6853. doi: 10.1093/nar/gkr330.

Titmarsh, G.J., Duncombe, A.S., McMullin, M.F., O'Rorke, M., Mesa, R., De Vocht, F., et al. (2014). How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *Am J Hematol* 89(6)**,** 581-587. doi: 10.1002/ajh.23690.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR

splicing factor phosphorylation. *Mol Cell* 39(6), 925-938. doi: 10.1016/j.molcel.2010.08.011.

Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S.M., Wu, X., et al. (2013). Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet* 9(3), e1003368. doi: 10.1371/journal.pgen.1003368.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., et al. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329(5992), 689-693.

Turhan, A.G., Hwang, J.W., Chaker, D., Tasteyre, A., Latsis, T., Griscelli, F., et al. (2021). iPSC-Derived Organoids as Therapeutic Models in Regenerative Medicine and Oncology. *Front Med (Lausanne)* 8, 728543. doi: 10.3389/fmed.2021.728543.

Ullah, S., John, P., and Bhatti, A. (2014). MicroRNAs with a role in gene regulation and in human diseases. *Mol Biol Rep* 41(1), 225-232. doi: 10.1007/s11033-013-2855-1.

Vainchenker, W., and Kralovics, R. (2017). Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood* 129(6), 667-679. doi: 10.1182/blood-2016-10-695940.

van Rhee, F., Hochhaus, A., Lin, F., Melo, J.V., Goldman, J.M., and Cross, N.C. (1996). p190 BCR-ABL mRNA is expressed at low levels in p210-positive chronic myeloid and acute lymphoblastic leukemias. *Blood* 87(12), 5213-5217.

Vannucchi, A.M., Lasho, T.L., Guglielmelli, P., Biamonte, F., Pardanani, A., Pereira, A., et al. (2013). Mutations and prognosis in primary myelofibrosis. *Leukemia* 27(9), 1861-1869. doi: 10.1038/leu.2013.119.

Vasudevan, S. (2012). Posttranscriptional upregulation by microRNAs. *Wiley Interdiscip Rev RNA* 3(3), 311-330. doi: 10.1002/wrna.121.

Vasudevan, S., Tong, Y., and Steitz, J.A. (2007). Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318(5858), 1931-1934. doi: 10.1126/science.1149460.

Vishnoi, A., and Rani, S. (2017). MiRNA Biogenesis and Regulation of Diseases: An Overview. *Methods Mol Biol* 1509, 1-10. doi: 10.1007/978-1-4939-6524-3_1.

Wada, H., Mizutani, S., Nishimura, J., Usuki, Y., Kohsaki, M., Komai, M., et al. (1995). Establishment and molecular characterization of a novel leukemic cell line with Philadelphia chromosome expressing p230 BCR/ABL fusion protein. *Cancer Res* 55(14), 3192-3196.

Wagle, M., Eiring, A.M., Wongchenko, M., Lu, S., Guan, Y., Wang, Y., et al. (2016). A role for FOXO1 in BCR-ABL1-independent tyrosine kinase inhibitor resistance in chronic myeloid leukemia. *Leukemia* 30(7), 1493-1501. doi: 10.1038/leu.2016.51.

Wagner, L.A., Christensen, C.J., Dunn, D.M., Spangrude, G.J., Georgelas, A., Kelley, L., et al. (2007). EGO, a novel, noncoding RNA gene, regulates eosinophil granule protein transcript expression. *Blood* 109(12), 5191-5198. doi: 10.1182/blood-2006-06-027987.

Wang, H., Iacoangeli, A., Lin, D., Williams, K., Denman, R.B., Hellen, C.U., et al. (2005). Dendritic BC1 RNA in translational control mechanisms. *J Cell Biol* 171(5), 811-821. doi: 10.1083/jcb.200506006.

Wang, H., Li, Q., Tang, S., Li, M., Feng, A., Qin, L., et al. (2017). The role of long noncoding RNA HOTAIR in the acquired multidrug resistance to imatinib in chronic myeloid leukemia cells. *Hematology* 22(4), 208-216. doi: 10.1080/10245332.2016.1258152.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol Cell* 43(6), 904-914. doi: 10.1016/j.molcel.2011.08.018.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.P., and Li, W. (2013a). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* 41(6), e74. doi: 10.1093/nar/gkt006.

Wang, L., Su, Y., Huang, C., Yin, Y., Chu, A., Knupp, A., et al. (2019). NANOG and LIN28 dramatically improve human cell reprogramming by modulating LIN41 and canonical WNT activities. *Biol Open* 8(12), bio047225. doi: 10.1242/bio.047225.

Wang, P., Ren, Z., and Sun, P. (2012). Overexpression of the long non-coding RNA MEG3 impairs in vitro glioma cell proliferation. *Journal of cellular biochemistry* 113(6)**,** 1868-1874.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., et al. (2013b). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev Cell* 25(1)**,** 69-80. doi: 10.1016/j.devcel.2013.03.002.

Wang, Z., Zheng, J., Pan, R., and Chen, Y. (2021). Current status and future prospects of patient-derived induced pluripotent stem cells. *Hum Cell* 34(6)**,** 1601-1616. doi: 10.1007/s13577-021-00592-2.

Wattanapanitch, M. (2019). Recent Updates on Induced Pluripotent Stem Cells in Hematological Disorders. *Stem Cells Int* 2019**,** 5171032. doi: 10.1155/2019/5171032.

Wen, F., Cao, Y.X., Luo, Z.Y., Liao, P., and Lu, Z.W. (2018). LncRNA MALAT1 promotes cell proliferation and imatinib resistance by sponging miR-328 in chronic myelogenous leukemia. *Biochem Biophys Res Commun* 507(1-4)**,** 1-8. doi: 10.1016/j.bbrc.2018.09.034.

Wolfe, H.R., and Rein, L.A.M. (2021). The Evolving Landscape of Frontline Therapy in Chronic Phase Chronic Myeloid Leukemia (CML). *Curr Hematol Malig Rep* 16(5)**,** 448-454. doi: 10.1007/s11899-021-00655-z.

Wu, S., Zheng, C., Chen, S., Cai, X., Shi, Y., Lin, B., et al. (2015). Overexpression of long non-coding RNA HOTAIR predicts a poor prognosis in patients with acute myeloid leukemia. *Oncol Lett* 10(4)**,** 2410-2414. doi: 10.3892/ol.2015.3552.

Wylie, A.A., Schoepfer, J., Jahnke, W., Cowan-Jacob, S.W., Loo, A., Furet, P., et al. (2017). The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. *Nature* 543(7647)**,** 733-737. doi: 10.1038/nature21702.

Xiao, G., Li, Y., Wang, Y., Zhao, B., Zou, Z., Hou, S., et al. (2018). LncRNA PRAL is closely related to clinical prognosis of multiple myeloma and the bortezomib sensitivity. *Exp Cell Res* 370(2)**,** 254-263. doi: 10.1016/j.yexcr.2018.06.026.

Xiao, M., Li, J., Li, W., Wang, Y., Wu, F., Xi, Y., et al. (2017a). MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol* 14(10)**,** 1326-1334. doi: 10.1080/15476286.2015.1112487.

Xiao, Y., Jiao, C., Lin, Y., Chen, M., Zhang, J., Wang, J., et al. (2017b). lncRNA UCA1 Contributes to Imatinib Resistance by Acting as a ceRNA Against miR-16 in Chronic Myeloid Leukemia Cells. *DNA Cell Biol* 36(1)**,** 18-25. doi: 10.1089/dna.2016.3533.

Xing, C.Y., Hu, X.Q., Xie, F.Y., Yu, Z.J., Li, H.Y., Bin, Z., et al. (2015). Long non-coding RNA HOTAIR modulates c-KIT expression through sponging miR-193a in acute myeloid leukemia. *FEBS Lett* 589(15)**,** 1981-1987. doi: 10.1016/j.febslet.2015.04.061.

Xing, Z., Lin, A., Li, C., Liang, K., Wang, S., Liu, Y., et al. (2014). lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* 159(5)**,** 1110-1125. doi: 10.1016/j.cell.2014.10.013.

Xu, M., Chen, X., Lin, K., Zeng, K., Liu, X., Xu, X., et al. (2019). lncRNA SNHG6 regulates EZH2 expression by sponging miR-26a/b and miR-214 in colorectal cancer. *J Hematol Oncol* 12(1)**,** 3. doi: 10.1186/s13045-018-0690-5.

Xu, Q., and Xie, W. (2018). Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. *Trends Cell Biol* 28(3)**,** 237-253. doi: 10.1016/j.tcb.2017.10.008.

Yang, L., Froberg, J.E., and Lee, J.T. (2014). Long noncoding RNAs: fresh perspectives into the RNA world. *Trends in biochemical sciences* 39(1)**,** 35-43.

Ye, Z., Zhan, H., Mali, P., Dowey, S., Williams, D.M., Jang, Y.Y., et al. (2009). Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. *Blood* 114(27)**,** 5473-5480. doi: 10.1182/blood-2009-04-217406.

Yin, D.D., Liu, Z.J., Zhang, E., Kong, R., Zhang, Z.H., and Guo, R.H. (2015). Decreased expression of long noncoding RNA MEG3 affects cell proliferation and predicts a poor prognosis in patients with colorectal cancer. *Tumour Biol* 36(6)**,** 4851-4859. doi: 10.1007/s13277-015-3139-2.

Yogarajah, M., and Tefferi, A. (2017). Leukemic Transformation in Myeloproliferative Neoplasms: A Literature Review on Risk, Characteristics, and Outcome. *Mayo Clin Proc* 92(7)**,** 1118-1128. doi: 10.1016/j.mayocp.2017.05.010.

Yoon, J.H., Abdelmohsen, K., Kim, J., Yang, X., Martindale, J.L., Tominaga-Yamanaka, K., et al. (2013). Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. *Nat Commun* 4**,** 2939. doi: 10.1038/ncomms3939.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., et al. (2012). LincRNA-p21 suppresses target mRNA translation. *Mol Cell* 47(4)**,** 648-655. doi: 10.1016/j.molcel.2012.06.027.

Yoshimura, H., Matsuda, Y., Yamamoto, M., Kamiya, S., and Ishiwata, T. (2018). Expression and role of long non-coding RNA H19 in carcinogenesis. *Front Biosci (Landmark Ed)* 23**,** 614-625. doi: 10.2741/4608.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I.I., et al. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 324(5928)**,** 797-801.

Zhang, X., Chen, X., Lin, J., Lwin, T., Wright, G., Moscinski, L.C., et al. (2012). Myc represses miR-15a/miR-16-1 expression through recruitment of HDAC3 in mantle cell and other non-Hodgkin B-cell lymphomas. *Oncogene* 31(24)**,** 3002-3008. doi: 10.1038/onc.2011.470.

Zhang, X., Hong, R., Chen, W., Xu, M., and Wang, L. (2019). The role of long noncoding RNA in major human disease. *Bioorg Chem* 92**,** 103214. doi: 10.1016/j.bioorg.2019.103214.

Zhang, X., Lian, Z., Padden, C., Gerstein, M.B., Rozowsky, J., Snyder, M., et al. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 113(11)**,** 2526-2534. doi: 10.1182/blood-2008-06-162164.

Zhang, X., Subrahmanyam, R., Wong, R., Gross, A.W., and Ren, R. (2001). The NH(2)-terminal coiled-coil domain and tyrosine 177 play important roles in induction of a myeloproliferative disease in mice by Bcr-Abl. *Mol Cell Biol* 21(3)**,** 840-853. doi: 10.1128/MCB.21.3.840-853.2001.

Zhang, X., Zhou, Y., Mehta, K.R., Danila, D.C., Scolavino, S., Johnson, S.R., et al. (2003). A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. *The Journal of Clinical Endocrinology & Metabolism* 88(11)**,** 5119-5126.

Zhao, Q., Zhao, S., Li, J., Zhang, H., Qian, C., Wang, H., et al. (2019). TCF7L2 activated HOXA-AS2 decreased the glucocorticoid sensitivity in acute lymphoblastic leukemia through regulating HOXA3/EGFR/Ras/Raf/MEK/ERK pathway. *Biomed Pharmacother* 109**,** 1640-1649. doi: 10.1016/j.biopha.2018.10.046.

Zhou, J.D., Lin, J., Zhang, T.J., Ma, J.C., Li, X.X., Wen, X.M., et al. (2018). Hypomethylation-mediated H19 overexpression increases the risk of disease evolution through the association with BCR-ABL transcript in chronic myeloid leukemia. *J Cell Physiol* 233(3)**,** 2444-2450. doi: 10.1002/jcp.26119.

Zhou, X., Yuan, P., Liu, Q., and Liu, Z. (2017). LncRNA MEG3 Regulates Imatinib Resistance in Chronic Myeloid Leukemia via Suppressing MicroRNA-21. *Biomol Ther (Seoul)* 25(5)**,** 490-496. doi: 10.4062/biomolther.2016.162.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10(1)**,** 1523. doi: 10.1038/s41467-019-09234-6.

Zhou, Y.G., Kim, J., Yuan, X.J., and Braun, T. (2011). Epigenetic Modifications of Stem Cells A Paradigm for the Control of Cardiac Progenitor Cells. *Circulation Research* 109(9)**,** 1067-1081. doi: 10.1161/Circresaha.111.243709.

Zou, T., Jaladanki, S.K., Liu, L., Xiao, L., Chung, H.K., Wang, J.Y., et al. (2016). H19 Long Noncoding RNA Regulates Intestinal Epithelial Barrier Function via MicroRNA 675 by Interacting with RNA-Binding Protein HuR. *Mol Cell Biol* 36(9)**,** 1332-1341. doi: 10.1128/MCB.01030-15.