Identifying regulators of chronic myeloid leukaemic (CML) stem cell quiescence and discovering effective novel therapies for CML relapse

Thesis submitted for the degree of Master of Sciences (MSc)

Zhiyan Xu

Haematopoietic Stem Cell Biology

Weatherall Institute of Molecular Medicine

St Edmund Hall, University of Oxford

Hilary term 2022

Index

Abstract ----------(4)

Acknowledgements ----------(6)

Abbreviations ----------(7)

Chapter 1 ----------(10)

Introduction and Aims ----------(10)

1. Haematopoietic stem cells ----------(11)

2. Brief introduction of chronic myeloid leukaemia (CML) ----------(12)

3. CML stem cells in TKI treatment resistance ----------(14)

4. *GAS2* is significantly highly expressed in CML stem cells ----------(15)

5. Aims of the thesis and project outline ----------(18)

Chapter 2 ----------(22)

Materials and Methods ----------(22)

1. Bioinformatic analysis for candidate targets prioritisation ----------(23)

2. Optimisation of gene knock-down/knock-out strategy ----------(25)

3. Preliminary functional validation in CML stem cells ----------(32)

Chapter 3 ----------(37)

Results ----------(37)

1. *GAS2* expression is particularly elevated in CML MPPs and Stem cells ---(38)

2. Optimisation of *GAS2* knock-out/knock-down strategy ----------(41)

3. Preliminary functional validation of *GAS2* knock-down/knock-out in cell lines and CML stem cells ----------(58)

Chapter 4 ----------(64)

Discussion ----------(64)

1. Quiescence of CML LSCs in TKI treatment resistance ----------(65)
2. *GAS2* function in CML stem cells ----------(65)
3. Future directions ----------(69)
4. Conclusion ----------(70)

Appendices ----------(72)

1. Key resources table ----------(72)
2. Supplementary figures ----------(75)
3. Supplementary tables ----------(80)
4. Bioinformatic scripts ----------(81)

References ----------(97)

Approximate word count = 17164

Abstract

Chronic myeloid leukaemia (CML) is a myeloproliferative neoplasm characterised by the presence of Philadelphia (Ph) chromosome. The standard treatment is tyrosine kinase inhibitors (TKI), while they achieve remission in many patients, the majority will relapse on cessation of therapy. This indicates that TKI do not ablate the CML stem cells. In this study, I analysed a single-cell transcriptome dataset of CML stem cells published in our lab [1] and other public datasets trying to identify candidate targets that may mediate stem cells resistance to TKI. *Growth Arrest Specific 2 (GAS2)* was the 'top hit' and therefore the candidate I took forward for functional validation.

I first sought to develop an efficient gene knock-down (KD) or knock-out (KO) strategy. Four approaches were compared: 1. Inducible lentiviral shRNA, 2. Constitutive lentiviral shRNA, 3. Single Cas9-sgRNA ribonucleoprotein (RNP) complex, 4. Dual Cas9-sgRNA RNPs. Comparing all approaches, constitutive shRNA achieved 67.2% of successfully transduced eYFP$^+$ cells, 86.2% reduction of *GAS2* mRNA level in CML CD34$^+$ cells and was used for further experiments.

My preliminary functional analysis suggested that *GAS2*^{KD} reduced quiescence and increased cell proliferation in both cell lines and primary CML CD34$^+$ cells. Then I repeated lentiviral shRNA transduction in different CML samples. Slightly higher proliferation index occurred in the *GAS2*^{KD} group than control^{KD} group. No significant changes in apoptosis or lineage differentiation were noted in *GAS2*^{KD} group *vs.* control^{KD} group. In summary, knock-down of *GAS2* modestly reduced the number of quiescent CML stem cells, suggesting potential therapeutic efficacy against TKI resistant CML stem cells. Further *in vitro* and *in vivo* studies are needed to validate and extend this observation.

Keywords: Quiescence, CML stem cell, Tyrosine kinase inhibitor resistance

Acknowledgements

I thank my supervisors Adam Mead and Bethan Psaila for their mentorship and support over the last three years, and for giving me this precious opportunity to perform this research. My research was funded by the Chinese Academy of Medical Sciences Oxford Institute.

I thank all members of the Mead and Psaila labs, for their kind help, suggestions, and friendship. I am sincerely grateful to Rong Li, the postdoc who guided me to many of these techniques when I started in the lab.

I would like to thank the WIMM FACS Facility team, for their help and advice on the antibody staining panel and cell sorting experiments. And the WIMM Genome Engineering Facility team, especially for their guidance and help on the shRNA plasmid design and lentivirus production.

I also thank Prof. Tao Dong, the director of Chinese Academy of Medical Sciences Oxford Institute, for organizing the program and offering me precious support and guidance along my time in Oxford.

Abbreviations

Abbreviations	Full form
7AAD	7-Amino-Actinomycin D
AGM	Aorta-gonad meso-nephros
AP	Accelerated phase
ATCC	American Type Culture Collection
BC	Blastic phase
BFU/CFU-E	Burst-forming unit and colony forming unit erythrocyte
BP	Blast phase
CFC	Colony forming cells
CFSE	Carboxyfluorescein diacetate succinimidyl ester
CFU	Colony forming unit
CFU-G	Colony forming unit granulocyte
CFU-GEMM	Colony forming unit granulocyte, erythrocyte, monocyte/macrophage, megakaryocyte
CFU-GM	Colony forming unit granulocyte & monocyte
CFU-M	Colony forming unit monocyte
CML	Chronic myeloid leukaemia
CP	Chronic phase
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CV	Coefficient of variation
DA	Dasatinib
DAPI	4', 6-diamidino-2-phenylindole
DMSO	Dimethyl sulfoxide
DN-GAS2	Dominant negative form of *Growth Arrest Specific 2*
ECL	Electrochemiluminescence
EDTA	Ethylenediamine tetra-acetic acid
ET	Essential thrombocythemia
eYFP	Enhanced yellow fluorescent protein
FACS	Fluorescence-activated cell sorting
FcR	Fc receptor

FDR	False discovery rate
FMO	Fluorescence minus one
Foxo3a	Forkhead O transcription factors 3a
G-CSF	Granulocyte colony-stimulating factor
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAS2	Growth arrest specific 2
GLM	General linear model
GM-CSF	Granulocyte macrophage colony-stimulating factor
H3K27triM	Polycomb-based histone H3 lysine 27 trimethylation
HDAC	Histone deacetylase
hEF1α	Human elongation factor-1 alpha
$hOCT_1$	Human organic cation transporter 1
HRP	Horseradish peroxidase
hSCF	Human stem cell factor
HSCs	Haematopoietic stem cells
HSPCs	Haematopoietic stem / progenitor cells
ICE	Inference of CRISPR edits
IDT	Integrated DNA technologies
IL-3	Interleukin-3
IL-6	Interleukin-6
IM	Imatinib
KD	Knock-down
KO	Knock-out
LSCs	Leukaemic stem cells
LTC-IC	Long-term culture initiating cells
MDR_1	Multi-Drug Resistance 1
MOI	Multiplicity of infection
MPN	Myeloproliferative neoplasms
MPPs	Multipotent progenitors
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next-generation sequencing
PBS	Phosphate-buffered saline
Ph	Philadelphia

PI	Proliferation index
PML	Promyelocytic leukaemia protein
PP2A	Protein phosphatase 2A
PS	Penicillin-Streptomycin
PV	Polycythaemia vera
RNAi	RNA interference
RNP	Ribonucleoprotein
RT	Room temperature
RT-qPCR	Real-time quantitative polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SFM	Serum-free medium
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
TBSt	Tris-buffered saline tween 20
TGF-β	Transforming growth factor β
TKI	Tyrosine kinase inhibitors
TNF-α	Tumour necrosis factor α
TPO	Thrombopoietin
tracrRNA	Trans-activating CRISPR RNA
tRFP	Turbo red fluorescent protein
tSNE	T-distributed stochastic neighbour embedding

CHAPTER 1.

Introduction

1. Haematopoietic stem cells (HSCs) in physiological haematopoiesis

HSCs are multipotent cells that self-renew and give rise to progeny that differentiates to replenish the different types of mature blood cells. Early studies of HSCs started with attempts to reverse the bone marrow failure of irradiated mice by injecting cells from unirradiated bone marrow [2]. Self-renewal versus (*vs.*) differentiation of HSCs results from their cell division. An individual HSC can undergo either symmetric division into 2 stem cells, or give rise to 2 differentiating progenitors, or can undergo asymmetric division into 1 stem cell and 1 differentiating cell. Some HSCs remain dormant, and only divide rarely. These cells can act as a reservoir for stress haematopoiesis. These properties enable stem cells to become tolerant to external stimulants and survive under environmental pressures. Literature shows that stem cells that remain in quiescence for a long time before division have a higher repopulation capacity [3]. Therefore, cell cycle state represents one aspect of heterogeneity among HSCs. Lineage-biased differentiation represents another aspect of the heterogeneity. Mice genetic studies and the advent of advancing genome sequencing technologies have revealed that HSCs' transcriptome and expression changes are under strict control of methylation and chromatin accessibility, independent of lineage selection [4]. With the advances of single-cell transplantation, the Müller-Sieburg and Jacobsen groups demonstrated that single HSC has distinct myeloid or lymphoid fate, they only execute limited patterns of lineage fates [5, 6]. Overall, both cell cycle status and lineage choices are regulated by both intrinsic factors as well as extrinsic regulators in the local bone marrow microenvironment to keep HSCs heterogenous and resilient to various stimulus [7].

2. Brief introduction of chronic myeloid leukaemia (CML)

Myeloproliferative neoplasms (MPN) are a group of haematologic diseases characterised by abnormally increased proliferation of bone marrow cells leading to increased numbers of myeloid cells in the peripheral blood. MPN are divided into 2 subgroups depending on the presence or absence of the 'Philadelphia (Ph) chromosome' – t (9; 22) (q34.1; q11.2), *i.e.*, the *BCR-ABL1* translocation. *BCR-ABL1*$^+$ MPN is chronic myeloid leukaemia (CML), which is characterised by uncontrolled production of granulocytes at different maturing stages. Symptoms of CML can be variable and may include fever, weight loss, night sweats, bone pain and discomfort from an enlarged spleen.

Maguer-Satta *et al.* [8] first proved that CML disease is developed by the clonal amplification of *Ph*$^+$ primitive cells in patients. They found that cells originating from all the myeloid lineages carry the leukaemic clone of *Ph*$^+$ in patients diagnosed with CML, and very high frequency of *Ph*$^+$ long-term culture-initiating cells (LTC-IC) was detected in 2 CML patients. This paper implicated that the presence of *BCR-ABL* transcripts in the most primitive cell populations can deregulate their normal function in patients.

The *BCR-ABL1* translocation is pivotal for the initiation and progression of CML. ABL is a non-receptor tyrosine kinase. BCR has multiple modular domains. The fusion of BCR to ABL leads to increasing tyrosine-kinase activity, proliferation, and growth-factor-independent survival [9]. Retroviral transduction of *BCR-ABL* into murine haematopoietic stem/progenitor cells (HSPCs) and mice bone marrow transplantation could induce the myeloproliferative disorder that resembled CML [10],

including increasing granulocytes in peripheral blood and splenomegaly. Furthermore, secondary transplantation still generates the same CML phenotype.

Besides its structural function, *BCR-ABL1* is involved in many pathways. Many signalling proteins become phosphorylated in *BCR-ABL* expressing cells and in turn activate downstream proteins and transcription factors, such as *RAS, AKT, STATs* and *MYC* [11]. They work together for the pathogenesis of CML, which lead to the altered cell adhesion/migration, regulation of myeloid cells proliferation and survival, and negative regulation of monocytic cell differentiation [12].

Given the fact that *BCR-ABL1* translocation is important in CML, inhibition of the tyrosine-kinase activity of BCR-ABL1 by tyrosine kinase inhibitors (TKI) are highly effective and have dramatically improved survival for patients with CML, enabling a complete cytogenetic response and major molecular response (MMR), *i.e.*, blood BCR-ABL levels decreased to no more than 0.1% on an international scale [13] using quantitative polymerase chain reaction, in a majority of patients. Around 40% of patients who achieve a prolonged and deep molecular remission on TKI therapy are able to discontinue TKI treatment without subsequent relapse for median gaps up to 8 years [14, 15]. However, most patients need to rely on long-term TKI treatment due to the persistence of CML stem cells, which increases the economic burden to patients as well as the health care systems. Corresponding to the theory that leukaemic stem cells serve as a reservoir for CML progression, researchers detected consistent existing *BCR-ABL1*$^{+}$ primitive cells in the bone marrow from TKI-treated CML patients, although they achieve MMR [16, 17]. Furthermore, researchers applied genetic engineering to let mice bone marrow cells express *BCR-ABL1*, those with low level of *BCR-ABL1* expression were more insensitive to

imatinib [18]. These findings point to the existence of *BCR-ABL1*$^{+}$ primitive cells as the major source of disease relapse under TKI treatment.

3. TKI treatment resistance – role of CML HSC quiescence

The resistance to TKI and subsequent relapse of CML have been a challenging issue in the haematology area. Kinase binding domain mutation is the most investigated mechanism of TKI resistance. Overexpression of efflux pump MDR_1 (Multi-Drug Resistance 1) and down-modulation of influx pump $hOCT_1$ (human Organic Cation Transporter 1) are other approved *BCR-ABL*-dependent mechanism of TKI resistance [19-21]. Another essential source of TKI resistance is CML leukaemic stem cells (LSCs), which are not dependent on BCR-ABL for their survival and self-renewal and therefore less vulnerable to TKI therapy [22].

CML LSCs have constitutive expression of the fusion tyrosine kinase BCR-ABL1, but maintain the same properties of self-renewal, proliferation, and differentiation as normal HSCs [23]. CML LSCs show a higher engraftment level and be able to initiate leukaemia in immunocompromised mice. A primary mechanism of resistance to treatment is that the quiescent CML LSCs are able to persist despite TKI therapy, thereby constituting a reservoir for disease recurrence if/when treatment is discontinued (**Figure 1**).

Holyoake *et al.* first isolated a highly quiescent subpopulation of HSCs in the bone marrow of CML patients, around 4% of primitive leukaemic cells in CML [24], which could spontaneously exit G_0 and proliferate independent of added growth factors [25]. Furthermore, imatinib, with its anti-proliferative activity, would increase re-entry to quiescence, by increasing the proportion of cells changing from proliferating to non-cycling state [26]. In addition, Roeder *et al.* raised a dynamic

model that quiescent CML LSCs were able to evade imatinib-induced apoptosis and degradation [27]. The mechanism of resistance of quiescent CML LSCs to imatinib differs from mechanisms of "acquired resistance", where permanent genetic changes result in imatinib-resistant subclones. In CML LSCs, imatinib-induced apoptosis would be restored once quiescent cells exit G_0 to enter cycling.

Several studies have aimed to design strategies to target quiescent stem cells in CML, either by promoting entry of quiescent CML stem cells into cycling, or by targeting *BCR-ABL* independent pathways which maintain stem cell quiescence. Jorgenson *et al.* [28] found that intermittent exposure to granulocyte colony-stimulating factor (G-CSF) during standard imatinib treatment could reduce the proportion of CML $CD34^+$ cells that were quiescent *in vitro*. Holtz *et al*. [29] also reported that granulocyte macrophage colony-stimulating factor (GM-CSF) and/or G-CSF exposure significantly reduced primitive progenitors validated in LTC-IC assays. Several *BCR-ABL* independent pathways were reported to help maintain stem cells in CML, *i.e.*, the forkhead box transcription factors 3a (Foxo3a) [30], histone deacetylase (HDAC) [31], inhibition of the tumour suppressor protein phosphatase 2A (PP2A) [32]. There are many ongoing clinical trials evaluating the therapeutic effects of exemplar inhibitors or activators of candidate [23].

4. *GAS2* is highly expressed in CML stem cells and may mediate HSC quiescence

There are not many studies validating targets involved in regulating cell cycle status in *BCR-ABL1*$^+$ *vs. BCR-ABL1*$^-$ stem cells, which could identify targets selectively expressed in CML LSCs but not healthy HSCs in patients with CML. Our group has successfully identified a subgroup of *BCR-ABL1*$^+$ CML LSCs that persist

through TKI treatment by combining single cell mutation detection with whole-transcriptome analysis in CML stem cells [1]. These cells clustered distinctly from healthy donors' HSCs, characterised by marked quiescence-related gene expression, such as transforming growth factor β (TGF-β) and tumour necrosis factor α (TNF-α) (via nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB))-associated gene expression. This study showed that over 1000 differentially expressed genes were detected in *BCR-ABL1*$^{+}$ CML LSCs compared to healthy donors' HSCs. Expression of the growth arrest specific 2 (*GAS2*) gene had the highest relative fold change [1]. Notably, *GAS2* was also among the top over-expressed genes in other datasets generated by our laboratory focused on HSPCs in *BCR-ABL*$^{-}$ MPNs, suggesting a potential role in other subtypes of myeloid cancers as well as CML [33].

GAS2 was first found specifically expressed and distributed along stress fibres in growth-arrested NIH 3T3 cells (mouse embryonic fibroblast cells). GAS2 protein was hyperphosphorylated and specifically localised with actin filaments in the membrane ruffling during cell transition from G_0 to G_1/S phase in NIH 3T3 cells [34, 35], suggesting that the post-transcriptional phosphorylation of GAS2 correlates highly with rearrangements of actin and cell cycle transition. *GAS2* is a component of the microfilament system in cells [36], whilst cell motility relying on the microfilament system is restricted in quiescent cells [37]. These observations imply that *GAS2* could be one of the key regulators for cell quiescence. Moreover, the function of *GAS2* in the cell cycle is cell type specific. GAS2 protein is upregulated in murine F9 embryonal carcinoma cells upon serum starvation, and *GAS2* was shown to facilitate G_1 to S cell cycle transition in Jurkat (human immortalized T lymphocyte cell line) and CCRF-CEM (human T lymphoblastoid) cell lines [38, 39].

GAS2 also acts as a dual regulator in apoptosis in different cells. GAS2 protein was proved to interact and inhibit the protease activity of calpain, thus enhancing the stability of p53 and apoptosis in U2OS (human bone osteosarcoma epithelial cell line) and MCF7 (human breast cancer epithelial luminal cell) cell line [40, 41]. Similar findings were also noted in prostate and breast cancers cell line, where *GAS2* was targeted by EZH2-mediated H3K27triM (polycomb-based histone H3 lysine 27 trimethylation) [42]. There is structural evidence that C terminal of GAS2 can be incised by caspase 3 and triggers alternation of microfilament for apoptosis induced cell morphology changes [43]. But recently a study by Kong *et al*. [39] reported that *GAS2* could inhibit apoptosis in Jurkat and CCRF-CEM cells relative to normal T lymphocytes. Huang *et al*. [44] validated that *GAS2* led to decreased calpain activity, stabilizing β-catenin protein, and apoptosis resistance in CML mice model. *GAS2* expression was also found related to the Fas-induced apoptosis resistance in Bcr-Abl$^+$ mice myeloid progenitor cells [44]. Overall, previous studies suggested that *GAS2* may act as apoptosis inhibitor in human haematopoietic cells.

There are several *Gas2* related pathways studies available using mice models. Huang *et al.* [45] used a retroviral vector expressing *p210 Bcr/abl* to transduce myeloid progenitor cells isolated from bone marrow of wild-type (WT) mice, they found that *Bcr/abl* increased *Gas2* mRNA and β-catenin protein expression in mice myeloid progenitors. Then they examined the same *Bcr/abl*$^+$ mice myeloid progenitors and validated the increased Stat5 protein expression caused by Calpain inhibition by *Gas2* [46]. **Figure 2** summaries all currently proposed signalling pathways in relating to *GAS2*.

In terms of *GAS2* related studies in CML from human samples, Janssen *et al.* [36] first detected significantly higher expression of *GAS2* in peripheral blood mononuclear cell pellets from patients at blast crisis *vs.* diagnosis and late chronic phase. This was further confirmed by DNA microarray analysis of bone marrow from 91 CML cases [47], *GAS2* as among the top 10 up-regulated genes in accelerated and blast phases *vs.* chronic phase. Another study tested gene expression level using $CD34^+$ cells purified from peripheral blood in patients with chronic myeloproliferative disorder and bone marrow in healthy donors, and *GAS2* was also significantly up regulated in idiopathic myelofibrosis, polycythaemia vera (PV) and essential thrombocythemia (ET), compared with healthy donors [48]. In addition to the detection of high expression of *GAS2*, several gene manipulation was applied as well. Zhou *et al.* once transduced human CML bone marrow $CD34^+$ cells with lentiviral vectors expressing a dominant negative form of *GAS2* (*DN-GAS2*), and revealed that transduced CML cells had higher inhibition of colony forming cells (CFC) compared with healthy donors [49]. Hjort *et al.* [46] transduced human CML bone marrow $CD34^+$ cells with retroviral vectors expressing *DN-GAS2*, the calpain activity was increased by inhibition of *GAS2*. *In vitro* culture of transduced $CD34^+$ CML cells suggested that *GAS2* inhibition combining imatinib treatment could increase cell apoptosis [44]. To date, no studies have elucidated the relation of *GAS2* to cell quiescence exit in human CML stem cells.

5. Aims of the thesis and project outline

CML stem cells resistance to TKI treatment remains a challenging question in CML therapy. The function of the candidate target, *GAS2,* in terms of TKI resistance, cell quiescence, apoptosis and lineage differentiation in CML stem cell are largely

unknown. This leads to our hypothesis that the high expression of *GAS2* within CML stem cells could be essential for cell quiescence and persistence during TKI treatment. The overarching goal of my project was to test this hypothesis, to determine if targeting *GAS2* could be a successful approach to target CML LSCs. Specifically, my aims were:

(1) To confirm *GAS2* as a 'top hit' in CML stem cells by bioinformatic analysis of in-house and published transcriptomic analyses of samples from patients with CML.
(2) To establish and optimise a platform for *GAS2* knock-down/knock-out that can be used in primary HSCs from patients.
(3) To test the functional role of *GAS2* in CML HSCs on cell cycle, differentiation kinetics and apoptosis at steady-state and in response to TKI treatment.

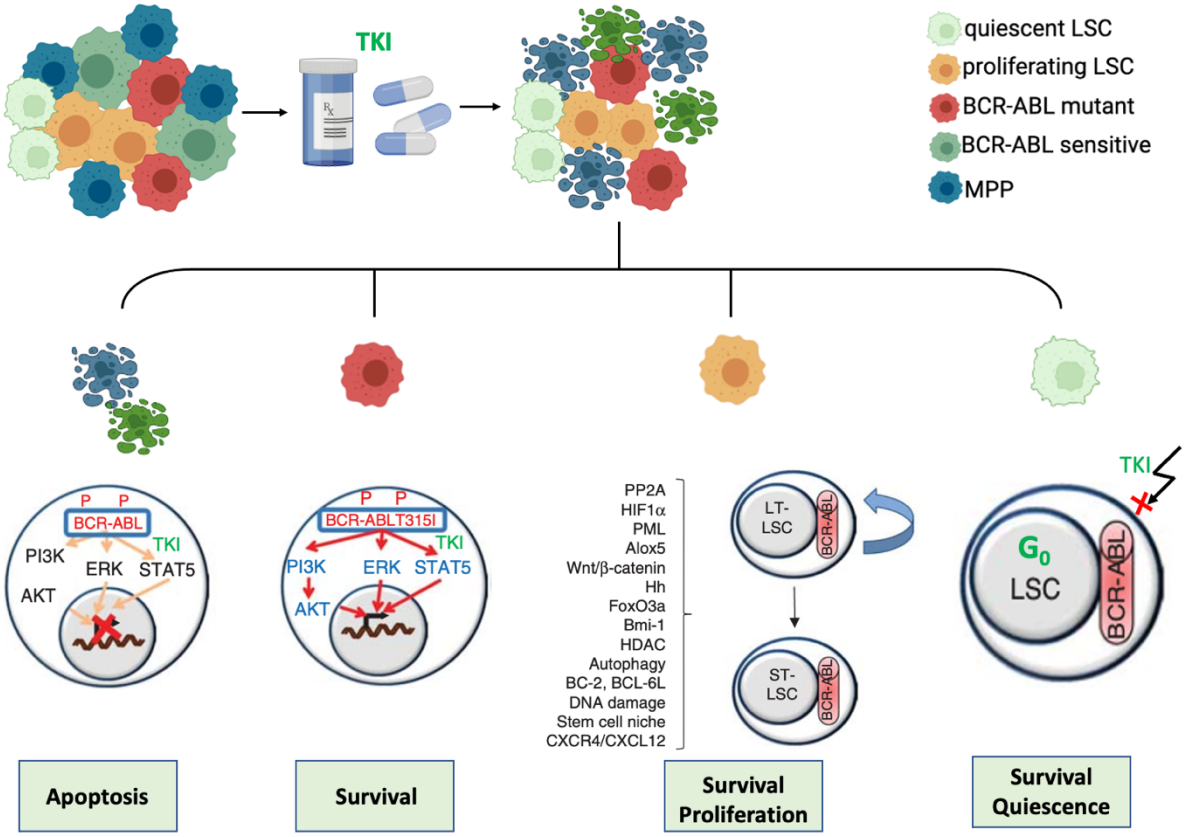


Figure 1. Mechanism of CML cells resistance to tyrosine kinase inhibitor (TKI).

Schematic diagram of heterogenous CML cells responses to TKI. Treatment of TKI leads to apoptosis of *BCR-ABL* sensitive CML cells and multipotent progenitors by inactivating *BCR-ABR* signalling pathways. CML cells with *BCR-ABL* mutation, *T315I*, evades TKI treatment and survives. Proliferating LSCs rely on *BCR-ABL* independent pathways for survival under TKI treatment. Quiescent LSCs are inactive and dormant, the proliferative effect caused by *BCR-ABL* signal is low, which makes TKI hard to target. LSC: leukaemic stem cells, MPP: multipotent progenitors.

Adapted from Morotti A, 2014 [50].

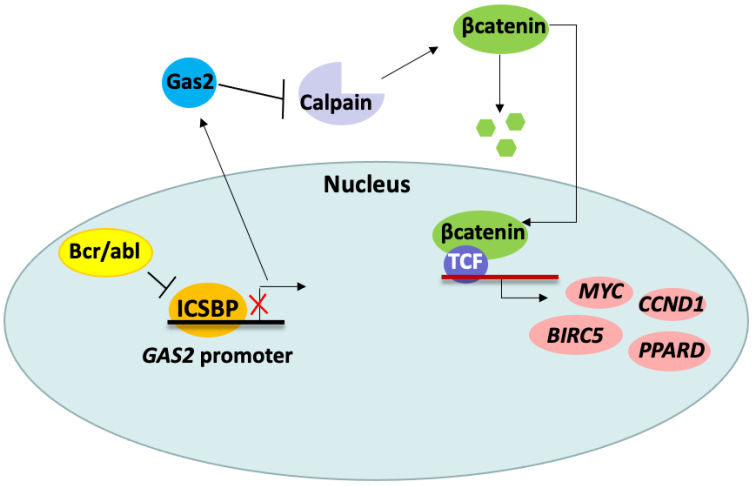


Figure 2. Schematic summaries of *GAS2* interactions with Bcr/abl.

Bcr/abl protein inhibits the transcription factor of Interferon Consensus Sequence Binding Protein (ICSBP / IRF8), which acts as a suppressor of *GAS2*. Increased expression of Gas2 protein can decrease Calpain activity, which further prevents β-catenin from being degraded. Nuclear translocation of β-catenin promotes transcription of target genes: *e.g., MYC, CCND1, PPARD, BIRC5*, and modulates proliferation, differentiation, cell cycle, apoptosis, *et al*.

Adapted from Huang W, 2010 [45]

CHAPTER 2.

Materials and Methods

1. Bioinformatic analysis to prioritise candidate targets that may mediate HSC quiescence

I first re-analysed a published dataset from a previous study in our lab [1]. In this study, single-cell transcriptomics were used to identify the distinct molecular signatures of CML HSCs by comparing the gene expression profile of *BCR-ABL1*$^+$ HSCs to wild-type HSCs within the same patient, as well as to HSCs from healthy donors.

I extracted two groups of differentially expressed genes (DEG) from this dataset by comparing:

(1) *BCR-ABL1*$^+$ CML HSCs *vs.* healthy donor HSCs,
(2) *BCR-ABL1*$^+$ *vs. BCR-ABL1*$^-$ HSCs within CML patients.

These DEGs were directly downloaded from the "Supplementary Table 2" from the paper published [1]. I then used a volcano plot to visualize the top up-regulated or down-regulated genes using Graphpad Prism (v9.0.0). The general linear model (GLM) was used to model coefficient of variation (CV) and mean expression of genes to estimate highly variable genes from each type of cells. I directly downloaded the R-object data files of highly variable genes and cell annotation text on Github (https://github.com/supatt-lab/Giustacchini-Thongjuea-et-al.-Nat.Med.2017) for my analysis. Then I performed T-distributed stochastic neighbour embedding (tSNE) analysis using those highly variable genes for distinct clustering of *BCR-ABL1*$^+$, *BCR-ABL1*$^-$ and healthy donors' HSCs. The "Limma" package was applied first to combine the annotation and "all cells R-object data" to remove batch effect, a large matrix of all cells with annotated gene expression was generated. "Rtsne" package was applied to the matrix for tSNE analysis, arguments were set

as: (dims, initial_dims, perplexity) = (2, 20, 20). *GAS2* expression values were imposed on every cell in tSNE clustering, I used "AUCell" package to calculate the gene set score of quiescence and cell cycle S phase in each cell and imposed them in tSNE clustering. "ggplot2" package was used for figures plotting. The detailed programming scripts and genes in each set are available in **Appendix 4**.

To help prioritise genes from these lists for target validation, I also analysed other publicly available datasets. I extracted data from the Affymetrix Human Gene 1.0 ST array (GEO accession number GSE47927), which was first published in *Blood* 2013 [51]. This dataset compared gene expression in HSCs and progenitors from patients with CML in chronic phase (CP), accelerated phase (AP) and blastic phase (BC) and from healthy donors. The microarray data was downloaded using "GEOquery" package, followed by probe identification and annotation using "hugene10sttranscriptcluster.db" package. Values of *GAS2* expression in all cells were extracted to "csv" format. Only HSCs were filtered for downstream analysis using "dplyr" package. After data annotation and cleaning, the data frame was transformed into matrix before differential expressed gene analysis using "Limma" package. Volcano plot of differential expressed genes in CML HSCs *vs.* healthy donors' HSCs was depicted using "ggpubr" package, genes with p value > 0.01 were marked as stable genes, otherwise logFC > 1 were up-regulated genes, logFC < -1 were down-regulated genes, *GAS2* was labelled in the volcano plot. Boxplot of *GAS2* expression in CML HSCs and healthy donors' HSCs was plotted using "ggpubr" package as well. Heatmap of top up-regulated genes (logFC > 2) and down-regulated genes (logFC < -2) was plotted using "pheatmap" package. The detailed programming scripts are available in **Appendix 4 Script 2**. The software used for analysing these datasets are R (v3.6.3) and R studio (v1.2.5033).

2. Optimisation of *GAS2* knock-down/knock-out method in cell lines and CML CD34$^+$ cells

Culture and maintenance of cell lines

SET-2 (human megakaryoblastic cell line), HEL (human erythroleukaemia cell line), K562 (human erythroleukaemia cell line), Jurkat (leukaemic T cell lymphoblast) and HEK-293T (human embryo kidney epithelial cell line) were obtained from the American Type Culture Collection (ATCC), U266 (lymphoblast from multiple myeloma) and RPMI8226 (lymphoblast from multiple myeloma) cells were kindly provided by Dr Sarah Gooding (Radcliffe Department of Medicine, University of Oxford), All cells except for HEK-293T were maintained in RPMI1640 medium (Gibco, Cat#22409031) containing 10% (v/v) fetal calf serum (FCS), 1% (v/v) Penicillin-Streptomycin (PS) and 1mM L-glutamine. RPMI8226 cells were cultured in the same medium but supplemented with 2mM L-glutamine. HEK-293T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco, Cat#31885049) containing 10% (v/v) FCS and 1% (v/v) PS.

SMARTvector inducible lentiviral shRNA vectors

Three short-hairpin RNAs (shRNA) targeting different sequences within *GAS2* were purchased from Horizon Discovery Ltd, the full virus production was performed by Genome Engineering Facility of Weatherall Institute of Molecular Medicine. They were "SMARTvector Inducible Lentiviral shRNA vectors" (**Supplementary Figure 1**), the vector size was 10,090bp. I ordered vectors containing human elongation factor-1 alpha (hEF1α) promoter and turbo red fluorescent protein (tRFP) reporter. Inclusion of the puromycin resistance gene following the hEF1α promoter can

permit antibiotic selection for successfully transduced cells. “Tet-On 3G” element can be regulated by the presence of doxycycline to activate TRE3G promoter (inducible promoter with tetracycline response elements), after which tRFP is expressed for visual tracking of transduction and shRNA targeting sequence is expressed for RNA interference (RNAi). A set of 3 lentiviral shRNA constructs targeting different locations of *GAS2* (A9, A10, B10) and 1 construct for nontargeting control were used, details of the sequence information are listed in **Appendix 1**.

Constitutively expressed lentiviral shRNA vectors

I further used a smaller size lentiviral vector, pLKO_eYFP (enhanced yellow fluorescent protein), 7120bp. It was a constitutively expressed vector with U6 promoter and eYFP reporter. Three different shRNA sequences targeting *GAS2* were incorporated into it individually to form 3 plasmids. The plasmid map is shown in **Supplementary Figure 2**. I designed the shRNA sequences according to the guideline on https://www.addgene.org/protocols/plko/. The online design tool, http://sirna.wi.mit.edu/show_oligo.cgi, was used. Details of the shRNA sequences are listed in **Appendix 1**.

Titre determination of lentivirus

Virus titration was tested via transduction on HEK-293T cells. 3 x 10^5 HEK-293T cells were plated per well in a 12-well plate and cultured overnight for adherence. On the next day, the medium was carefully withdrawn from the side, 0.5, 1, and 2 µl of thawed lentiviral stock were added separately into each well with 500 µl new media containing 4 µg polybrene, the virus dilutions were 1:1000, 1:500 and 1:250 respectively. The plate was returned to the incubator for at least 3 hours, then 1 ml of complete culture medium was added with 1 µg/ml doxycycline and without

polybrene per well for another 2 days. The cell culture plate was retrieved and washed with phosphate-buffered saline (PBS) and trypsinised, centrifuged at room temperature at 500 × g for 5 minutes, live lentiviral particles were removed by washing with PBS 3 times. Cells were analysed by flow cytometry to determine the percentage of RFP positive cells, *i.e.*, the rate of virus transduction (**Figure 1**). The viral titre was calculated by the following formula:

$$Viral\ titre\ (TU/ml) = \frac{\%RFP\ \times\ cell\ number\ \times\ dilution}{transduction\ volume\ (ml)}$$

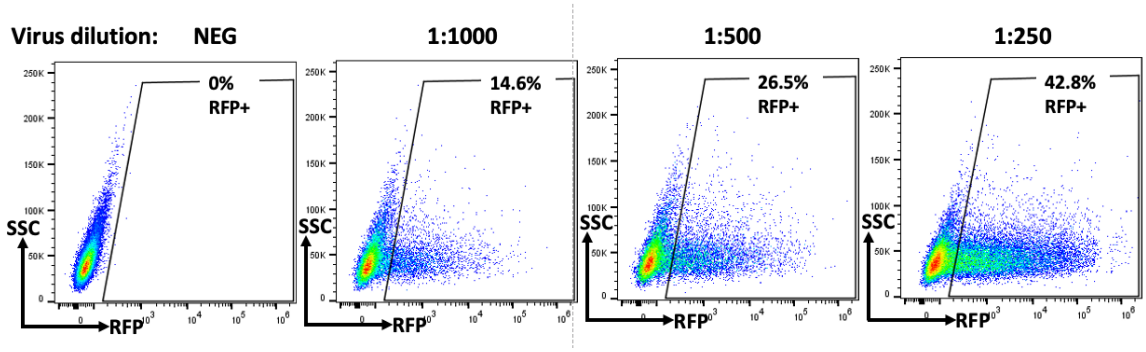


Figure 1. Example flow cytometry plots showing the RFP+ cells gating and percentage after transduction with different virus dilutions. NEG: negative control.

Transduction of SMARTvector inducible lentiviral shRNA

Viral transduction in cell lines was performed at a viral multiplicity of infection (MOI) around 0.3-0.5 (TU/cell) to produce a single-copy lentiviral genome integration, using the same culture medium with 8 μg/ml polybrene. Cells and viruses were centrifuged at 2000 rpm at 25 °C for 30 minutes and incubated for 48 hours. Live lentiviral particles were removed by washing with PBS 3 times.

Different concentrations of puromycin: 0, 0.5, 1, 2, 4, 6, 8 μg/ml, were added into 1 x 10^6 RPMI8226 cells in the 12-well plate for 3 days. The lowest concentration to cause all cell death on day 3 was the optimal concentration and was used in the proceeding experiment. Optimal concentration of puromycin was added for 3 days, so all cells without transduction were dead. The optimal concentration of doxycycline was determined using the same gradient doxycycline concentrations' culture strategy: the optimal concentration is the one that causes the highest turbo-RFP signal and the highest cell viability. Then optimal concentration of doxycycline was added for signal induction to enable turbo-RFP and target shRNA sequence expression.

Patient primary cells were pre-stimulated overnight in serum-free medium (SFM) (StemSpan) with 100 ng/ml recombinant human Fms-like tyrosine kinase 3 ligand (FL), 100 ng/ml human stem cell factor (hSCF), 100 ng/ml thrombopoietin (TPO), 2 mM glutamine. 0.3 ml of 25 μg/ml retronectin was dispensed (Takara Cat#T100B) into each well of the 48-well plate the night before and stored at 4 °C. Then the retronectin solution was removed. CML $CD34^+$ cells were seeded in 0.7 PS/ml for 400 μl per-well in a 48-well plate, virus was added with MOI = 2.19 (TU/cell), and spinoculation was performed for 1 hour at 25°C at 2000rpm. More virus with MOI = 1.09 (TU/cell) and 4 μg/ml polybrene were added immediately post spinoculation, cells were incubated in a 37 °C, 5% CO_2 incubator for 48 hours before washing out virus and continuing culturing in medium with 0.75 μg/ml doxycycline. Post 3 days of doxycycline induction, the transduction efficiency was analysed by detecting turbo-RFP percentage using flow cytometry.

Transduction of constitutive lentiviral shRNA

Patient primary cells were pre-stimulated overnight same as above. 8 x 10^5 CML $CD34^+$ cells were seeded in 1 ml medium (same as above) per-well in 12-well plate, virus was added with MOI = 20 (TU/cell), polybrene was added with a final concentration of 4 μg/ml, spinoculation was performed for 1.5 hour at 30 °C at 1000 x g. Cells were incubated in a 37 °C, 5% CO_2 incubator immediately post spinoculation, and 0.5 ml medium was added at least 3 hours later. 48 hours post transduction, cells were washed with PBS for 3 times to remove virus, and were resuspended in culture medium with a density of 8 × 10^5 /ml. The transduction efficiency was analysed by detecting eYFP percentage using flow cytometry.

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) knock-out of *GAS2*

A set of 3 single guide RNA (sgRNA) targeting *GAS2* and 1 sgRNA targeting intronic *AAVS1* as positive control were bought from Synthego company, their sequences were listed in **Appendix 1**. sgRNA is a single RNA that contains both the pre-designed CRISPR RNA (crRNA) and the trans-activating crRNA (tracrRNA). The synthetic format of sgRNA by chemical modification was used in experiments. The sequences of sgRNA were listed in **Appendix 1**. Cas9 nuclease protein was purchased from Integrated DNA Technologies (IDT, Cat#1081059). Cas9 protein and modified sgRNA were pre-assembled into ribonucleoprotein (RNP) complexes with a molar ratio of 1: 2.76 (Cas9: sgRNA) under 21 °C incubation for 10 minutes.

Nucleofection for $CD34^+$ HSPCs was performed using the Lonza 4D-Nucleofector (program EO-100) and the P3 Primary Cell 4D-Nucleofector Kit (Lonza, Cat#V4Xp-3024). Nucleofection conditions: 22 μl nucleofection solution, 10^5 cells, RNPs formed by 3.2 μg chemically modified sgRNA and 6 μg Cas9 protein. Cells

were incubated in 200 µl medium in a 96-well plate in a humidified 37 °C / 5 % CO_2 incubator immediately after nucleofection.

Nucleofection for K562 cells were performed using the Lonza 4D-Nucleofector (program FF-102) and the SF Cell line 4D-Nucleofector X kit (Lonza, Cat#V4XC-2012). Nucleofection conditions: 100 µl nucleofection solution, 1 x 10^6 cells, RNPs formed by 6.4 µg chemically modified sgRNA and 12 µg Cas9 protein. Cells were incubated in 1.5 ml medium in a 12-well plate in a humidified 37 °C / 5 % CO_2 incubator immediately after nucleofection.

Measuring allele modification frequencies using ICE

Genomic DNA was extracted from cells 2 days after nucleofection using Qiagen DNeasy Blood and Tissue Kit (QIAGEN, Cat#69506) following manufacturer's instructions. PCR amplicons spanning the genome target site by sgRNA were generated using KAPA2G Robust HotStart ReadyMix (Cat#KK5702) with primer pairs over target region of *GAS2* and *AAVS1,* sequences of primers were listed in **Appendix 1**. PCR amplicons were separated by gel electrophoresis, the band with expected sizes was cut and purified by QIAquick Gel Extraction Kit (Cat#28706). The purified PCR amplicons were Sanger-sequenced using both PCR primers and all sequencing chromatograms were uploaded and analysed by online inference of CRISPR edits (ICE) software https://ice.synthego.com/#/. The ICE software tool can offer accurate results which correlate strongly with the next-generation sequencing (NGS) based analysis [52]. The reference sequence used for analysis was from the positive control sample.

Gene expression analysis: reverse-transcription quantitative PCR

For total cell number > 5 x 10^4, RNA was extracted from the cell pellet using the RNeasy Mini Kit (QIAGEN, Cat#74104). For cell numbers < 5 x 10^4, RNA was extracted using the RNeasy Plus Micro Kit (QIAGEN, Cat#74034). cDNA was synthesized from RNA using EvoScript Universal cDNA Master (Roche, Cat#07912455001) for the real-time quantitative PCR (RT-qPCR). The RT-qPCR was the target amplification using cDNA as the template, which was implemented using TaqMan Gene Expression Assays (**Appendix 1**) and the TaqMan Universal PCR Master Mix (ThermoFisher, Cat#4304437). Housekeeping gene *glyceraldehyde-3-phosphate dehydrogenase (GAPDH)* was used as internal control. The comparative C_T method was used for relative quantification.

Western blotting

2 x 10^5 cells from cell lines were collected and washed twice with ice-cold PBS, the pellet was lysed in RIPA buffer containing 1X protease and phosphatase inhibitor cocktail and 1mM ethylenediamine tetra-acetic acid (EDTA). Samples were incubated on ice for 15 minutes, sonicated for 1 minute to ensure complete cell membrane lysis, and centrifuged for 15 minutes at 14000 rpm at 4 °C. The supernatant was collected as intracellular proteins. A Bradford assay was performed to quantify the protein concentration.

20 µg proteins from each sample were heated at 100 °C for 5 minutes and were separated on 4%–12% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gels with 1.0 mm wells (Invitrogen, Bolt™, Cat#NW04125BOX) alongside a page ruler pre-stained protein ladder (Biolabs, NEB, Cat#P7719). Gels were run at 200 V for 22 minutes. Proteins were then transferred from the electrophoresed gel onto a nitrocellulose membrane (Thermo

Scientific, Cat#77010). The nitrocellulose membrane was blocked in Tris-buffered saline Tween 20 (TBSt) with 5% skim milk powder for 1 hour at room temperature (RT), washed 3 times with TBSt, and incubated with a primary antibody (2 µg/ml mouse monoclonal antibody for GAS2 (Santa Cruz Biotech, Cat#sc-398669), a rabbit monoclonal antibody to GAS2 (1:1000 dilution, Abcam, Cat#ab109762), 0.06 µg/ml mouse monoclonal antibody for tubulin (Abcam, Cat#ab7291) in 1% skim milk) in the shaker at 4 °C overnight. On the next day, the membrane was washed 3 times with TBST for 5 minutes per wash and incubated with a corresponding secondary antibody (conjugated to horseradish peroxidase (HRP), CST, Cat#7076 and Cat#7074) for 1 hour at RT. The membrane was washed again for 3 times with TBSt for 5 minutes per wash, incubated with electrochemiluminescence (ECL) western blot detection reagents and visualized using the Bio-Rad system.

3. Preliminary functional validation in CML stem cells

Processing of human samples

Patients were consented to the InForMED Study (REC#199833). Peripheral blood samples were obtained from patients diagnosed with CML prior to treatment initiation. $CD34^+$ cells were enriched with “CD34 MicroBead Kit UltraPure, human” (Miltenyi Biotec), following the manufacturer’s instructions. Enriched $CD34^+$ cells were resuspended into freezing buffer containing 90% FCS and 10% dimethyl sulfoxide (DMSO) and stored into Mr. Frosty freezing container in - 80 °C freezer overnight before transferring to liquid nitrogen tank for long time cryopreservation.

Cryopreserved $CD34^+$ cells were thawed by warming at 37 °C water bath for 2 minutes, gradually added into RPMI-1640 media supplemented with 10% FCS and 0.1 mg/ml DNase I, centrifuged at 300 g for 10 minutes.

Human primary cells were cultured for 4 days in SFM (StemSpan) with 100 ng/ml recombinant human FL, 100 ng/ml hSCF, 20 ng/ml recombinant human IL-3, 20 ng/ml recombinant human IL-6, and 20 ng/ml G-CSF. The seeding density was 8 × 10^5 /ml, the medium was changed every 3 days.

Inhibitors

Imatinib (IM) and Dasatinib (DA) was purchased from Selleckchem. Stock solutions of 10 μM were dissolved into DMSO and stored at -20 °C. IM was treated with 10 μM in CML $CD34^+$ cells, and DA was treated with 150 nM, or the same volume of DMSO was added on CML $CD34^+$ cells as control for 3 days, then cells were washed and resuspended into the same culture medium without drugs, no additional inhibitors or cytokines were added.

Flow cytometry analysis

Cells were re-suspended in the fluorescence-activated cell sorting (FACS) buffer, which was composed of PBS, 2% (v/v) FCS, 2% (v/v) EDTA, and 1% (v/v) PS. The LSRII flow cytometer (BD Biosciences) was applied and calibrated as manufacturers' instructions.

Fluorescence-activated cell sorting was performed using a Becton Dickinson Fusion 1 machine. Cells were first washed by FACS buffer followed by centrifuging at 350 × g for 5 minutes at room temperature. 1 μl human Fc Receptor (FcR) blocking reagent per 200 μl FACS buffer were added in cell pellets and incubated on ice for 5 minutes. Single colour-stained control and fluorescence minus one (FMO) control were used for all experiments. Cells were stained with the following antibody cocktail for 15 minutes on ice in the dark prior to sorting: CD34-APC-eflour780 (1:50, Thermo Fisher Scientific); Lineage-BV510 (1:50, BioLegend);

CD38-Brilliant Violet 650 (1:50, BioLegend). 7-Amino-Actinomycin D (7AAD) was used for live/dead exclusion. Target cells were isolated into 1.5 ml Eppendorf tubes or PCR tubes depending on the experiment. All flow cytometry files were analysed via FlowJo (BD Life Sciences).

Cells division tracking and cell proliferation assay

Carboxyfluorescein diacetate succinimidyl ester (CFSE) is an amine-reactive dye. It enters live cells and is hydrolysed by cytoplasmic esterase to release fluorescent dyes. They form stable and consistent protein-conjugates with intracellular proteins on the amine groups. Once cells divide, the fluorescent dyes are also halved in daughter cells. After culturing for several days, cells are analysed by flow cytometry and different generations of cells can be differentiated according to their different fluorescence intensities.

CellTrace™ CFSE Cell Proliferation Kit (Invitrogen, Cat#C34554) and CellTrace™ Violet Cell Proliferation Kit (Invitrogen, Cat#C34557) were used in this study. 1×10^6 cells were labelled with 5 μM CFSE or CellTrace™ Violet in 37 °C water bath in dark for 30 minutes. After washing out exceeding dyes, CellTrace labelled cells were cultured in a humidified incubator at 37 °C in an atmosphere of 5-7% CO_2 for 30 minutes to allow the dye to undergo acetate hydrolysis. 5×10^4 cells were taken out and fixed in 1% formaldehyde solution. This was used as undivided control for generation 0. Quiescent and dividing cells could be differentiated by flow cytometry according to the histogram of CellTrace signal. The FlowJo's proliferation platform was used to analyse and quantify the proliferation dynamics over time. The built-in algorithm could simulate the right model of

proliferation. Proliferation index (PI) was calculated by dividing the total number of cell divisions from the number of cells that entered division.

Apoptosis assay

After surface markers staining, cells were washed once in PBS and then once in Annexin V binding buffer (Invitrogen, Cat#BMS500BB). 1×10^6 cells were labelled with 5 µl Annexin V-APC (Thermo Fisher Scientific, Cat#A35110) in 100 µl Annexin V binding buffer and were incubated on ice for 15 minutes in dark. Cells were then stained with 4', 6-diamidino-2-phenylindole (DAPI) (0.1 µg/ml) for 10 minutes before analysing on flow cytometry machine (FACS Calibur; BD).

Colony forming unit (CFU) assay

Three hundred to five hundred target cells were collected in Eppendorf tubes during FACS-sorting. They then were seeded in methylcellulose-based medium (MethoCult™ H4435 + 1% (v/v) PS) in 35 mm grid culture dishes. After 13 days of incubation, colonies were scored under an inverted microscope. Colony types were scored as below: CFU-GM, colony forming unit granulocyte & monocyte; CFU-G, colony forming unit granulocyte; CFU-M, colony forming unit monocyte; BFU / CFU-E, burst-forming unit and colony forming unit erythrocyte; CFU-GEMM, colony forming unit granulocyte, erythrocyte, monocyte/macrophage, megakaryocyte.

Colony identification by Giemsa stain

The target colony was plucked using the 20 µl tip into 100 µl PBS. The whole cell suspension was transferred into cytospin funnels, followed by centrifuging with low acceleration at 400 rpm for 5 minutes for fragile cells. Slides were taken out to let air dry before staining. Giemsa-May-Grünwald stain was prepared as per

manufacturers protocol. Slides were soaked in the fixation buffer (blue) for 30 seconds, followed by 2 minutes in the staining buffer 1 (red), 2 minutes in the staining buffer 2 (purple). Slides were flushed from back slide with a very low speed of water running. Then slides were air dried by standing on tissue paper. Cells on the slide can be inspected using an un-inverted microscope. Slides were stored in a plastic container at room temperature.

CHAPTER 3.

Results

1. *GAS2* expression is particularly elevated in multipotent progenitors (MPPs) and HSCs from CML patients

To identify genes that were differentially expressed within HSCs from CML patients, I analysed a previously published dataset from our lab[1]. This study reported the single-cell transcriptomes of 477 *BCR-ABL1*$^+$ HSCs and 377 *BCR-ABL1*$^-$ HSCs from 18 patients with chronic phase (CP) CML sampled at diagnosis, and 232 normal HSCs from 5 healthy donors. The comparison between both *BCR-ABL1*$^+$ HSCs *vs.* healthy donor HSCs and *BCR-ABL1*$^+$ HSCs *vs. BCR-ABL1*$^-$ HSCs revealed *GAS2* to be the highest differentially expressed gene (false discovery rate (FDR)-adjusted $P < 0.01$; **Figure 1.1A-1.1B**). tSNE visualization of the single-cell transcriptome using highly variable genes revealed distinct clustering of *BCR-ABL1*$^+$ HSCs, *BCR-ABL1*$^-$ HSCs and normal HSCs. *GAS2* expression level of each cell was imposed onto tSNE clustering. Elevated expression of *GAS2* was markedly striking in the cluster of *BCR-ABL1*$^+$ HSCs (**Figure 1.1C, 1.1D, 1.1G**). The gene set score of quiescence and cell cycle S phase were imposed on tSNE clustering as well. The AUC histogram of the gene sets and t-SNE plot based on the automatically assigned AUC score threshold are available in **Supplementary Figure 3**. Overall, the quiescence score in *BCR-ABL1*$^-$ HSCs and normal HSCs were higher than *BCR-ABL1*$^+$ HSCs. On the contrary, the cluster of *BCR-ABL1*$^+$ HSCs had higher S phase score (**Figure 1.1E-1.1F**). Within the cluster of *BCR-ABL1*$^+$ HSCs, the high expression of *GAS2* correlated with high score of quiescence (**Figure 1.1D- 1.1E**).

In addition, I analysed another gene expression array dataset (GEO accession number GSE47927) comparing gene expression in HSPCs from patients with CML and from healthy donors [51]. This dataset contains gene expression in

cryopreserved $CD34^+$ enriched cell populations from the leukapheresis sample in CML patients in CP, accelerated phase (AP) or blastic phase (BC) at diagnosis prior to treatment. Bulk cells were flow sorted into rigorously defined populations using surface phenotypes. Gene expression of HSCs ($Lin^-CD34^+CD38^-CD45RA^-CD90^+$) and MPPs ($Lin^-CD34^+CD38^-CD45RA^-CD123^+$) from 3 healthy donors and 12 CML patients were extracted in correlation to our differential gene analysis in HSCs. Comparison of gene expression in CML HSC/MPPs *vs.* normal HSC/MPPs in this dataset also identified *GAS2* as among the most differentially expressed gene (false discovery rate (FDR)-adjusted $P < 0.01$; **Figure 1.1H, Supplementary Figure 3**). I then compared the expression of *GAS2* in HSCs from CML patients and healthy donors, CML HSCs had significant higher *GAS2* expression than normal HSCs (**Figure 1.1I**, Wilcoxon test, $P < 0.01$). These analyses led me to prioritise *GAS2* for functional validation as a potential candidate for targeting CML HSCs.

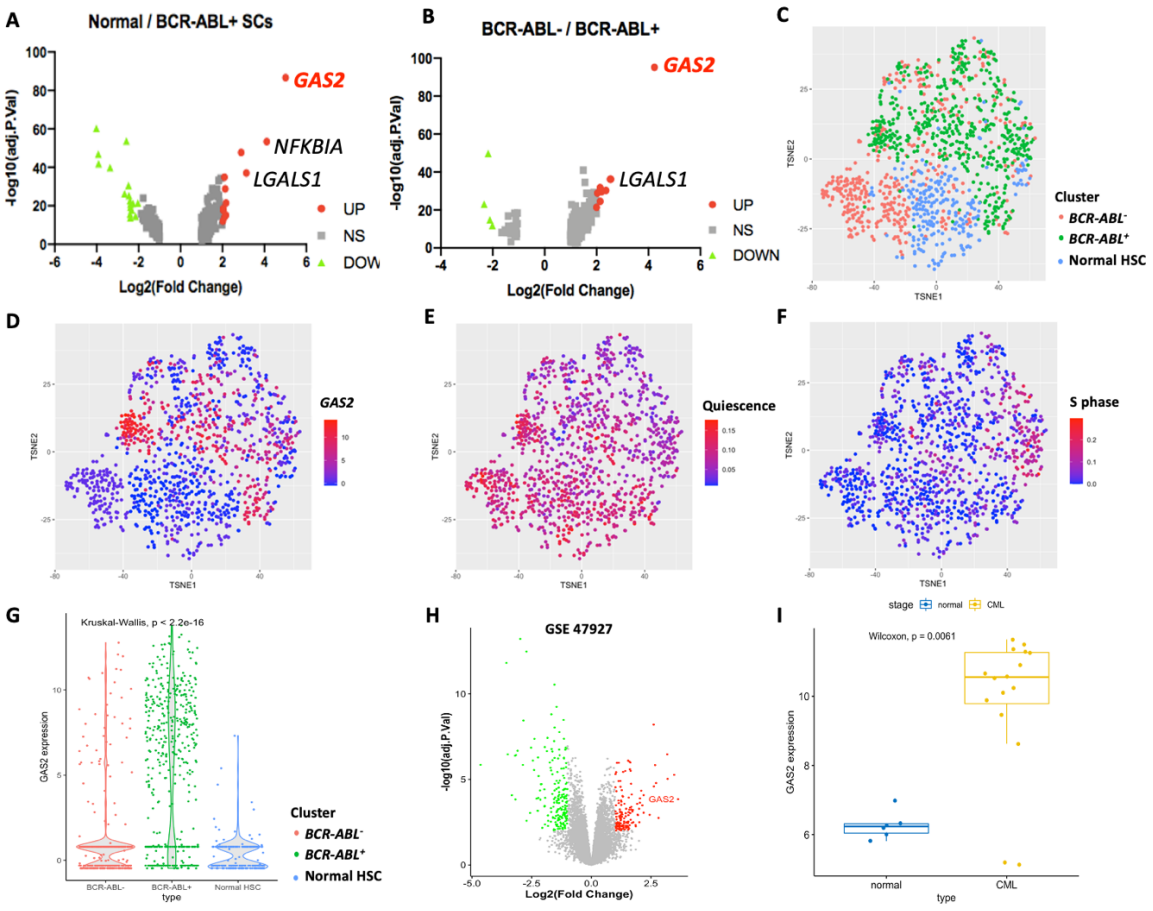


Figure 1.1 *Growth-Arrest Specific 2 (GAS2)* is highly expressed in CML HSCs

A. Volcano plot of differentially expressed genes with at least a 2-fold difference in expression and false discovery rate [FDR]-adjusted $P < 0.01$ between *BCR-ABL1*$^{+}$ HSCs *vs.* healthy donor HSCs.

B. Volcano plot of differentially expressed genes with at least a 2-fold difference in expression and false discovery rate [FDR]-adjusted $P < 0.01$ between *BCR-ABL1*$^{+}$ HSCs *vs. BCR-ABL1*$^{-}$ HSCs.

C. T-distributed stochastic neighbour embedding (tSNE) visualization of healthy donor HSCs (blue dots; n = 232), *BCR-ABL1*$^{-}$ HSCs (red dots; n = 377) and *BCR-ABL1*$^{+}$ HSCs (green dots; n = 477) using 8589 highly variable genes from the CML dataset [1].

D. Imposition of expression of *GAS2* on tSNE clustering of single-cell CML dataset [1].

E. Imposition of gene set score of quiescence on tSNE clustering of single-cell CML dataset.

F. Imposition of gene set score of cell cycle S phase on tSNE clustering of single-cell CML dataset.

G. Violin plot of *GAS2* expression in different clusters of HSCs from single-cell CML dataset [1]. Kruskal-Wallis test *P* values are shown above the graph.

H. Volcano plot of differentially expressed genes with a minimum 2-fold difference and false discovery rate [FDR]-adjusted $P < 0.01$ between CML and healthy donors HSCs (data extracted from GSE47927). *GAS2* shows highest differential expression.

I. Boxplot of *GAS2* expression in HSCs from CML patients and healthy donors (data extracted from GSE47927). The boxes represent the median and quartiles of gene expression levels. Wilcoxon test *P* values are shown above the graph.

2. Optimization of *GAS2* knock-out/knock-down in cell lines and CML CD34$^+$ cells

2.1. Optimization of inducible lentiviral transduction in cell lines and CML CD34$^+$ cells

2.1.1. *GAS2* knock-down in RPMI8226 cell line by inducible lentivirus transduction

To functionally validate the role of candidate genes in primary cells, a tool to efficiently manipulate gene expression is needed. First, I tested experimental techniques for transduction of lentiviral shRNA vectors using human malignant haematopoietic cell lines. For this I needed a cell line with high expression of *GAS2*.

Therefore, I performed RT-qPCR on 6 available human malignant haematopoietic cell lines: HEL (erythroblast from erythroleukaemia), SET2 (megakaryoblastic cell line), K562 (CML cell line), Jurkat (leukaemic T cell lymphoblast), U266 (lymphoblast from multiple myeloma), RPMI8226 (lymphoblast from multiple myeloma). Expression level of *GAS2* was the highest in U266 cell line, around 700 times of HEL cell line, followed by Jurkat and RPMI8226 cell lines. K562 cell line expressed the lowest level of *GAS2*, and SET2 had similar expression level as HEL (**Figure 2.1**), this mRNA level of *GAS2* correlated well with the database of human protein atlas (Details of the GAS2 protein expression levels in all cell lines can be seen in **Supplementary Figure 4**). https://www.proteinatlas.org/ENSG00000148935-GAS2/cell#rna.

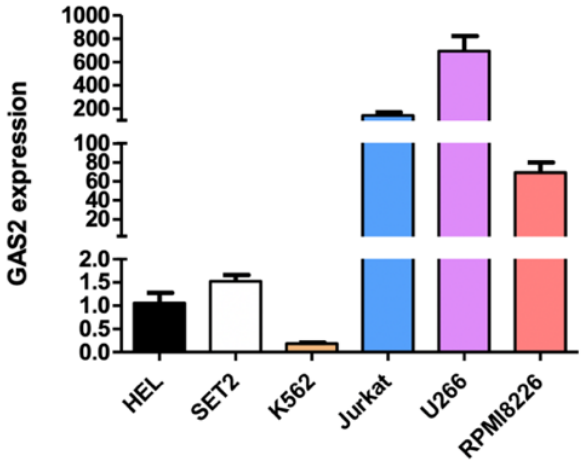


Figure 2.1. Relative expression of *GAS2* in human malignant haematopoietic cell lines (compared to HEL) (technical replicates n = 3, bars represent mean ± SD).

2.1.2. *GAS2* knock-down by 83% in RPMI8226 cell line by inducible lentivirus transduction

Details of inducible lentiviral shRNA vectors are in Methods section. A9, A10 and B10 are 3 shRNA targeting different sites of *GAS2,* sequences can be seen in

Appendix 1. To determine the efficiency of gene knock-down using lentiviral shRNA vectors, I first calibrated the titration of virus construct B10 by standard serial dilutions using the HEK-293T cell line (**Figure 2.2 A**). The titration of virus construct B10 was calculated by the formula listed in Methods, 9.82×10^7 TU/ml. The 1:250 dilution from all 3 lentiviral shRNA vectors targeting *GAS2* (A9, A10, B10) and the non-targeting control shRNA were used for transduction in RPMI8226 cell line for optimal MOI = 0.4. Cells with a 10-fold higher level of RFP intensity (relative to cells without virus transduction) were sorted from each virus transduction to determine the impact of shRNA transduction on *GAS2* expression level (**Figure 2.2 B**). Transduction with virus B10 resulted in the most dramatic decrease in *GAS2* expression, to 17% compared to virus nontarget control, while the other two shRNAs decreased expression to 31% and 67% respectively (technical replicates, n = 3, **Figure 2.2 C**).

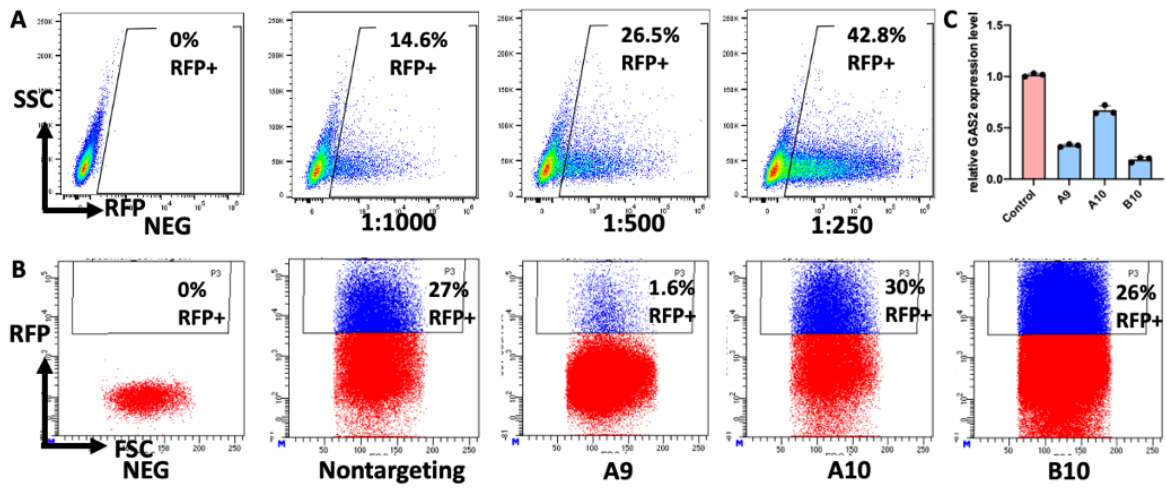


Figure 2.2 Serial dilution of virus transduction in HEK293-T and RPMI8226 cell line (technical replicates n = 3, bars represent mean ± SD).

A. Flow cytometry plots of RFP percentage post transduction of serial diluted virus construct B10 into HEK293-T cell line. Four conditions are tested: 1.NEG-cells

without virus transduction; 2. Cells transduction with 1:1000 virus dilution; 3. Cells transduction with 1:500 virus dilution; 4. Cells transduction with 1:250 virus dilution. The sorting gate of RFP+ cells is set according to NEG condition.

B. Flow cytometry plots showing sorting gate for transduced RFP^+ RPMI8226 cells with different lentiviral shRNA vectors. Five conditions are listed: 1. NEG represents cells without transduction, where the gate for RFP^+ signal is set 10 times higher than the top border of negative cells; 2. Cells transduction with shRNA for non-targeting control; 3. Cells transduction with shRNA A9 targeting *GAS2*; 4. Cells transduction with shRNA A10 targeting *GAS2*; 5. Cells transduction with shRNA B10 targeting *GAS2*.

C. *GAS2* expression mRNA levels (relative internal control of *GAPDH*) of different lentiviral vectors transduction in RPMI8226 cells by q-RTPCR, all cells are sorted by RFP signal after transduction for RNA extraction and q-RTPCR. Control sample is cells transduced with shRNA for non-targeting control. The test samples are cells transduced with shRNA A9, A10 or B10 targeting *GAS2*.

2.1.3. Low transduction efficiency in CML $CD34^+$ cells via inducible lentivirus transduction

Therefore, the efficiency of virus B10 and the non-targeting control virus were next tested on human CML $CD34^+$ cells (**Figure 2.3 A**), which are enriched from peripheral blood mononuclear cells by CD34 MicroBead Kit (Miltenyi Biotec), details are listed in the Methods Chapter. **Figure 2.3.A** showed the detailed experimental diagram: enriched $CD34^+$ cells were transduced with inducible lentiviral vectors on day 1; doxycycline was added to induce shRNA for *GAS2* or non-targeting control knock-down; 48 hours after transduction on day 3, CellTrace dye-CFSE was stained, followed by 150nM dasatinib (DA) or DMSO treatment on day 4 for 3 days; On day

7, transduced $CD34^+RFP^+$ cells were sorted into $CFSE^{max}$-quiescent cells, and $CFSE^{low}$-dividing cells for *GAS2* expression detection.

Because the human sample is very precious, I did this more complicated preliminary experiment instead of only validating the virus knock-down efficiency. *GAS2* expression level in quiescent and dividing cells can be compared, as well as the effect of TKI (Dasatinib (DA)) on *GAS2* expression in this preliminary experiment. It appeared that: **1.** *GAS2* expression was 3-fold higher in the quiescent than dividing $CD34^+$ cells (**Figure 2.3 B**); **2.** Despite the low transduction efficiency of this inducible lentivirus (7.06%, **Figure 2.3 C**), *GAS2* expression level was significantly lower in both quiescent and dividing CML $CD34^+$ cells following transduction with shRNA to *GAS2* than that in cells transduced with the non-targeting control (average *GAS2* mRNA levels: Quiescent $shRNA^{GAS2}$ + DMSO *vs.* quiescent $shRNA^{Ctrl}$ + DMSO = 0.35 *vs.* 1, dividing $shRNA^{GAS2}$ + DMSO *vs.* dividing $shRNA^{Ctrl}$ + DMSO = 0.12 *vs.* 0.33, **Figure 2.3 D**); **3.** DA did not have a significant effect on *GAS2* expression when *GAS2* was not knocked down, and *GAS2* expression was undetectable in cells when both DA and shRNA targeting *GAS2* were added. C_T values of *GAS2* were undetectable but values of *GAPDH* remained stable in $GAS2^{KD}$ + DA treatment group (**Figure 2.3 D**, **Appendix 3, Supplementary Table 1**); 4. The cell viability was not very good on the sorting day, resulting in few cells for expression detection by RT-qPCR. The percentage of DAPI- cell was 33.5% in "$shRNA^{Ctrl}$ + DMSO", 5.9% in "$shRNA^{Ctrl}$ + DA", 36.8% in "$shRNA^{GAS2}$ + DMSO", 10.5% in "$shRNA^{GAS2}$ + DA" (Example flow cytometry plot shown in **Figure 2.3 E**).

To solve the problem of low transduction efficiency using inducible lentivirus, I made several adjustments on transduction settings, including different transduction enhancer (retronectin coating or polybrene), spinoculation temperature, consecutive 2nd transduction. Other basic settings are all the same: All transductions were performed in 2.8 × 10^5 cells per-well in the 48-well plate with 400 μl medium, virus load MOI = 3.28 (detailed transduction steps in Methods Chapter). **Table 1** summaries all different Optimisation methods. The highest transduction efficiency achieved in CML CD34+ cells (7.06%) was still substantially lower than what had been achieved in cell lines (52.1%) (**Figure 2.3 C**, detailed transduction efficiencies under each condition are in **Table 1**).

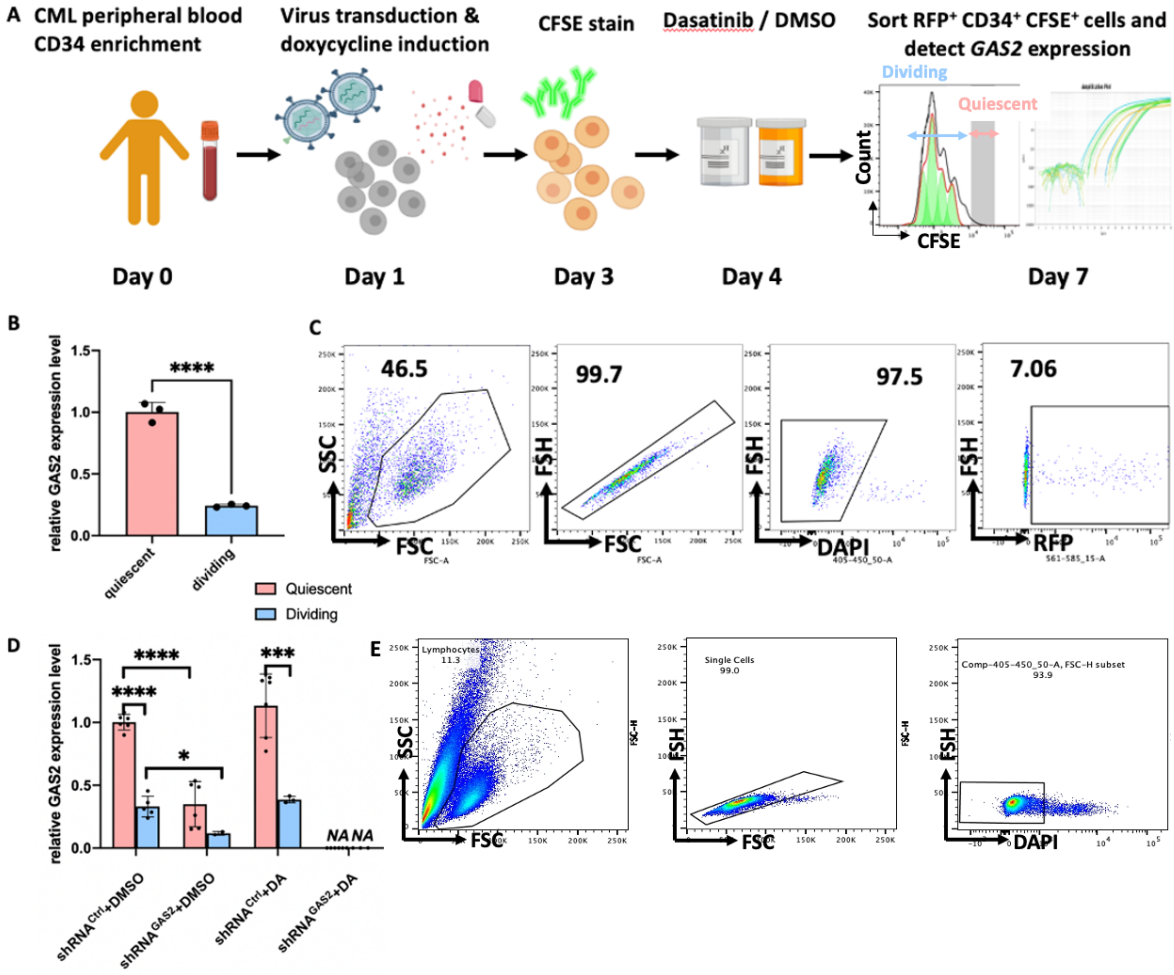


Figure 2.3. Detection of *GAS2* expression and virus transduction in CML $CD34^+$ cells (n=6, bars represent mean ± SD)

A. Experimental flowchart of CML $CD34^+$ cells. Peripheral blood $CD34^+$ cells were enriched from CML patient, virus transduction and induction were performed after cells were thawed, CFSE was stained for intracellular dynamic tracing and DA or DMSO was added for 3 days before the flow cytometry sort was performed, both quiescent (pink gating line) and dividing (blue gating line) RFP^+ $CD34^+$ cells were sorted for GAS2 expression detection.
B. Bar plot of relative *GAS2* expression levels of quiescent and dividing CML $CD34^+$ cells by RT-qPCR (technical replicates, n=3, bars represent mean ± SD, **** $p<0.0001$, unpaired t test).
C. Example flow cytometry plot showing the gating for the successfully transduced CML cells.
D. Bar plot of relative *GAS2* expression levels of quiescent and dividing RFP^+ CML $CD34^+$ cells in $shRNA^{Ctrl}$ + DMSO, $shRNA^{GAS2}$ + DMSO, $shRNA^{Ctrl}$ + DA, $shRNA^{GAS2}$ + DA by RT-qPCR (n=6, bars represent mean ± SD, * $p<0.5$, *** $p<0.001$, **** $p<0.0001$, unpaired t test).
E. Example flow cytometry plot of $DAPI^-$ cells percentage in experiment condition $shRNA^{GAS2}$ + DA.

Table 1. Optimisation of transduction and efficiencies in CML $CD34^+$ cells.

Group	Transduction condition	RFP%
1	Retronectin coating, +2/3 virus + cells, centrifuge at 25°C 2000rpm 1hr; immediately + 1/3 virus + 1X polybrene after centrifuge	5.22
2	Retronectin coating, +2/3 virus + cells, centrifuge 25°C 2000rpm 1hr; the 2^{nd} day + 1/3 virus + 1X polybrene, centrifuge 25°C 2000rpm 1hr	6.13
3	Retronectin coating, +virus + cells, centrifuge 25°C 2000rpm 1hr,	2.45
4	No retronectin, + 1X polybrene + virus + cells, centrifuge 25°C 2000rpm 1hr	7.06
5	Retronectin coating, +2/3 virus + cells, centrifuge 4°C 2000rpm 1hr; immediately + 1/3 virus + 1X polybrene after centrifuge	5.11
6	Retronectin coating, +2/3 virus + cells, centrifuge 4°C 2000rpm 1hr; the 2^{nd} day + 1/3 virus + 1X polybrene, centrifuge 25°C 2000rpm 1hr	6.41
7	Retronectin coating, +virus + cells, centrifuge 4°C 2000rpm 1hr,	4.14
8	No retronectin, + 1X polybrene + virus + cells, centrifuge 4°C 2000rpm 1hr	5.23

2.1.4. Troubleshooting for low transduction efficiency of *GAS2*-virus in CML CD34$^+$ cells

I first tried to determine whether the low transduction efficiency in CML primary cells was related to the specific patient sample I had tested or to the virus. The same transduction assay was repeated in the same sample CML-1 and a new sample CML-2 with the same virus targeting *GAS2*. In addition, a smaller size of 5kb GFP virus was transduced into these two samples. Flow cytometry plots showed that the transduction efficiency in CML-2 sample was lower than CML-1 sample in both *GAS2*-virus group (0.88% *vs.* 2.6%) and GFP virus group (31.3% *vs.* 43.6%). Much higher transduction efficiency was found with the GFP virus (43.6% & 31.3%) than *GAS2*-virus (2.6% & 0.88%) in both samples (**Figure 2.4**).

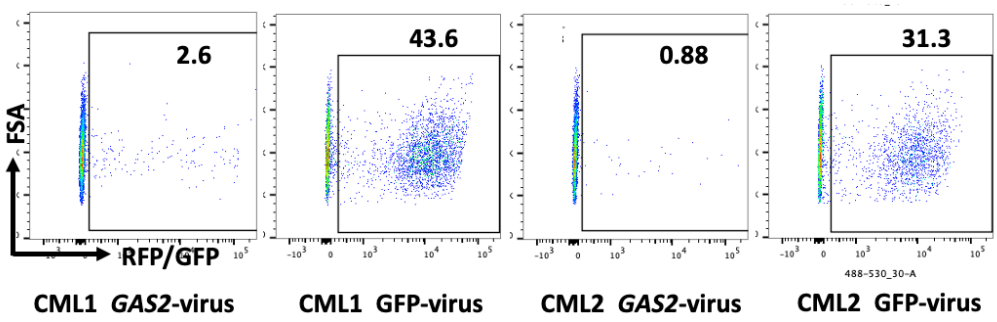


Figure 2.4. Flow cytometry plot showing percentage of transduced cells within CML CD34$^+$ cells (n=1)

Therefore, the same efficiency of transduction was not achieved in CML CD34$^+$ cells as in cell lines. A likely reason was that the lentiviral vector was too large for CML primary cells, given the fact that a smaller size virus could achieve much higher percentage of RFP$^+$ CD34$^+$ cells (43.6%) in the same patient. Human primary cells

are more complicated in cell metabolism, nucleus-cytoplasm exchange and are harder for large virus to transfect. Other gene manipulation approaches were needed to get efficient knock-down. Then I designed a smaller size virus shRNA vector targeting *GAS2*. During the time of producing the new virus, CRISPR-Cas9 genome editing was therefore next explored, and the most efficient delivery method by pre-assembled Cas9-sgRNA RNP complex was applied. The knock-out strategy of targeting an essential exon of *GAS2* using one sgRNA was applied first.

2.2. Optimisation of GAS2 knock-out using single RNP CRISPR-Cas9 nucleofection

I tested 3 sgRNAs targeting different locations of exon 3 in *GAS2* from Lonza using an optimised nucleofection protocol on K562 cell line. Cells remained 75% viability post nucleofection, which was much better than what I achieved previously in lentivirus transduction of CML $CD34^+$ cells (39% viability) and cell lines (44% viability). Equivalent high level of allele modification frequency was noted in both *AAVS1* knock-out (set as positive control) and one of the *GAS2* knock-out sgRNAs (**Figure 2.5 A**). The other 2 sgRNAs had much lower editing efficiency (**Figure 2.5 A**).

I next sought to find out whether *GAS2* mRNA and protein levels were also effectively reduced in K562 cells after nucleofection. Six days after nucleofection, RNA and protein were extracted from K562 cells for q-RTPCR and western blotting. There were only 20% decrease in *GAS2* mRNA and 28% decrease in GAS2 protein

level in *GAS2*^{KO} cells compared to *AAVS1*^{KO} cells (**Figure 2.5 B, 2.5 C,** the uncropped western blot figures are shown in **Supplementary Figure 5**).

Nucleofection in CML CD34^{+} cells with Cas9-gRNA RNP targeting *GAS2* was performed as well, using the same processing protocol as in K562 except that the specific K562 Lonza buffer and program (SF kit, FF-120 program) were switched to specific human CD34^{+} cells buffer and program (P3 kit, EO-100 program). Similar allele modification frequency of three sgRNAs was found in primary cells compared to cell lines (**Figure 2.5.A**).

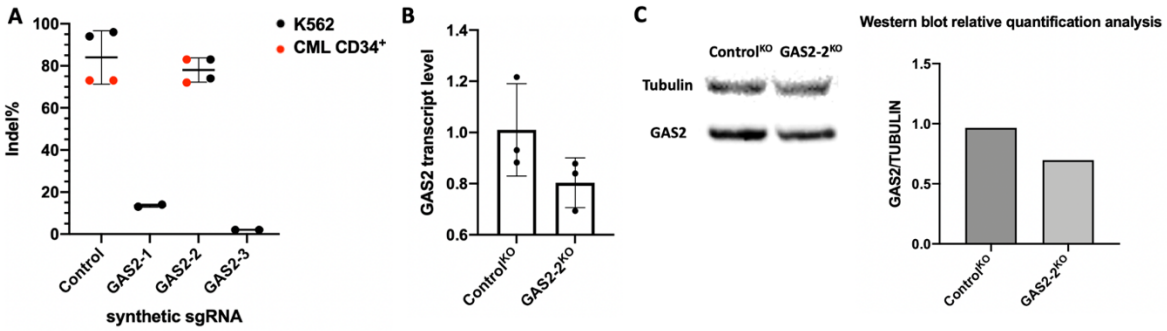


Figure 2.5. *GAS2* knock-out by CRISPR RNP approach and validation by mRNA, protein levels.

A. Allele modification frequency of cells post nucleofection of Cas9-gRNA RNP including changes of both small insertion and deletion, dots represent biological replicates, bars represent mean ± SD.
B. Bar plot of *GAS2* expression level (internal control *GAPDH*) in K562 *GAS2*^{KO} relative to *AAVS1*^{KO} (positive control) cells by RT-qPCR (technical replicates, n=3, bars represent mean ± SD).
C. Western blot analysis of GAS2 level in K562 cells transfected with a AAVS1 sgRNA (positive control) or the most efficient GAS2-targeting sgRNA quantified and compared to TUBULIN 7 days post-nucleofection. (The uncropped western blot figures are shown in **Supplementary Figure 5**)

2.3. GAS2 knock-out using dual RNPs CRISPR-Cas9 nucleofection

Since single RNP CRISPR knock-out in K562 cells could not achieve a significant reduction in *GAS2* mRNA level or protein level, then I tried dual RNPs targeting both the head and tail of exon 3 region in *GAS2* on K562 cells (**Figure 2.6 A**). Forty-eight hours after nucleofection, DNA was extracted from cells and amplified using primers over exon 3. If both head and tail of exon 3 were successfully knocked-out within 1 cell, then genotype would show a clear shorter band compared to wild-type control. Single cells were sorted into 60-well Terasaki plate from knock-out cells. 21 days after expanding the culture, 71 single cell clones were genotyped, 33 clones had shorter band, among which clone 2 and clone 21 had the clearest shorter bands (**Appendix 2, Supplementary Figure 6**). Then I picked up 4 heterozygous clones (clone 3, 4, 20, 23) together with clone 2 and clone 21. After expanding the colonies, genotyping was repeated, and clone 21 was found to have the purest shorter band (**Figure 2.6 B**). GAS2 protein level was assessed among the knock-out cell clones. Although better than previous single RNP knock-out clones (**Figure 2.5 C**), very weak expression of *GAS2* was still observed (**Figure 2.6 C**).

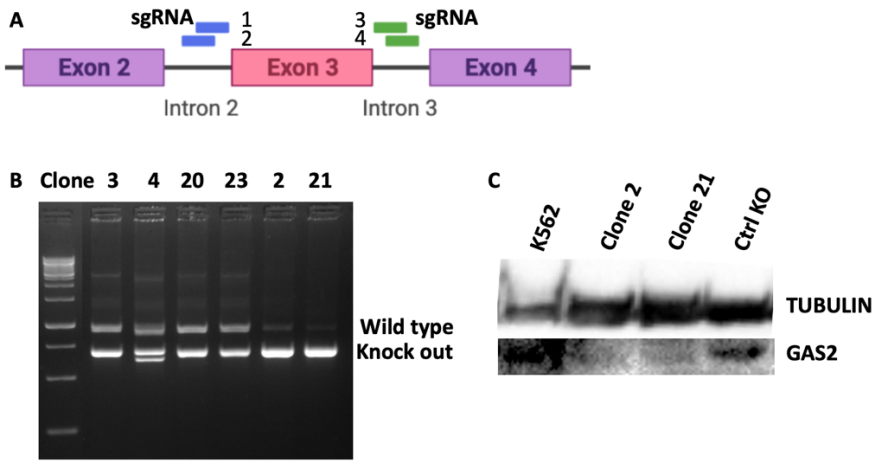


Figure 2.6. *GAS2* knock-out by CRISPR dual RNPs approach.

A. Diagram of CRISPR dual RNPs approach. One sgRNA targeting intron 2 and another sgRNA targeting intron 3 add Cas9 nucleases separately to form RNPs, and these 2 RNPs work together for nucleofection.

B. Genotyping of *GAS2* Exon 3 regions covering RNP1 and RNP3 cutting sites in *GAS2* knock-out single cell clones. Wild-type band is 934 bp, and knock-out band is 663 bp.

C. Western blot analysis of GAS2 levels in wild-type, knock-out clone 2, clone 21 and AAVS1-control knock-out K562 cells, TUBULIN is used as internal controls.

2.4. Optimisation of GAS2 knock-down using constitutively expressed lentiviral shRNA vectors

To get a pure *GAS2*$^{KO/KD}$ population, a selection marker is needed. Although the CRISPR-Cas9 approach was able to achieve a high percentage of knock-out in cell lines, the knock-out population cannot be selected. Therefore, I next used a smaller size lentiviral vector (pLKO_eYFP, **Appendix 2, Supplementary Figure 2**) with an e-YFP reporter for constitutive *GAS2* knock-down. Three shRNA vectors were

tested first on K562 cells (sequences see **Appendix 1**). The transduction efficiency was over 90% (**Figure 2.7 A**). The e-YFP$^+$ cells were sorted and *GAS2* mRNA level was measured via q-RT-PCR. The shRNA3-*GAS2* achieved over 90% knock-down in *GAS2* mRNA in K562 cells (**Figure 2.7 B**).

I next tested whether this smaller lentiviral vector could achieve a higher transduction efficiency in primary CML CD34$^+$ cells. 48 hours after transduction, I analysed the cells by flow cytometry: over 76% cells remained viable, and the percentage of e-YFP$^+$ cells, reflecting the transduction efficiency, was over 62.5% (sh-Ctrl 57.7%, sh3-*GAS2* 67.2%, **Figure 2.7 C**). *GAS2* mRNA level in sh3-*GAS2* group was 86% lower than that in sh-Ctrl group (in quiescent CD34$^+$ cells under DMSO treatment) (**Figure 2.7 D**). And cell viability on the sorting day in all treatment groups were relatively good, the percentage of 7-AAD$^-$ cell was 54.6% in "sh-Ctrl + DMSO", 25.9% in "sh-Ctrl + Imatinib", 48.1% in "sh3-*GAS2* + DMSO" and 24.3% in "sh3-*GAS2* + Imatinib" (the example flow cytometry plot was shown in **Figure 2.7 E**). Therefore, I successfully developed a strategy to significantly knock down *GAS2* in primary CD34$^+$ cells from CML patients, with a fluorescent reporter to enable sorting of cells for downstream functional assays.

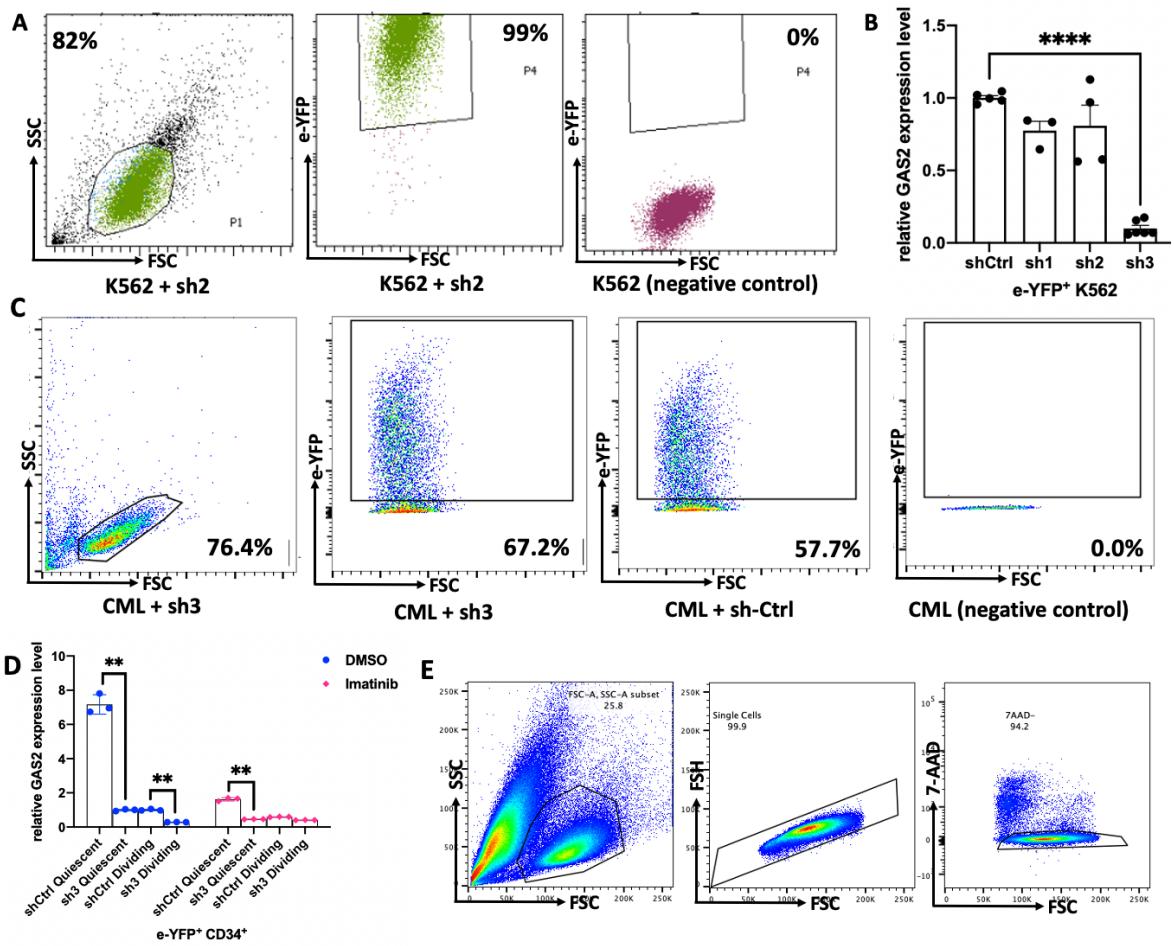


Figure 2.7. *GAS2* expression level in K562 and peripheral blood CML CD34$^+$ cells by constitutive shRNA knock-down.

A. Example flow cytometry plots of K562 transduction with sh2 vector and negative control (K562 without shRNA vector). There were over 80% live cells after transduction, and over 90% e-YFP$^+$ cells *vs.* the negative control.

B. Bar plot of relative *GAS2* expression levels (internal control *GAPDH*) in K562 cells after transduction. Different lentiviral vectors targeting *GAS2* were compared, *GAS2* level of sh1, sh2 and sh3 vector groups was normalized to shCtrl group (set as 1). e-YFP$^+$ cells were sorted for RNA extraction and followed by RT-qPCR (n=6, bars represent mean ± SD, **** $p<0.0001$, unpaired t test).

C. Example flow cytometry plots of 2 days after CML $CD34^+$ cells transduction with sh3 vector targeting *GAS2*, sh-Ctrl vector targeting *AAVS1*, and negative control (CML $CD34^+$ cells without shRNA vector).

D. Bar plot of relative *GAS2* expression level (internal control *GAPDH*) detected by RT-qPCR in different treatment groups, all experimental groups were normalized to "shCtrl Dividing" cells group (set as 1) (n=3, bars represent mean ± SD, ** $p<0.01$, unpaired t test).

E. Example flow cytometry plot of cell viability ($7\text{-}AAD^-$ cells) in "sh3 + Imatinib" group in figure D.

In summary, 4 gene manipulation strategies were applied to *GAS2* knock-down or knock-out (**Table 1**), *i.e.*, **1.** Inducible lentiviral shRNA transduction, **2.** Constitutively expressed lentiviral shRNA transduction, **3.** Single RNP nucleofection, **4.** Dual RNPs nucleofection.

The following factors are taken into consideration to select the promptest strategy: ① Transduction efficiency, ② Cell viability, ③ *GAS2* knock-down or knock-out level, ④ Selection markers.

Constitutively expressed lentiviral shRNA best fits all the selection criteria, so I finally chose constitutively expressed lentiviral shRNA transduction for further experiments.

Table1. Summary of different gene manipulation strategies in cell lines and CML CD34$^+$ cells.

Method	Cell type	Efficiency	Viability	*GAS2* reduction level
Inducible lentiviral shRNA transduction	RPMI8226	52.1%	50.4%	83% reduction in mRNA
	CML CD34$^+$	7.06%	46.5%	65% reduction in mRNA
Constitutive lentiviral shRNA transduction	K562	97.9%	75.3%	90% reduction in mRNA
	CML CD34$^+$	67.2%	76.4%	86.2% reduction in mRNA
Single RNP nucleofection	K562	Indel 78.5%	76.5%	20% reduction in mRNA and protein
	CML CD34$^+$	Indel 77.5%	73.8%	NA
Dual RNPs nucleofection	K562	Indel 48%	86.4%	NA, requires single cell clone selection

3. Preliminary functional validation of *GAS2* knock-down/knock-out in cell line and CML primary cells

3.1. Preliminary test of GAS2 knock-down in cell line

Cell cycle analysis was performed in RPMI8226 cells following *GAS2* shRNA knock-down (**Figure 3.1 A**). Following *GAS2* knock-down, RPMI8226 cells had a slightly higher percentage of G_2-S-M phase cells compared to control (44.8% *vs.* 32.8%, n = 1) (**Figure 3.1 A**). Following *GAS2* knock-out, K562 cells had a slightly higher proliferation index (PI) and division index (DI) than control (PI: 2.61 *vs.* 2.32, DI: 2.61 *vs.* 2.29, n = 1, **Fig 3.1 B**) in cell proliferation assay. These suggested that *GAS2* knock-down could cause cell cycle changes and increase cell proliferation, though more replicates were needed for the validity. *GAS2* knock-out did not alter early or late apoptosis in K562 cells (early apoptosis: 4.39% *vs.* 5.48%, late apoptosis: 2.38% *vs.* 1.75%, n = 1, **Fig 3.1 C**).

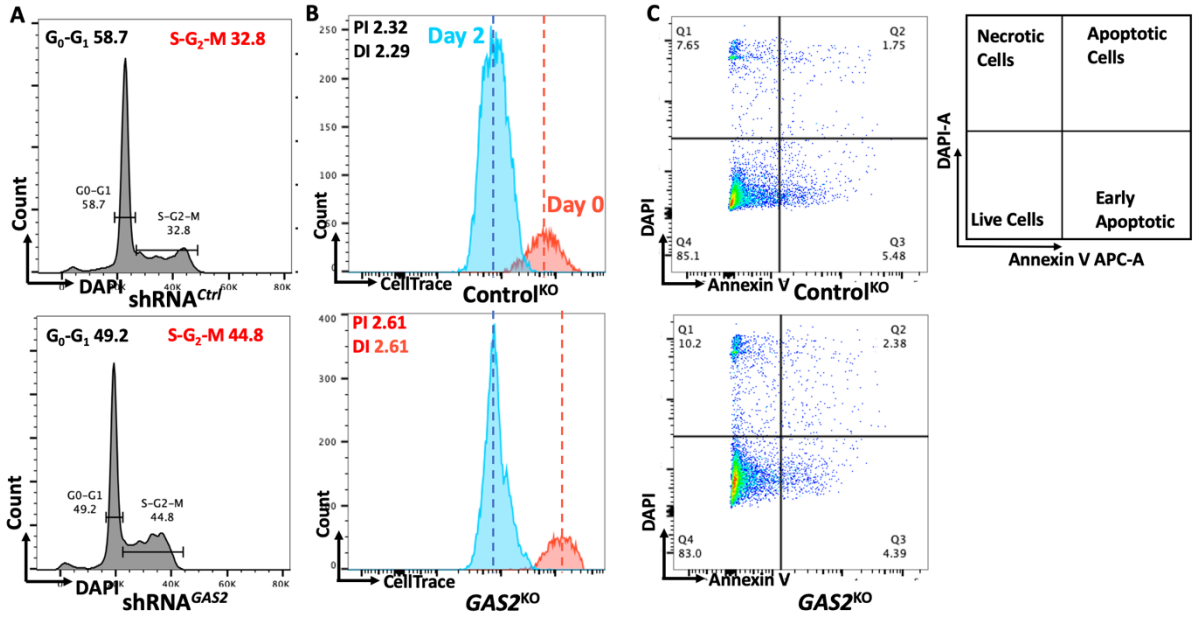


Figure 3.1. Functional validation of *GAS2* in cell lines (n = 1)

A. Histogram of cell cycle analysis in *GAS2*^{KD} RPMI8226 cells, the x axis is DAPI signal in the cells, the first peak represents G_0-G_1 phase, the rest is S-G_2-M phase (n = 1).
B. Histogram of cell proliferation analysis in *GAS2*^{KO} K562 cells, the x axis is CellTrace dye signal, the red peak on the right represents day 0 signal, the blue peak on the left is day 2 signal within the cell, the dashed line represents the mean fluorescence index, the PI and DI are calculated by FlowJo Cell proliferation platform (n = 1).
C. Dot plot of apoptosis analysis in *GAS2*^{KO} K562 cells. Annexin-V$^+$ DAPI$^-$ cells (lower right) were early apoptotic cells, Annexin-V$^+$ DAPI+ cells (upper right) were late apoptotic cells (n = 1).

3.2. GAS2 knock-down increased proliferation in CML stem cells

To evaluate the combined effect of *GAS2* knock-down and TKI treatment on CML HSPCs, I designed and performed the following experiment: CML CD34$^+$ cells were thawed and transduced with the lentiviral shRNA vectors (sh-control or sh3-*GAS2*). 2 days after transduction, cells were washed twice to remove the virus. Then, cells were stained with CellTrace dye, which can enter cells and halves as cell divides, so that the quiescent cells obtain the strongest signal and are easily recognized. TKI or DMSO treatment was added for 3 days before FACS-sorting (**Figure 3.2.A**). Cells' viability was reasonable and a good transduction efficiency (eYFP$^+$ percentage) was achieved – around 39%. This experiment was repeated 3 times using 3 different CML patient samples, and the gating strategy is shown in **Appendix 2 Supplementary Figure 6**. CML Lin$^-$CD34$^+$CD38$^-$ cells showed a higher proliferation index (PI) following *GAS2* knock-down as compared to the sh-control group (average PI: 1.73 *vs.* 1.53, $p < 0.05$, paired t test, **Figure 3.2 B**). No significant change in apoptosis was noted in the *GAS2*^{KD} *vs.* control^{KD} cells (**Figure 3.2 C**).

Cells were counted using trypan blue before sorting, and the absolute number of quiescent CML Lin$^-$CD34$^+$CD38$^-$ cells was calculated based on the FACS gating and CellTrace dye. The *GAS2*^{KD} + Imatinib (IM) group had the lowest quiescent CML Lin$^-$CD34$^+$CD38$^-$ cell numbers among all groups, although this change was not statistically significant (**Figure 3.2 D**). Cells' viability was relatively good, **Figure 3.2 E** summarised the percentage of 7-AAD$^-$ cells in 4 treatment groups during FACS-sorting on Day 7. The DMSO groups had relatively higher percentage of viable cells *vs.* IM groups, no statistically significant differences were noted among 4 groups.

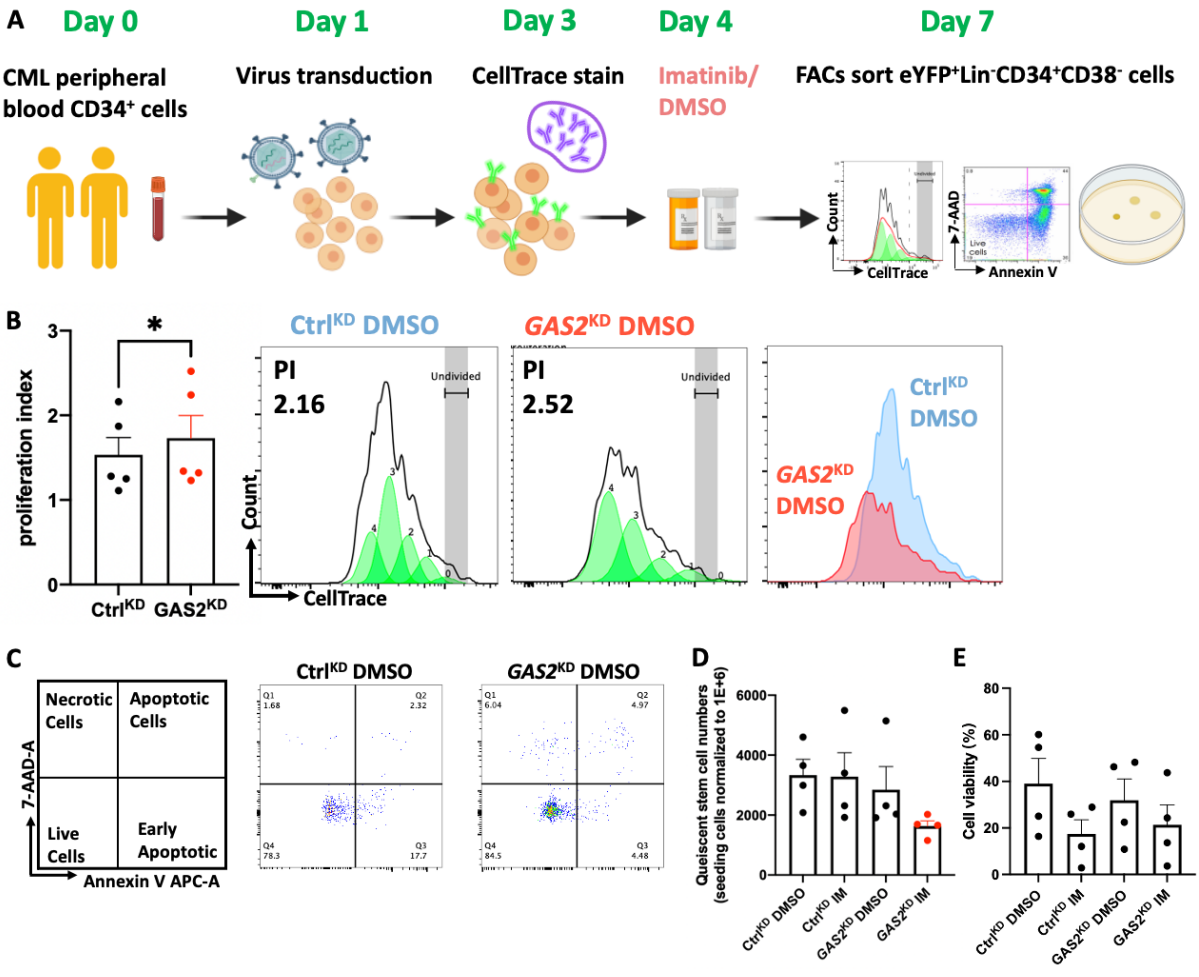


Figure 3.2. Functional validation of *GAS2* knock-down in CML stem cells (Lin$^-$ CD34$^+$CD38$^-$ cells)

A. Experiment flowchart of *GAS2* knock-down and downstream functional analysis in primary cells.

B. Bar plot of proliferation index of CML stem cells in $Ctrl^{KD}$ and $GAS2^{KD}$ group. (n=5, bars represent mean ± SD, * $p<0.05$, paired t test). Example figure of the proliferation assay histogram of stem cells from Sample-150 by FlowJo platform, the right figure is the stack view of $Ctrl^{KD}$ + DMSO (blue) and $GAS2^{KD}$ + DMSO (red) group figure.

C. Example dot plot of apoptosis analysis of CML stem cells from $Ctrl^{KD}$ + DMSO and $GAS2^{KD}$ + DMSO group of sample-201. Annexin-V^+ 7-AAD^- cells (lower right) were early apoptotic cells, Annexin-V^+ 7-AAD^+ cells (upper right) were late apoptotic cells.

D. Bar plot of absolute quiescent stem cell numbers from different CML patient samples, the final quiescent stem cell numbers were normalized to the same 1,000,000 seeding cells on day 0 in each group and each replicate. (n = 4, bars represent mean ± SD, unpaired t test not statistically significant).

E. Bar plot of cell viability (percentage of 7-AAD^- cells) in 4 treatment groups during FACS-sorting on Day 7. (n = 4, bars represent mean ± SEM, unpaired t test not statistically significant).

3.3. GAS2 knock-down had no significant effect on lineage differentiation of CML stem cells

To test lineage differentiation potential following *GAS2* knock-down, CML $Lin^-CD34^+CD38^-$ e-YFP^+ cells in $Ctrl^{KD}$ + DMSO, $Ctrl^{KD}$ + IM, $GAS2^{KD}$ + DMSO, and $GAS2^{KD}$ + IM group on day 7 (**Figure 3.2 A**) were sorted into methylcellulose-based medium (MethoCult™ H4435 + 1% (v/v) PS) for CFU assays (500 cells per dish, 3 replicates per group). The majority of colonies were CFU-M, with very few CFU-GM,

CFU/BFU-E, CFU-GEMM (**Figure 3.3 A-C, example morphologies in Figure 3.4**). There was a slightly trend that *GAS2*^{KD} + IM group on average grew more colonies per dish than DMSO groups after 14 days culture (**Figure 3.3.D**). *GAS2*^{KD} + IM cells also had the lowest percentage of CFU-M colonies among all colonies (**Figure 3.3 E**), although this difference was not statistically significant.

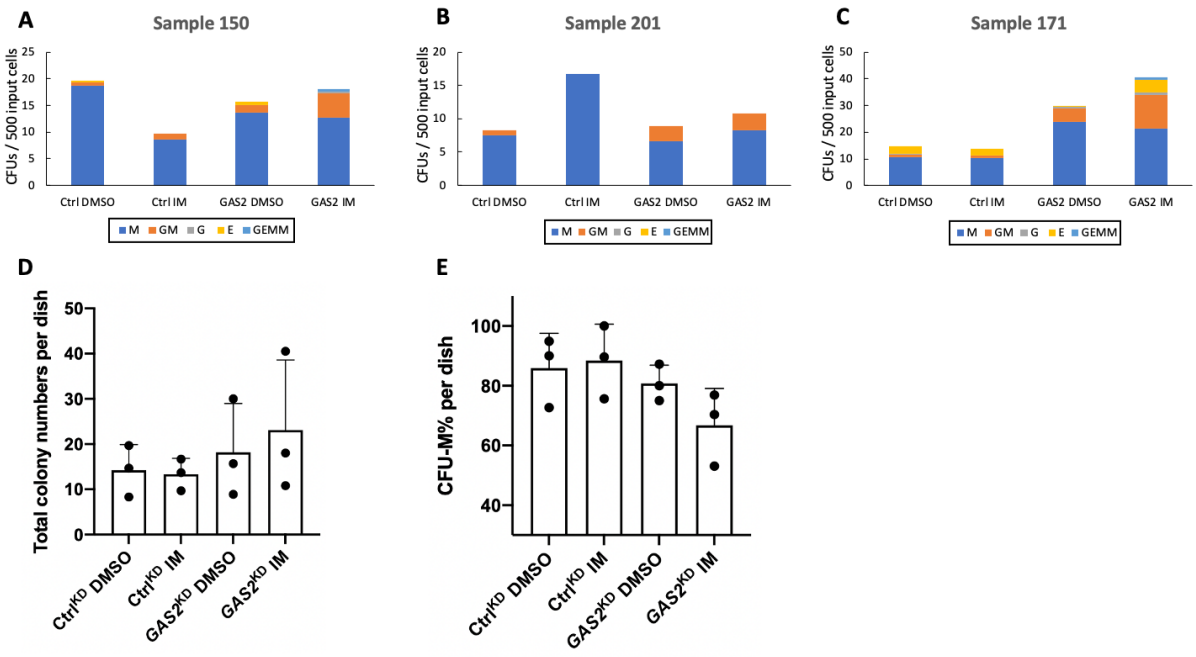


Figure 3.3 Colony forming unit (CFU) assays from CML stem cells

A-C. Total colonies per 500 input cells CML patient (Sample-150 (A), 201 (B), 171 (C),) after 13 days culture. Different colour bars represent different quantity of colonies: dark blue is CFU-M, orange is CFU-GM, grey is CFU-G, yellow is CFU/BFU-E, light blue is CFU-GEMM, the mean value from triplicates is used. Stack bar plot represents the sum-up of different types of colonies.

D. Bar plot of total colony numbers per 500 input cells from 3 CML patients (sample-150, 201, 171) under different treatment groups. (n=3, bars represent mean ± SD, not significant, unpaired t test)

E. Bar plot of CFU-M% among total colonies in 3 samples under different treatment groups. (n = 3, bars represent mean ± SD, not significant, unpaired t test)

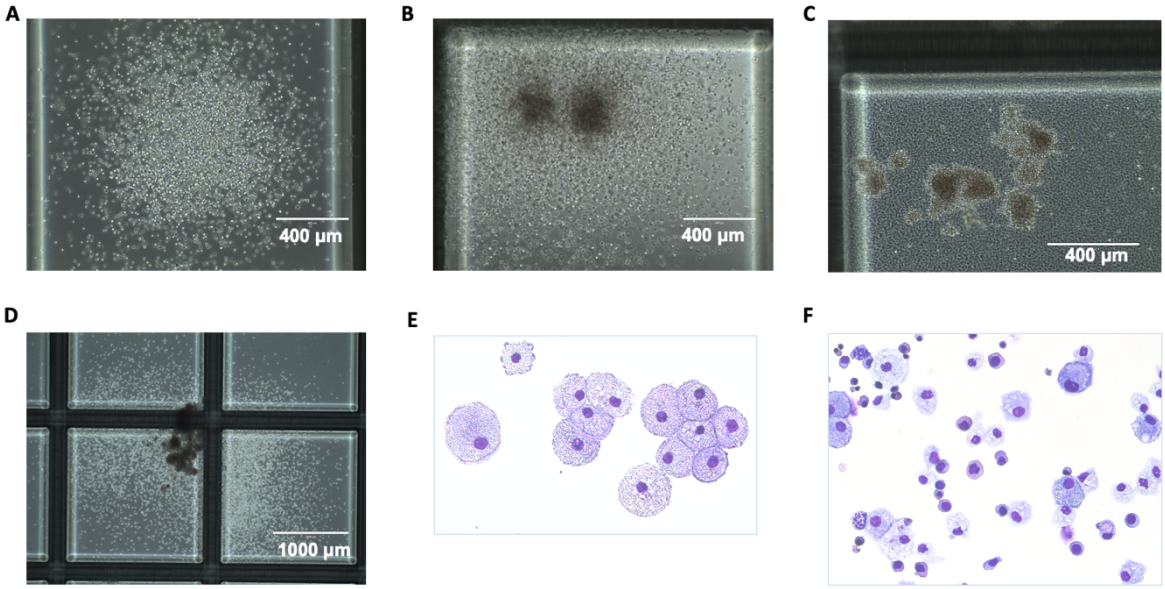


Figure 3.4 Morphology of colonies from CFU assay.

(A-D) Example colonies under microscope:

(A) CFU-M (from Sample-150 $Ctrl^{KD}$ IM group),

(B) 2 CFU-GM colonies (from Sample-150 $Ctrl^{KD}$ DMSO group),

(C) BFU-E (from Sample-150 *GAS2*^{KD} DMSO group),

(D) GEMM (from Sample-150 *GAS2*^{KD} IM group).

(E-F) Example Giemsa stain of cells under microscope, 400X magnification:

(E) Cells from CFU-M (from Sample-201 $Ctrl^{KD}$ DMSO group),

(F) Cells from CFU-GEMM (from Sample-150 *GAS2*^{KD} IM group).

CHAPTER 4.

Discussion

1. Quiescence of CML LSCs in TKI treatment resistance

CML has long been a classical disease model for targeted cancer therapy due to the characteristic *BCR-ABL1* fusion gene, which results in the CML cells being selectively vulnerable to inhibition. Like healthy HSCs, CML stem cells (LSCs) maintain self-renewal and quiescence properties. CML LSCs have a higher engraftment level and can induce CML in secondary transplantation in immunocompromised mice. While CML LSCs carry the BCR-ABL1 fusion protein and are targeted by traditional TKIs [23], TKI treatment promotes entry of LSCs into a quiescent state. Quiescent LSCs are less sensitive to TKI and can evade TKI-induced apoptosis and degradation.

Early in 2008, Ito *et al.* [53] found that As_2O_3 (Arsenic Trioxide) treatment significantly decreased promyelocytic leukaemia protein (PML) expression in leukaemia-initiating cells (LICs) and decreased the number of quiescent LICs by inducing cell-cycle entry in CML mice models. There are many clinical trials evaluating therapeutic effects of As_2O_3 in CML patients (*e.g.,* NCT00250042, NCT00053248, *et al.*). Another important finding from Essers *et al*. [54] revealed that IFNα treatment can promote dormant HSCs to move into proliferation status in mice models. It would be highly beneficial to identify a novel agent that targets stem cell quiescence and be used as adjuvant therapy with TKI to prevent disease resistance and relapse in CML.

2. *GAS2* validation in CML stem cells

Molecular signatures of CML LSCs from our single-cell transcriptomic dataset revealed *GAS2* as the most over-expressed gene in *BCR-ABL1*$^+$ CML stem cells

compared to either *BCR-ABL1*$^{-}$ CML stem cells or HSCs from healthy donors. Furthermore, the group of *BCR-ABL1*$^{+}$ CML stem cells with high *GAS2* expression was enriched for genes associated with cell quiescence, despite that the overall quiescence score in *BCR-ABL1*$^{+}$ cluster was the lowest compared to *BCR-ABL1*$^{-}$ CML stem cells or HSCs from healthy donors, which could be explained that *BCR-ABL1*$^{+}$ could promote proliferation. In turn, the *BCR-ABL1*$^{+}$ cluster had higher gene set score of S-phase. High *GAS2* expression in primitive CML stem cells were also noted in other bulk gene-array analysis. This led to us proposing that *GAS2* could act as a significant role in maintaining stem cell quiescence and drug resistance.

Owing to the scarcity of patient samples, cell lines were used for optimisation of the protocol of gene manipulation. After comparing *GAS2* mRNA expression level in our available haematopoietic cell lines, RPMI8226 had relative high expression, it was applied first. K562 is a classical cell line model for researchers to study CML, though it expresses the lowest level of *GAS2* among all cell lines. Since all characteristics of CML stem cells are quite distinct from cell lines, it's rational that the high expression of *GAS2* in CML stem cells can't reflect on CML cell lines.

The transduction efficiency of inducible lentivirus in primary cells could not be as high as in cell lines (7.06% *vs.* 52.1%). Many factors can affect transduction efficiency, such as cell division rate, cell metabolic activity and exchange rates between nucleus and cytoplasm. Primary cells are more complicated than cell lines in terms of cell metabolism, differentiation, and division, as a result, they are harder to be transfected. Another potential reason might be the large size of transduction vector. The length of the lentiviral vector carrying shRNA targeting *GAS2* is around 10kb, which made it hard to enter cell nucleus for transduction. We confirmed this

hypothesis by validating the relative high transduction efficiency (44%) of a smaller GFP-virus (5kb) in CML $CD34^+$ cells. In conclusion, this large size shRNA targeting *GAS2* cannot be applied efficiently for transduction in primary cells.

The CRISPR-Cas9 approaches with either single or dual RNPs are hard to apply on human HSCs. Pure knock-out cells can only be selected after single-cell expansion clones. Human HSCs will lose the stemness after long time single cell clone expanding culture. And if special media is applied to maintain the stemness of HSCs, this would add another confounding factor of the gene functional analysis outcome. The constitutively expressed lentiviral shRNA vector (pLKO_eYFP) could achieve high transduction efficiency and *GAS2* knock-down level in both K562 cell line and CML primary cells. Successfully transduced cells can be easily sorted by e-YFP signal. In summary, I chose this constitutively expressed lentivirus for gene manipulation and downstream functional analysis in human HSCs.

As expected, $GAS2^{KD}$ could increase the proliferation of CML stem cells. Quantitative analysis of absolute quiescent CML stem cells indicated that $GAS2^{KD}$ alone could reduce the quiescent stem cells, and $GAS2^{KD}$ plus IM treatment could obtain the lowest quiescent CML stem cells, though not statistically significant. This can be attributed to many factors: such as low sample numbers, batch effect, and subjective counting deviation, *etc*. More samples need to be tested and more concise quantification tools such as counting beads can be applied to further validate this conclusion.

Previous literature reviews indicated that *GAS2* acts as a dual regulator in apoptosis in other cell lines. I found that $GAS2^{KO}$ K562 single cell clones had similar live cells and apoptotic cells percentage compared to control. No significant

changes in apoptosis were also noted in *GAS2*^{KD} CML stem cells *vs.* scramble control^{KD} CML stem cells. But this conclusion needs to be validated further due to low sample numbers in my experiment, and apoptosis stimulation could be added in the experimental design to specifically validate apoptosis effect.

In CFC assays, *GAS2*^{KD} groups had more total colonies than control groups, either alone or combining IM treatment, though not statistically significant, this result correlated with the proliferative effect of *GAS2*^{KD} observed using CFSE stain in CML stem cells. *GAS2*^{KD} groups appealed to have lower CFU-M percentages and more other colony types than control groups (also not statistically significant). Though no direct research reported *GAS2* in relation to haematopoiesis, Guglielmelli *et al*. [55] reported *GAS2* was among a set of 8 gene clusters including *WT1*, whose mutated form in human CD34^{+} cord blood cells could block myeloid differentiation *in vitro* [56]. This could be possible supporting evidence that *GAS2*^{KD} in CML peripheral blood stem cells could decrease myeloid output, but more replicates of the experiments are needed for the validity. CFU-M acted as the main proportion of colonies grown from all treatment groups. This could attribute to the methylcellulose-based medium, MethoCult™ H4435, which is myeloid biased. Cells under such *ex vivo* condition can differentiate more myeloid colonies than *in vivo*, let alone all samples are CML, which can pathologically differentiate more into myeloid lineages. Overall, these factors would affect the veracity of the lineage differentiation potential outcome.

In my project, I successfully set up a stable and efficient gene knock-down strategy for *GAS2* in human HSCs using a constitutively expressed lentiviral shRNA vector, *GAS2*^{KD} cells can be sorted by e-YFP marker with stable gene knock-down. Quiescent and dividing stem cells can be distinguished by cell surface markers and

cell trace staining. This optimised efficient gene knock-down and viable cell-tracing system can be ready-to-use in future for any candidate target validation in human HSCs.

There are some limitations as well. First, patient samples included in the experiment are limited, which significantly affect the power in statistical analysis. More samples are needed to get convincing conclusions. Second, there is still difference between *ex vivo* and *in vivo* condition. To confirm if *GAS2* can act as a candidate target in CML patients, mice models are needed to test the $GAS2^{KD}$ effect *in vivo*, the efficacy of combining treatment of $GAS2^{KD}$ with TKIs, as well as toxicity to healthy HSCs and other organs. Technically, the virus transduction I used cannot achieve on/off control of *GAS2* expression due to the vector design. It would be hard for me to restore *GAS2* expression in cells for functional analysis. Over-expression of *GAS2* also needs to be validated in CML stem cells. And I only validated part of biological functions of *GAS2* in CML stem cells, other impact of *GAS2* can be further explored by RNA-sequencing cells from $Ctrl^{KD}$ DMSO, $Ctrl^{KD}$ IM, $GAS2^{KD}$ DMSO, and $GAS2^{KD}$ IM treatment groups.

3. Possible future directions

Future plans for my project are as follows: First, I need to validate the GAS2 protein reduction level in $eYFP^+$ CML $CD34^+$ cells, because I only validated that *GAS2* mRNA level was decreased by 86% in $eYFP^+$ CML $CD34^+$ cells compared to control knock down. To test GAS2 protein level, the gold standard is the western blot, which requires large number of cells to extract enough protein for the antibody conjugation. This is hard to achieve in the precious and rare patient primary cells' samples. Therefore, flow cytometry test of GAS2 protein signal can be applied. A

specific GAS2 antibody conjugated to fluorescence protein is required, and the protocol for intracellular protein detection by permeabilization and fixation needs to be optimised to get the solid conclusion.

Second, only mild increase of proliferation index and a trend of absolute quiescent cell number reduction was noted from my function analysis in *GAS2*^{KD} CML stem cells, this is probably due to limited replicates in this experiment. More patient samples are needed to decrease the beta-error. Furthermore, healthy donors' primary cell samples are needed to test if *GAS2*^{KD} could achieve the same effects as in CML patients or could cause lethality in healthy samples.

Third, I already sorted CML stem cells from 4 treatment groups into lysis buffer on day 7, they will be further processed for bulk RNA-sequencing. By performing gene sets and pathway enrichment analysis, I can check if this correlates with my functional assays' outcome, and I can explore other promising biological changes in *GAS2*^{KD} alone or with IM treatment in CML stem cells.

Last and the most important, if the RNA-sequencing results correlate quite well with my results that *GAS2*^{KD} can promote stem cells proliferation and into dividing phase, and *GAS2*^{KD} in healthy donors' samples do not cause severe lethality. *In vivo* experiments can be implemented using CML mice models to observe the effects of *GAS2*^{KD} alone or with TKI treatment. The toxicity and safety, the effectiveness on long term survival can all be evaluated.

4. Conclusions

This study has provided preliminary validation for a potential novel target therapy to help eliminate quiescent human CML stem cells. By optimising different gene

manipulation protocols, I have generated a stable and efficient gene knock-down strategy in human CML primary cells which can be used for future studies into any target validation in all sorts of human haematopoietic malignant diseases.

Using this gene knock-down strategy, I successfully generated *GAS2*^{KD} human CML stem cells which proliferate faster compared to control^{KD} cells, and the absolute cell number of quiescent CML stem cells is reduced in *GAS2*^{KD} in combination with Imatinib. Apoptosis and lineage differentiation are not much affected by *GAS2*^{KD} in CML stem cells under *ex vivo* culture. Although larger sample sizes are needed to increase the reliability of these functional analysis, this provided us insights into potential therapy by promoting quiescent CML stem cells re-entry into cell cycle and proliferate to eliminate TKI insensitive stem cells.

Appendices

Appendix 1. Key Resources Table

Reagent	Source	Identifier
SMARTvector Inducible Lentiviral shRNA		
Nontargeting control	Horizon Discovery Ltd	VSC11657 CACACAACATGTAAACCAGGGA
GAS2 (ORF) (B10)	Horizon Discovery Ltd	V3SH11252-230547613 GCCGCTCCACACAGAACTT
GAS2 (ORF) (A10)	Horizon Discovery Ltd	V3SH11252-228290050 TAAATCTCGGCACCAGGAT
GAS2 (3'UTR) (A9)	Horizon Discovery Ltd	V3SH11252-224759413 CGTATGTCAGCTCATGTAT
Constitutive Lentiviral shRNA		
Scramble control	Genome Engineering Facility of Weatherall Institute of Molecular Medicine	GATGTTTAGTACGGGCAGTAT
shRNA1-GAS2		AAGGAGATTACAGCAGAAACT
shRNA2-GAS2		AAGTTGGACAATGGTGCCTTG
shRNA3-GAS2		GGTGGATGAAACGTGTCTATT
Synthetic sgRNAs		
AAVS1	Merck Ltd	GGGGCCACTAGGGACAGGAT
GAS2 (within exon3)	Merck Ltd	TTTATGGAGAAGTTGGACAA
GAS2 (5'-1 upstream exon3)	Merck Ltd	TCTAGCTGAGTTACATAACT
GAS2 (5'-2 upstream exon3)	Merck Ltd	GTTCCATCATTTAGTAATCA
GAS2 (3'-1 downstream exon3)	Merck Ltd	ATGCAAAAATCACTCTTAGG

GAS2 (3'-2 downstream exon3)	Merck Ltd	TATTGTGAACGGATGCATCA
GAS2 (1-within exon4)	Merck Ltd	ATGCAAAACCAGTGCACCCT
GAS2 (2-within exon4)	Merck Ltd	ACGTGTCTATTTGAATCGGA
DNA primers (5' to 3')		
GAS2 exon3_fw1	Invitrogen	ACTGGGGATGAGAATGGAATGA
GAS2 exon3_rv1	Invitrogen	GCTAGATCTGTGCATTGCCTAA
AAVS1_fw	Invitrogen	CTTGCTTTCTTTGCCTGGAC
AAVS1_rv	Invitrogen	GATCAGTGAAACGCACCAGA
GAS2 exon3_fw2	Invitrogen	TGGGCCACTGGGGATGAGAATGG
GAS2 exon3_rv2	Invitrogen	CTTGGCTAGATCTGTGCATTGCC
GAS2 exon3_fw3	Invitrogen	ACGGCGTTCTTGTAGTCAGA
GAS2 exon3_rv3	Invitrogen	GGCTAGATCTGTGCATTGCC
GAS2 exon4_fw1	Invitrogen	GCTTAGATGAAAGGGAAGGCC
GAS2 exon4_rv1	Invitrogen	AGAACATGCACAGCTACCCT
GAS2 exon4_fw2	Invitrogen	GCCTGAGCCAAGAAAAGTACT
GAS2 exon4_rv2	Invitrogen	ATGGCTAGAACATGCACAGC
TaqMan® Gene Expression Assays		
GAS2	Thermo Fisher Scientific	Cat#4331182 Hs00169477_m1
GAPDH	Thermo Fisher Scientific	Cat#4331182 Hs02786624_g1
Antibodies		
Lineage Cocktail-BV510	BioLegend	Cat#348807
CD34-APC efluor780	Thermo Fisher Scientific	Cat#47-0349-42 RRID AB_2573956
CD38-BV650	BioLegend	Cat#356620

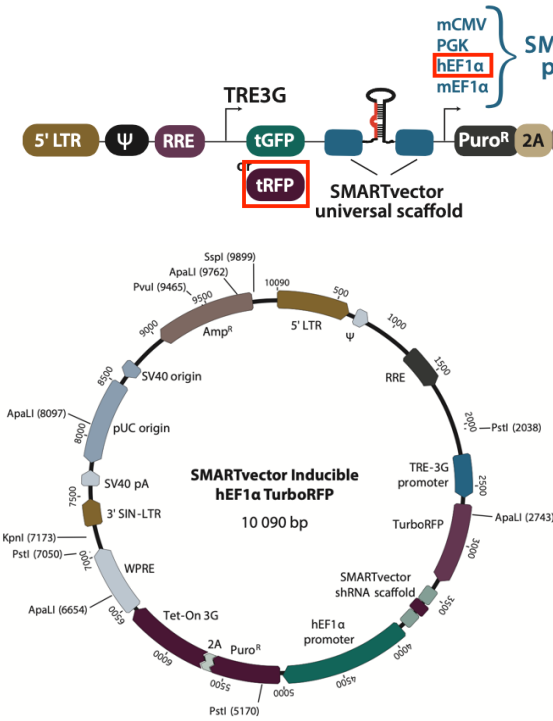
Annexin-V APC conjugate	Thermo Fisher Scientific	Cat#A35110
CellTrace CFSE	Thermo Fisher Scientific	Cat#C34554
CellTrace Violet	Thermo Fisher Scientific	Cat#C34557
Ki-67-AF700	Thermo Fisher Scientific	Cat#56-5698-82
Gas2 (F-12) (mice monoclonal)	Santa Cruz Biotech	Cat#sc-398669
Gas2 (Rabbit monoclonal)	Abcam	Cat#ab109762
Tubulin (mice monoclonal)	Abcam	Cat#ab7291
Anti-mouse IgG, HRP-linked Antibody	Cell Signaling Tech	Cat#7076
Anti-rabbit IgG, HRP-linked Antibody	Cell Signaling Tech	Cat#7074

Biological samples

Study	Patient ID	Age	Mutation	Treatment	Source
INForMeD Study (IRAS 199833; REC 16/LO/1376 University of Oxford)	001/150	69	BCR-ABL	Baseline	Peripheral blood
	001/201	32	BCR-ABL	Baseline	Peripheral blood
	001/171	87	BCR-ABL	Baseline	Peripheral blood
	001/198	55	BCR-ABL	Baseline	Peripheral blood
	001/162	48	BCR-ABL	Baseline	Bone marrow

Appendix 2. Supplementary figures

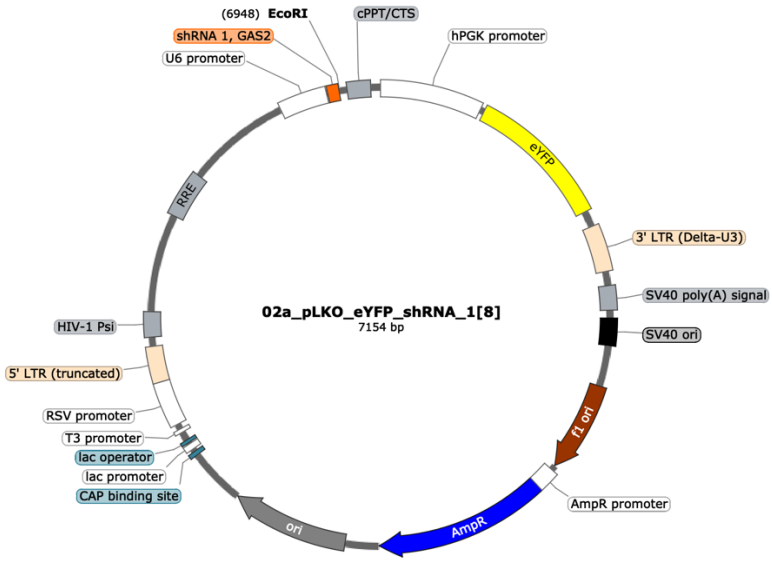
Supplementary figure 1. Elements and plasmid map of the Dharmacon SMARTvector inducible lentiviral shRNA (Horizon Discovery Ltd).



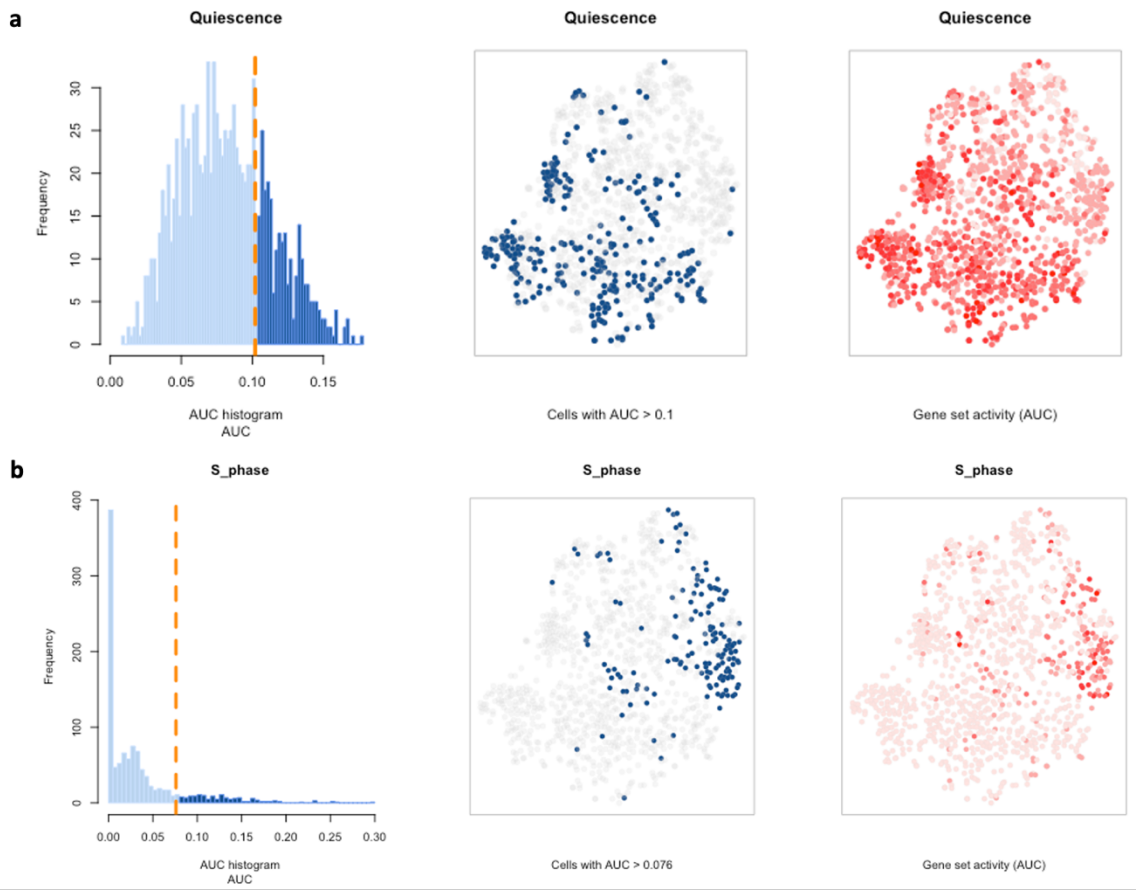
Plasmid features	Position
5' LTR	1–635
Ψ	685–822
RRE	1252–1605
TRE-3G promoter	2240–2615
TurboRFP	2668–3363
SMARTvector shRNA scaffold	3458–3864
hEF1α promoter	3885–5002
Puro^R	5021–5623
2A	5624–5683
Tet-On™ 3G	5684–6430
WPRE	6444–7035
3' SIN-LTR	7239–7473
SV40 pA	7582–7712
pUC origin	7783–8508
SV40 origin	8578–8713
Amp^R	9021–9881

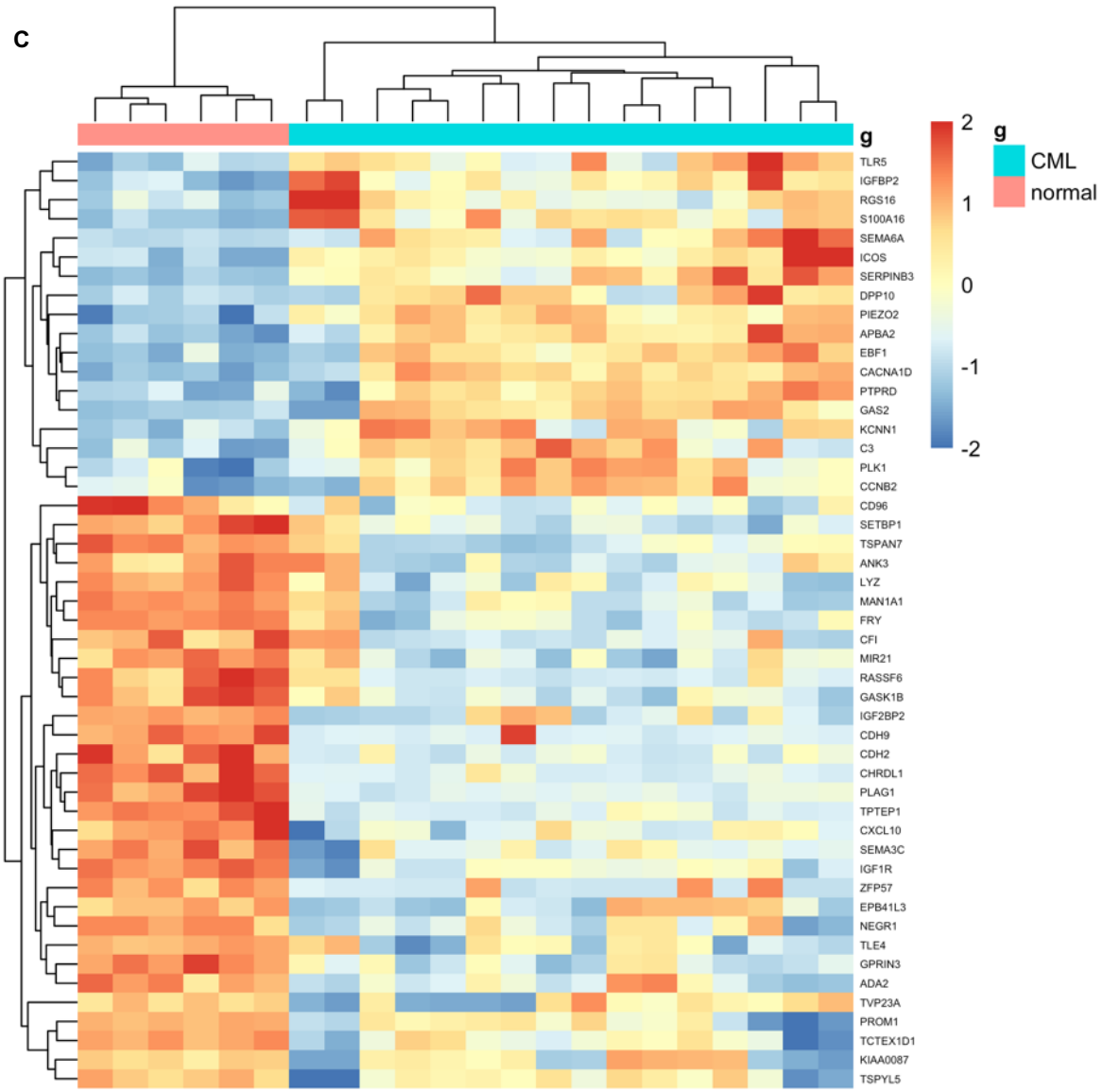
Restriction sites	Position
ApaLI	2743, 6654, 8097, 9762
KpnI	7173
PstI	2038, 5170, 7050
PvuI	9465
SspI	9899

Supplementary figure 2. Plasmid map of Constitutive lentiviral shRNA (Genome Engineering Facility of Weatherall Institute of Molecular Medicine)

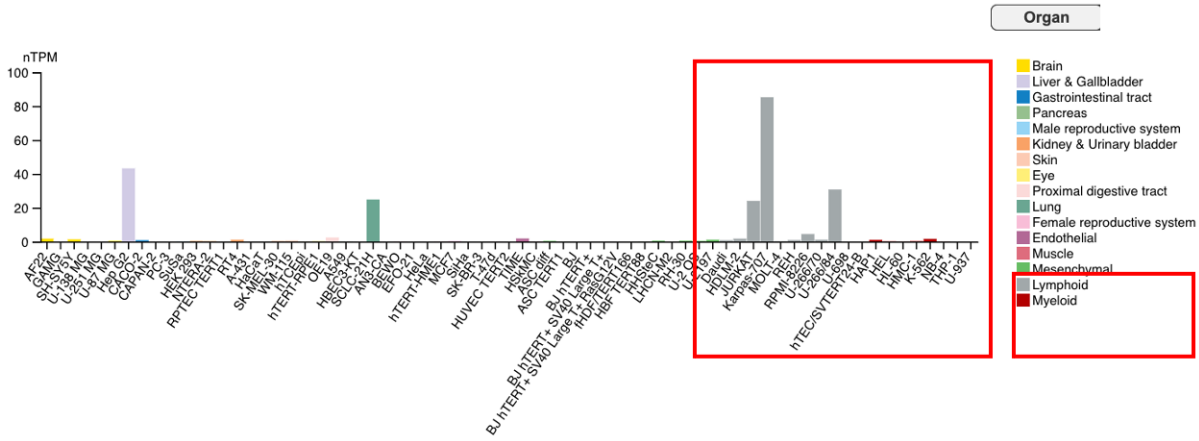


Supplementary figure 3. (3a)(3b).The left figure is AUC histogram showing the proportion of genes in the gene set that are highly expressed in each cell, the dotted line represented the automatically assigned threshold for gene set “active” by “AUCell” package; The middle figure shows the cluster of cells with AUC score passing the assignment threshold on TSNE plot. The right figure imposed AUC score of each cell by gradient red colour on TSNE plot. (3c). Heat map illustrating the hierarchical clustering of CML HSCs (n=16) or normal HSCs (n=6) (samples from GSE47927) showing the top 49 differentially expressed genes (absolute value of logFC ≥ 2).

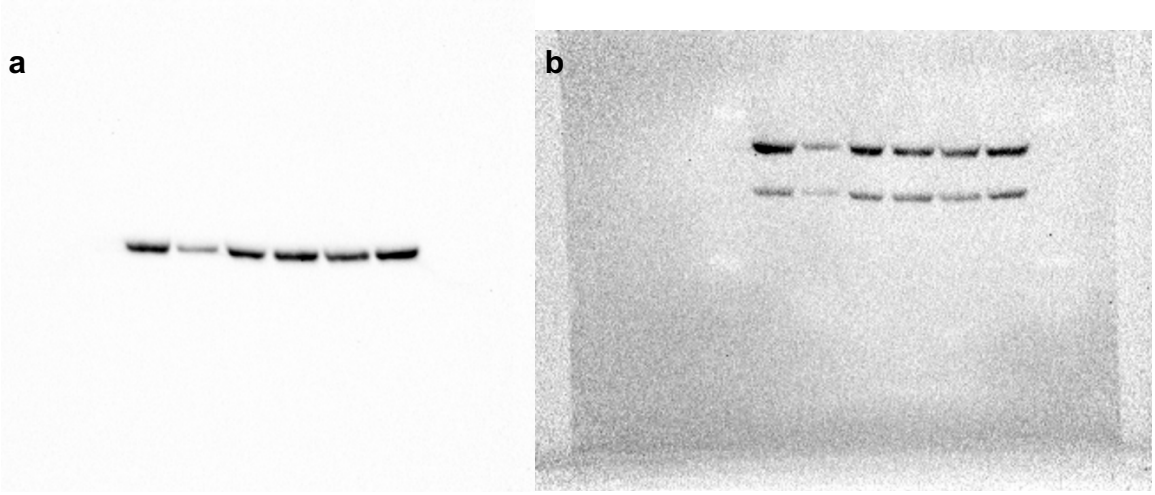




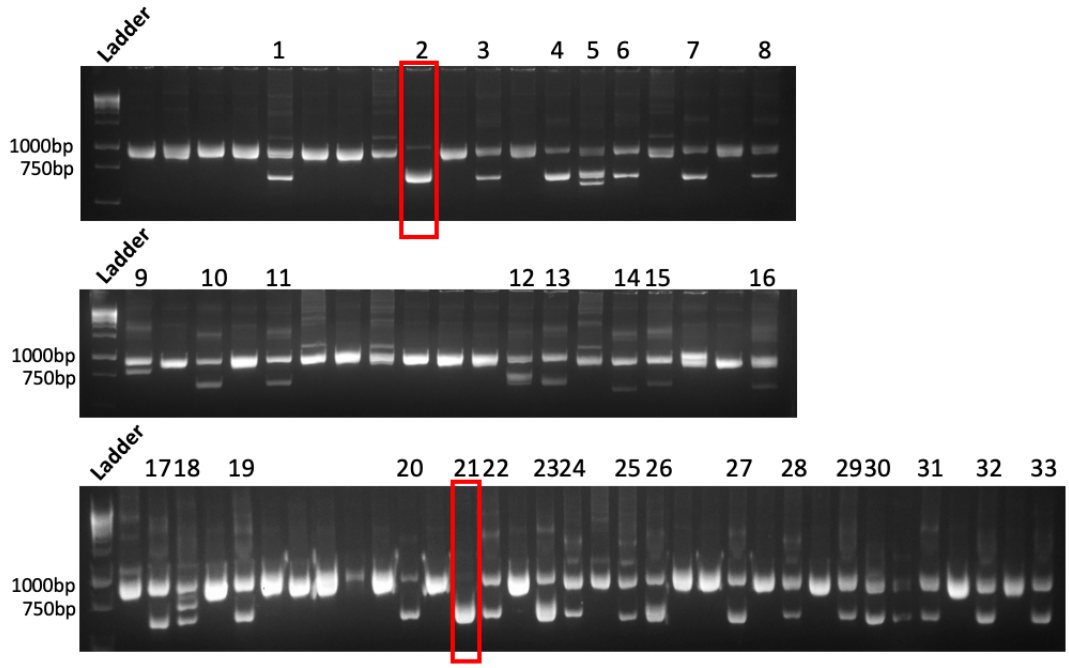
Supplementary figure 4. *GAS2* RNA expression levels in different cell lines analysed in the Human Protein Atlas. Normalized transcript per million (nTPM) values are used for all cell lines expression, nTPM value of 1.0 is defined as the threshold for the expression of the corresponding protein. Different colours represent different organs where cell lines are categorized. The lymphoid (grey) and myeloid (red) cell lines are highlighted in the red square, ranging from Daudi to U-937 cell lines. (Figure cited from https://www.proteinatlas.org/ENSG00000148935-GAS2/subcellular#rna)



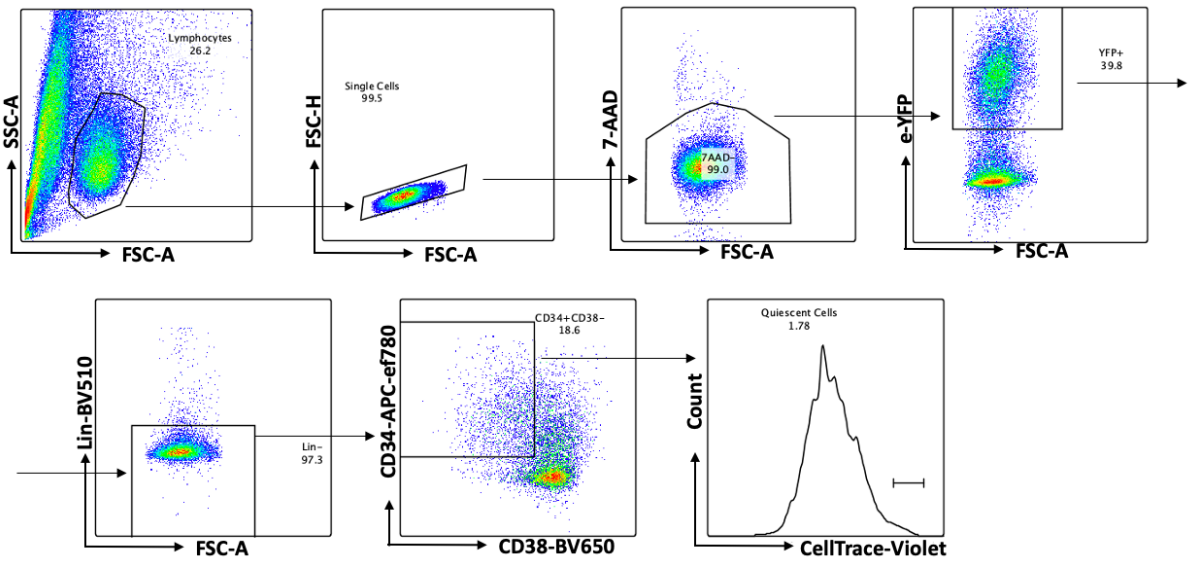
Supplementary figure 5. Uncropped western blot membrane shown in Figure 2.5C. (5a) is the western blot of GAS2 protein, (5b) is the western blot of TUBULIN. Samples from the left to right are cell lysates from: K562, RPMI8226, Jurkat, Ctrl^{KO} of K562, *GAS2*-2^{KO} of K562, *GAS2*-2^{KO} of K562.



Supplementary figure 6. Genotyping of 71 single cell clones from K562 cells after *GAS2* knock-out by dual RNPs nucleofection. 33 clones are heterozygous, clone 2 and clone 21(red rectangles) had the clearest shorter band.



Supplementary figure 7. Gating strategy of CML e-YFP^{+} Lin^{-}CD34^{+}CD38^{-} cells on day 7. The example figure is from the *GAS2*^{KD} DMSO group of sample-201.



Appendix 3 Supplementary table

Supplementary table 1. Reverse-transcription quantitative PCR of quiescent and dividing cells in different treatment groups by relative C_T value comparison. The C_T value of house-keeping gene *GAPDH* was used as internal control. The grey values are omitted due to high deviation from the other 2 replicates.

		GAPDH(C_T)	GAS2(C_T)		
Ctrl DMSO	quiescent	7.085	15.937	15.851	15.925
Ctrl DMSO	dividing	7.529	18.036	18.413	17.087
Ctrl DA	quiescent	6.914	15.293	15.320	15.540
Ctrl DA	dividing	6.405	16.497	16.616	16.684
B10 DMSO	quiescent	6.335	16.140	16.195	16.058
B10 DMSO	dividing	6.928	18.701	18.975	18.000
B10 DA	quiescent	6.830	NA	NA	NA
B10 DA	dividing	6.804	NA	NA	NA

Appendix 4 Bioinformatic analysis scripts

Script 1: R code to generate Figure 1.1C – 1.1G

```
library(limma)

##################################################################

my.anno<-read.delim("../Data/CML_PROJECT_ALL_CELLs.Freezed-
5.anno.txt",header=T)

my.anno<-subset(my.anno,used4analysis==1)

my.anno<-subset(my.anno,Stage_2=="normal_hsc"| Stage_2=="diagnosis")

##################################################################

load(file="../Data/CML_PROJECT_ALL_CELLs.rdata")

my.cells<-CML_PROJECT_ALL_CELLs[,as.character(my.anno$Cell)]

##################################################################

min_expr<-1

S.m<-as.matrix(my.cells)

S.m[S.m < min_expr]<- 0

S.m[S.m >= min_expr]<-log2(S.m[S.m >= min_expr])

#################With Batch effect#################################

my_bcr_abl<-as.character(my.anno$BCR_ABL)

my_bcr_abl[my_bcr_abl=="negative_low_gapdh"]<-"negative"

my_bcr_abl<-as.factor(my_bcr_abl)

my.design<-model.matrix(~my_bcr_abl)

rownames(my.design)<-colnames(S.m)

batch.x<-as.character(my.anno$process_date)

batch.x[which(batch.x=="batch1" | batch.x=="batch2")]<-"A"

batch.x[which(batch.x=="batch5")]<-"B"

batch.x[which(batch.x=="batch6")]<-"C"
```

```
S.m.no.batch = removeBatchEffect(S.m,batch=batch.x,design=my.design)

#################################################################

load(file="Diagnosis.Pos.HSC.highVariable.Genes.rdata")

load(file="Diagnosis.Neg.HSC.highVariable.Genes.rdata")

load(file="Normal_HSCs.highVariable.Genes.rdata")

#################################################################

my.genes<-
c(Normal_HSCs.highVariable.Genes,Diagnosis.Pos.HSC.highVariable.Genes,Dia
gnosis.Neg.HSC.highVariable.Genes)

my.genes<-unique(my.genes)

#################################################################

S.m2<-subset(S.m.no.batch,rownames(S.m.no.batch) %in% my.genes)


# tsne:

install.packages("Rtsne")

library(Rtsne)

tsne_data <- Rtsne( t(S.m2),2,20,20 )


# cluster:

clust <- as.character( my.anno$BCR_ABL)

clust[clust=="negative_low_gapdh"] <- "negative"

clust[clust=="negative"] <- "BCR-ABL-"

clust[clust=="positive"] <- "BCR-ABL+"

clust[clust=="normal"] <- "Normal HSC"


#tsne figure draw

input_original <- data.frame( cbind(tsne_data$Y, clust ) )

colnames(input_original) <- c("TSNE1","TSNE2","cluster")
```

```
input_original$TSNE1 <- as.double( as.character(input_original$TSNE1) )
input_original$TSNE2 <- as.double( as.character(input_original$TSNE2) )
library("ggplot2")

png("plot_all.png", units="in", width=5, height=5, res=300)
ggplot(input_original, aes(x=TSNE1,y=TSNE2,col=cluster) ) + geom_point()
dev.off()

#use AUCell method to plot tsne
# to plot other gene sets AUCell score, replace the quiescence gene set accordingly
library(GSEABase)
library("AUCell")
genes <- quiescence[[1]]
geneSets <- GeneSet(genes, setName="quiescence")
geneSets

cells_rankings <- AUCell_buildRankings(S.m2, nCores=1, plotStats=TRUE)
cells_rankings
save(cells_rankings, file="cells_rankings.RData")

# Calculate enrichment for the gene signatures (AUC)
cells_AUC <- AUCell_calcAUC(geneSets, cells_rankings)
save(cells_AUC, file="cells_AUC.RData")

# impose quiescence score onto tsne plot
```

```
cells_AUC
auc <- getAUC(cells_AUC)
auc <- t(auc)

input <- data.frame(input_original,auc)
colnames(input) <- c("TSNE1","TSNE2","cluster","Quiescence")

input$TSNE1 <- as.double( input$TSNE1 )
input$TSNE2 <- as.double( input$TSNE2 )

input$Quiescence <- as.numeric( as.character(input$Quiescence ) )

png("plot_quiescence.png", units="in", width=6, height=5, res=300)

ggplot(input, aes(x=TSNE1,y=TSNE2,col=Quiescence) ) +
  geom_point() +
  scale_color_gradient(low="blue", high="red")

dev.off()

# ## find row number of GAS2 in the matrix
which(rownames(S.m2) == "GAS2")

# gene level: gas2
library(data.table)
input <- data.table( cbind(cellsTsne, S.m2[2638,], clust) )
gas2 <- data.table(cbind(S.m2[2638,], clust))
```

```
class(gas2)
write.csv(gas2, "GAS2_expression.csv", row.names = FALSE)

#plot violin plot of selected gene
gas2 <- read_csv("GAS2_expression.csv", col_names = TRUE)
colnames(gas2) <- c("V1", "cluster")
gas2 %>% group_by(cluster) %>%
  select(V1) %>%
  summarise_all(list(mean = mean,
                     std = sd,
                     min = min,
                     max = max))

png("gas2_violin.png", units="in", width=5, height=5, res=300)

ggplot(gas2, aes(cluster, V1, color = cluster) ) +
  geom_violin(fill = "grey80", alpha=.5) +
  geom_point(position = "jitter", alpha=1, cex = 0.5) +
  labs(x = "type", y = "GAS2 expression" ) +
  stat_compare_means() +
  theme_classic()

dev.off()

# gene level: gas2
library(data.table)
input <- data.table( cbind(tsne_data$Y, S.m2[2638,], clust) )
```

```
colnames(input) <- c("TSNE1","TSNE2","GAS2","cluster")
input$TSNE1 <- as.double( input$TSNE1 )
input$TSNE2 <- as.double( input$TSNE2 )
input$GAS2 <- as.numeric( as.character(input$GAS2 ) )

png("plot_gas2.png", units="in", width=5, height=5, res=300)
ggplot(input, aes(x=TSNE1,y=TSNE2,col=GAS2) ) +
  geom_point() +
  scale_colour_gradient(low="blue", high="red")
dev.off()
```

Script 2: R code for GSE47927 analysis

```
options()$repos
options()$BioC_mirror
options(BioC_mirror="https://mirrors.ustc.edu.cn/bioc/")
options("repos" = c(CRAN="https://mirrors.tuna.tsinghua.edu.cn/CRAN/"))
options()$repos
options()$BioC_mirror

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("KEGG.db",ask = F,update = F)
BiocManager::install(c("GSEABase","GSVA","clusterProfiler" ),ask = F,update =
F)
BiocManager::install(c("GEOquery","limma","impute" ),ask = F,update = F)
BiocManager::install(c("genefu","org.Hs.eg.db","hgu133plus2.db"  ),ask  =
F,update = F)
```

```
options()$repos
install.packages('WGCNA')
install.packages(c("FactoMineR", "factoextra"))
install.packages(c("ggplot2", "pheatmap","ggpubr"))
library("FactoMineR")
library("factoextra")

library(GSEABase)
library(GSVA)
library(clusterProfiler)
library(genefu)
library(ggplot2)
library(ggpubr)
library(hgu133plus2.db)
library(limma)
library(org.Hs.eg.db)
library(pheatmap)

options(stringsAsFactors = F)
f='GSE47927_eSet.Rdata'
library(GEOquery)
#Setting options('download.file.method.GEOquery'='auto')
#Setting options('GEOquery.inmemory.gpl'=FALSE)
if(!file.exists(f)){
  gset <- getGEO('GSE47927', destdir=".",
                 AnnotGPL = F,
```

```
                getGPL = F)
  save(gset,file=f)
}
load('GSE47927_eSet.Rdata')
class(gset)
length(gset)
class(gset[[1]])
gset

a=gset[[1]]
dat=exprs(a)
dim(dat)
dat[1:4,1:4]
class(dat)

pd=pData(a)

qx <- as.numeric(quantile(dat, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
  (qx[6]-qx[1] > 50 && qx[2] > 0) ||
  (qx[2] > 0 && qx[2] < 1 && qx[4] > 1 && qx[4] < 2)

if (LogC) { dat[which(dat <= 0)] <- NaN
dat <- log2(dat)
print("log2 transform finished")}else{print("log2 transform not needed")}

library(hugene10sttranscriptcluster.db)
```

```
ids=toTable(hugene10sttranscriptclusterSYMBOL)
head(ids)

head(ids)
colnames(ids)=c('probe_id','symbol')
ids=ids[ids$symbol != '',]
ids=ids[ids$probe_id %in%  rownames(dat),]

dat[1:4,1:4]
dat=dat[ids$probe_id,]

ids$median=apply(dat,1,median)
ids=ids[order(ids$symbol,ids$median,decreasing = T),]
ids=ids[!duplicated(ids$symbol),]
dat=dat[ids$probe_id,]
rownames(dat)=ids$symbol
dat[1:4,1:4]

library(stringr)
library("tidyr")
group_list=str_split(pd$title,'\\.',simplify = T)[,1]
table(group_list)

save(dat,group_list,file = 'step1-output.Rdata')

## find row number of GAS2 in the matrix
which(rownames(dat) == "GAS2")
```

```
gas2express <- dat[13003,]
class(gas2express)
class(group_list)
df <- data.frame(group_list, gas2express)
write.csv(df,file = 'gas2_expression.csv')

# filter HSC output
rm(list = ls())
options(stringsAsFactors = F)
load(file = 'step1-output.Rdata')
library(dplyr)
dat_tag <- as.data.frame(t(dat))
dat_tag <- cbind( dat_tag, group_list)
dat_tag_1 <- dat_tag %>%
  dplyr:: filter (group_list %in% c("HSC_normal", "MPP_normal", "HSC_CP-CML", "HSC_AP-CML", "MPP_AP-CML", "HSC_BC-CML", "MPP_BC-CML"))
dat_tag_1$type = str_match(dat_tag_1$group_list,"normal|CML" )
View(dat_tag_1$type)

## transform df to matrix
dat_tag_1 <- dat_tag_1[,-which(names(dat_tag_1)%in%c("group_list"))]
type <- as.character(dat_tag_1$type)
dat_tag_1 <- dat_tag_1[,-which(names(dat_tag_1)%in%c("type"))]

dat_hsc =as.matrix(dat_tag_1)
dat_hsc <- t(dat_hsc)
class(dat_hsc)
```

```
dat_hsc[1:4, 1:4]
class(dat)

bp=function(g){
  library(ggpubr)
  df=data.frame(gene=g,stage=type)
  p <- ggboxplot(df, x = "stage", y = "gene",
                 colour = "stage", palette = "jco",
                 add = "jitter")
  #  Add p-value
  p + stat_compare_means()
}

png("boxplot_gas2.png", units="in", width=5, height=5, res=300)
bp(dat_hsc[13003,])
dev.off()

library(limma)
design=model.matrix(~factor(type))
fit=lmFit(dat_hsc,design)
fit=eBayes(fit)
options(digits = 4)
topTable(fit,coef=2,adjust='BH')

design <- model.matrix(~0+factor(type))
colnames(design)=levels(factor(type))
head(design)
```

```
exprSet=dat_hsc
rownames(design)=colnames(exprSet)

contrast.matrix<-makeContrasts("CML-normal",
                               levels = design)
contrast.matrix

deg = function(exprSet,design,contrast.matrix){
  ##step1
  fit <- lmFit(exprSet,design)
  ##step2
  fit2 <- contrasts.fit(fit, contrast.matrix)



  fit2 <- eBayes(fit2)  ## default no trend !!!
  ##eBayes() with trend=TRUE
  ##step3
  tempOutput = topTable(fit2, coef=1, n=Inf)
  nrDEG = na.omit(tempOutput)
  #write.csv(nrDEG2,"limma_notrend.results.csv",quote = F)
  head(nrDEG)
  return(nrDEG)
}

deg = deg(exprSet,design,contrast.matrix)
head(deg)
save(deg,file = 'deg.Rdata')
```

```
## for volcano plot
if(T){
  nrDEG=deg
  head(nrDEG)
  attach(nrDEG)
  plot(logFC,-log10(P.Value))
  library(ggpubr)
  df=nrDEG
  df$v= -log10(P.Value)
  ggscatter(df, x = "logFC", y = "v",size=0.5)

  df$g=ifelse(df$P.Value>0.01,'stable',
              ifelse( df$logFC >1,'up',
                      ifelse( df$logFC < -1,'down','stable') )
  )
  table(df$g)
  df$name=rownames(df)
  head(df)
  ggscatter(df, x = "logFC", y = "v",size=0.6,colour = 'g')
  ggscatter(df, x = "logFC", y = "v", colour = "g",size = 0.6,
            label = "name", repel = T,
            xlab = "Log2(Fold Change)", ylab = "-log10(adj.P.Val)",
            font.x = c(14, "bold", "black"), font.y = c(14, "bold", "black"),
            #label.select = rownames(df)[df$g != 'stable'] ,
            label.select = "GAS2",
            palette = c("green", "gray", "red") )
```

```
  ggsave('volcano.png')


}

## for heatmap
if(T){
  x=deg$logFC
  names(x)=rownames(deg)
  cg=c(names(head(sort(x),31)),
       names(tail(sort(x),18)))
  library(pheatmap)
  pheatmap(dat_hsc[cg,],show_colnames = F,show_rownames = T)
  n=t(scale(t(dat_hsc[cg,])))

  n[n> 2]= 2
  n[n< -2]= -2
  n[1:4,1:4]
  ##n <- na.omit(n)
  pheatmap(n,show_colnames = F,show_rownames = T)
  ac=data.frame(g=type)
  rownames(ac)=colnames(n)
  pheatmap(n,show_colnames = F,
           show_rownames = T,
           cluster_cols = T,
           border = FALSE,
           treeheight_row = 30,
           fontsize_row = 5,
```

```
annotation_col=ac,filename = 'heatmap_abslogfc2_DEG.png')

}
```

Genes in quiescence gene set and cell cycle S phase gene set

Quiescence gene set [33]**:** *ADAM28 ADAM8 AIF1 AKR1C3 ALDH1A1 ALOX5ANKRD12 ANXA5 APP AREG ARMCX2 BAALC BCL11A BIRC3 C3orf14 CCL19 CCL20 CCL8 CCNG2 CD200 CD40 CD44 CD52 CD59 CH25H CLEC2B CREM CRHBP CTSO CXCL1 CXCL10 CXCL11 CXCL13 CXCL2 CXCL3 CXCL5 CXCL6 CYTIP DDX17 DMD DST DUSP6 EBI3 EMP1 ENPP4 ERG EVI2A FAM129A FHL1 FLJ10038 FNBP1 FTH1 FXR1 GBP2 GLIPR1 GPR126 GUCY1A3 H1F0 HCP5 HIST1H2AC HIST1H2BK HIST2H2BE HLA-DMA HLA-DMB HLA-DPA1 HLA-DPB1 HLA-DQA1 HLA-DQB1 HLA-DRB4 HLF HLX IDS IFI16 IL18R1 IL1B IL8 INHBA JUN KIAA012 KYNU LIMCH1 MAFF MIR22HG MLLT3 MPZL2 MYO5C NEAT1 NFAT5 NR4A2 NRIP1 OPTN PCDH9 PMCH PPFIBP1 PROM1 PTGER4 PTGS2 PTPRC RABGAP1 RBPMS RGS1 RPL31 SCHIP1 SELL SH3BP5 SOD2 SORL1 SPINK2 SPTBN1 STAT4 SVIL TCEAL2 TCF4 TCF7L2 TFPI TGFB1I1 TM4SF1 TNFSF10 TNFSF4 TPBG TRA2A TRA2B TRIM22 TSPAN31 TSPAN6 VNN1 ZBTB10*

Cell cycle S phase gene set [33, 57]: *MCM5 PCNA TYMS FEN1 MCM2 MCM4 RRM1 UNG GINS2 MCM6 CDCA7 DTL PRIM1 UHRF1 MLF1IP HELLS RFC2 RPA2 NASP RAD51AP1 GMNN WDR76 SLBP CCNE2*

UBR7 POLD3 MSH2 ATAD2 RAD51 RRM2 CDC45 CDC6 EXO1

TIPIN DSCC1 BLM CASP8AP2 USP1 CLSPN POLA1

CHAF1B BRIP1 E2F8

References:

[1] GIUSTACCHINI A, THONGJUEA S, BARKAS N, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia [J]. Nat Med, 2017, 23(6): 692-702.

[2] FORD C E, HAMERTON J L, BARNES D W, et al. Cytological identification of radiation-chimaeras [J]. Nature, 1956, 177(4506): 452-4.

[3] BENVENISTE P, FRELIN C, JANMOHAMED S, et al. Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential [J]. Cell Stem Cell, 2010, 6(1): 48-58.

[4] CABEZAS-WALLSCHEID N, KLIMMECK D, HANSSON J, et al. Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis [J]. Cell Stem Cell, 2014, 15(4): 507-22.

[5] MÜLLER-SIEBURG C E, CHO R H, THOMAN M, et al. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation [J]. Blood, 2002, 100(4): 1302-9.

[6] SANJUAN-PLA A, MACAULAY I C, JENSEN C T, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy [J]. Nature, 2013, 502(7470): 232-6.

[7] SCHOFIELD R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell [J]. Blood Cells, 1978, 4(1-2): 7-25.

[8] MAGUER-SATTA V, PETZER A, EAVES A, et al. BCR-ABL expression in different subpopulations of functionally characterized Ph+ CD34+ cells from patients with chronic myeloid leukemia [J]. Blood, 1996, 88(5): 1796-804.

[9] ZHAO R C, JIANG Y, VERFAILLIE C M. A model of human p210(bcr/ABL)-mediated chronic myelogenous leukemia by transduction of primary normal human CD34(+) cells with a BCR/ABL-containing retroviral vector [J]. Blood, 2001, 97 8(2406-12.

[10] ZHANG X, REN R. Bcr-Abl efficiently induces a myeloproliferative disease and production of excess interleukin-3 and granulocyte-macrophage colony-stimulating factor in mice: a novel model for chronic myelogenous leukemia [J]. Blood, 1998, 92(10): 3829-40.

[11] WONG S, WITTE O N. The BCR-ABL story: bench to bedside and back [J]. Annu Rev Immunol, 2004, 22(247-306.

[12] REN R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia [J]. Nat Rev Cancer, 2005, 5(3): 172-83.

[13] HUGHES T, DEININGER M, HOCHHAUS A, et al. Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results [J]. Blood, 2006, 108(1): 28-37.

[14] ETIENNE G, GUILHOT J, REA D, et al. Long-Term Follow-Up of the French Stop Imatinib (STIM1) Study in Patients With Chronic Myeloid Leukemia [J]. J Clin Oncol, 2017, 35(3): 298-305.

[15] ROSS D M, MASSZI T, GÓMEZ CASARES M T, et al. Durable treatment-free remission in patients with chronic myeloid leukemia in chronic phase following frontline nilotinib: 96-week update of the ENESTfreedom study [J]. J Cancer Res Clin Oncol, 2018, 144(5): 945-54.

[16] CHOMEL J C, BONNET M L, SOREL N, et al. Leukemic stem cell persistence in chronic myeloid leukemia patients with sustained undetectable molecular residual disease [J]. Blood, 2011, 118(13): 3657-60.

[17] BHATIA R, HOLTZ M, NIU N, et al. Persistence of malignant hematopoietic progenitors in chronic myelogenous leukemia patients in complete cytogenetic remission following imatinib mesylate treatment [J]. Blood, 2003, 101(12): 4701-7.

[18] KUMARI A, BRENDEL C, HOCHHAUS A, et al. Low BCR-ABL expression levels in hematopoietic precursor cells enable persistence of chronic myeloid leukemia under imatinib [J]. Blood, 2012, 119(2): 530-9.

[19] SOVERINI S, MANCINI M, BAVARO L, et al. Chronic myeloid leukemia: the paradigm of targeting oncogenic tyrosine kinase signaling and counteracting resistance for successful cancer therapy [J]. Mol Cancer, 2018, 17(1): 49.

[20] MAHON F X, BELLOC F, LAGARDE V, et al. MDR1 gene overexpression confers resistance to imatinib mesylate in leukemia cell line models [J]. Blood, 2003, 101(6): 2368-73.

[21] WHITE D L, SAUNDERS V A, DANG P, et al. Most CML patients who have a suboptimal response to imatinib have low OCT-1 activity: higher doses of imatinib may overcome the negative impact of low OCT-1 activity [J]. Blood, 2007, 110(12): 4064-72.

[22] HAMILTON A, HELGASON G V, SCHEMIONEK M, et al. Chronic myeloid leukemia stem cells are not dependent on Bcr-Abl kinase activity for their survival [J]. Blood, 2012, 119(6): 1501-10.

[23] HOLYOAKE T L, VETRIE D. The chronic myeloid leukemia stem cell: stemming the tide of persistence [J]. Blood, 2017, 129(12): 1595-606.

[24] HOLYOAKE T, JIANG X, EAVES C, et al. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia [J]. Blood, 1999, 94(6): 2056-64.

[25] HOLYOAKE T L, JIANG X, JORGENSEN H G, et al. Primitive quiescent leukemic cells from patients with chronic myeloid leukemia spontaneously initiate factor-independent growth in vitro in association with up-regulation of expression of interleukin-3 [J]. Blood, 2001, 97(3): 720-8.

[26] GRAHAM S M, JØRGENSEN H G, ALLAN E, et al. Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro [J]. Blood, 2002, 99(1): 319-25.

[27] ROEDER I, HORN M, GLAUCHE I, et al. Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications [J]. Nat Med, 2006, 12(10): 1181-4.

[28] JØRGENSEN H G, COPLAND M, ALLAN E K, et al. Intermittent exposure of primitive quiescent chronic myeloid leukemia cells to granulocyte-colony stimulating factor in vitro promotes their elimination by imatinib mesylate [J]. Clinical cancer research, 2006, 12(2): 626-33.

[29] HOLTZ M, FORMAN S J, BHATIA R. Growth factor stimulation reduces residual quiescent chronic myelogenous leukemia progenitors remaining after imatinib treatment [J]. Cancer Res, 2007, 67(3): 1113-20.

[30] NAKA K, HOSHII T, MURAGUCHI T, et al. TGF-β–FOXO signalling maintains leukaemia-initiating cells in chronic myeloid leukaemia [J]. Nature, 2010, 463(7281): 676-80.

[31] ZHANG B, STRAUSS A C, CHU S, et al. Effective targeting of quiescent chronic myelogenous leukemia stem cells by histone deacetylase inhibitors in combination with imatinib mesylate [J]. Cancer Cell, 2010, 17(5): 427-42.

[32] NEVIANI P, HARB J G, OAKS J J, et al. PP2A-activating drugs selectively eradicate TKI-resistant chronic myeloid leukemic stem cells [J]. J Clin Invest, 2013, 123(10): 4144-57.

[33] PSAILA B, WANG G, RODRIGUEZ-MEIRA A, et al. Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets [J]. Mol Cell, 2020, 78(3): 477-92.e8.

[34] BRANCOLINI C, BOTTEGA S, SCHNEIDER C. Gas2, a growth arrest-specific protein, is a component of the microfilament network system [J]. J Cell Biol, 1992, 117(6): 1251-61.

[35] BRANCOLINI C, SCHNEIDER C. Phosphorylation of the growth arrest-specific protein Gas2 is coupled to actin rearrangements during Go-->G1 transition in NIH 3T3 cells [J]. J Cell Biol, 1994, 124(5): 743-56.

[36] JANSSEN J J W M, KLAVER S M, WAISFISZ Q, et al. Identification of genes potentially involved in disease transformation of CML [J]. Leukemia, 2005, 19(6): 998-1004.

[37] CONRAD P A, GIULIANO K A, FISHER G, et al. Relative distribution of actin, myosin I, and myosin II during the wound healing response of fibroblasts [J]. Journal of Cell Biology, 1993, 120(6): 1381-91.

[38] MANZOW S, BRANCOLINI C, MARKS F, et al. Expression of growth arrest-specific (Gas) genes in murine keratinocytes: Gas2 is specifically regulated [J]. Exp Cell Res, 1996, 224(1): 200-3.

[39] KONG Y, ZHAO S, TIAN H, et al. GAS2 Promotes Cell Proliferation and Invasion and Suppresses Apoptosis in Pediatric T-Cell Acute Lymphoblastic Leukemia and Activates Wnt/β-Catenin Pathway [J]. OncoTargets and therapy, 2020, 13(1099-108.

[40] BENETTI R, DEL SAL G, MONTE M, et al. The death substrate Gas2 binds m-calpain and increases susceptibility to p53-dependent apoptosis [J]. The EMBO journal, 2001, 20(11): 2702-14.

[41] PETROULAKIS E, PARSYAN A, DOWLING R J, et al. p53-dependent translational control of senescence and transformation via 4E-BPs [J]. Cancer Cell, 2009, 16(5): 439-46.

[42] KONDO Y, SHEN L, CHENG A S, et al. Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation [J]. Nat Genet, 2008, 40(6): 741-50.

[43] SGORBISSA A, BENETTI R, MARZINOTTO S, et al. Caspase-3 and caspase-7 but not caspase-6 cleave Gas2 in vitro: implications for microfilament reorganization during apoptosis [J]. J Cell Sci, 1999, 112(23): 4475-82.

[44] HUANG W, BEI L, HJORT E E, et al. Decreased calpain activity in chronic myeloid leukemia impairs apoptosis by increasing survivin in myeloid progenitors and xiap1 in differentiating granulocytes [J]. Oncotarget, 2017, 8(31): 50629-41.

[45] HUANG W, ZHOU W, SABERWAL G, et al. Interferon consensus sequence binding protein (ICSBP) decreases beta-catenin activity in myeloid cells by repressing GAS2 transcription [J]. Mol Cell Biol, 2010, 30(19): 4575-94.

[46] HJORT E E, HUANG W, HU L, et al. Bcr-abl regulates Stat5 through Shp2, the interferon consensus sequence binding protein (Icsbp/Irf8), growth arrest specific 2 (Gas2) and calpain [J]. Oncotarget, 2016, 7(47): 77635-50.

[47] RADICH J P, DAI H, MAO M, et al. Gene expression changes associated with progression and response in chronic myeloid leukemia [J]. Proc Natl Acad Sci U S A, 2006, 103(8): 2794-9.

[48] GUGLIELMELLI P, ZINI R, BOGANI C, et al. Molecular profiling of CD34+ cells in idiopathic myelofibrosis identifies a set of disease-associated genes and reveals the clinical significance of Wilms' tumor gene 1 (WT1) [J]. Stem Cells, 2007, 25(1): 165-73.

[49] SUN L, ZHOU H, LIU H, et al. GAS2–Calpain2 axis contributes to the growth of leukemic cells [J]. Acta Biochimica et Biophysica Sinica, 2015, 47(10): 795-804.

[50] MOROTTI A, PANUZZO C, FAVA C, et al. Kinase-inhibitor–insensitive cancer stem cells in chronic myeloid leukemia [J]. Expert Opinion on Biological Therapy, 2014, 14(3): 287-99.

[51] CRAMER-MORALES K, NIEBOROWSKA-SKORSKA M, SCHEIBNER K, et al. Personalized synthetic lethality induced by targeting RAD52 in leukemias identified by gene mutation and expression profile [J]. Blood, 2013, 122(7): 1293-304.

[52] HSIAU T, CONANT D, ROSSI N, et al. Inference of CRISPR Edits from Sanger Trace Data [J]. bioRxiv, 2019, 251082.

[53] ITO K, BERNARDI R, MOROTTI A, et al. PML targeting eradicates quiescent leukaemia-initiating cells [J]. Nature, 2008, 453(7198): 1072-8.

[54] ESSERS M A, OFFNER S, BLANCO-BOSE W E, et al. IFNalpha activates dormant haematopoietic stem cells in vivo [J]. Nature, 2009, 458(7240): 904-8.

[55] GUGLIELMELLI P, ZINI R, BOGANI C, et al. Molecular profiling of CD34+ cells in idiopathic myelofibrosis identifies a set of disease-associated genes and reveals the clinical significance of Wilms' tumor gene 1 (WT1) [J]. Stem Cells, 2007, 25(1): 165-73.

[56] SINHA S, THOMAS D, YU L, et al. Mutant WT1 is associated with DNA hypermethylation of PRC2 targets in AML and responds to EZH2 inhibition [J]. Blood, 2015, 125(2): 316-26.

[57] TIROSH I, IZAR B, PRAKADAN S M, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq [J]. Science, 2016, 352(6282): 189-96.