The long non-coding RNA landscape in juvenile myelomonocytic leukemia

Juvenile myelomonocytic leukemia (JMML) is a rare and aggressive myelodysplastic and myeloproliferative disorder of early childhood. It is characterized by proliferation of granulocytic and monocytic cells.[1] Currently, hematopoietic stem cell transplantation (HSCT) is the standard of care and results in long-term overall survival (OS) of only 50-60%.[2] In approximately 90-95% of patients, hyperactivation of the *RAS/MAPK* pathway can be observed through typical mutations in *NRAS* and *KRAS* (20-25%), *PTPN11* (35%), *NF1* (10-15%) or *CBL* (10-15%) genes. This underlying molecular defect can in part be used to guide patient management (e.g. no HSCT in patients with mutations in *CBL*).[2] Additional somatic mutations, including mutations in *SETBP1, JAK3* and members of the polycomb repressive complex 2 (*PRC2*), are associated with worse outcome.[3,4] In addition, *LIN28B* overexpression and DNA hypermethylation profiles can identify JMML patients with a more aggressive clinical course.[5-7]

The non-coding transcriptome consists of a variety of different RNA types such as tRNAs, snoRNAs, microRNAs, circRNAs and long non-coding RNAs (lncRNAs). LncRNAs are a class of RNA genes with a minimum length of 200 nucleotides.[8] Although only a fraction of known lncRNAs have been functionally characterized, there is growing evidence of their involvement in a variety of biological processes, human diseases and malignancies.[9]

To explore lncRNA expression in JMML, RNA from isolated mononuclear cell preparations from 44 previously untreated JMML patients (median age: 2.02 years) and 7 bone marrow samples from healthy controls (siblings screened for transplantation; median age 6.83 years) was profiled on a custom designed Agilent (8x60K) microarray platform described by Volders *et al.*[8] The dataset (GEO Series accession n. GSE71449) had been previously used in an mRNA expression analysis, leading to the identification of overexpression of *LIN28B* in a fetal-like subgroup of JMML patients.[5] Details of patients' characteristics can be found in the *Online Supplementary Table S1*.

Only probes (n=23,728) representing known lncRNAs were withheld for further analyses. Therefore, the probes were mapped against the most up-dated version of different lncRNA databases, including lncipedia (*www.lncipedia.org*), Hugo Gene Nomenclature Committee (HGNC), and Long Noncoding RNA Database v.2.0 (*www.lncrnadb.org*). The lncipedia 3.1 annotations for lncRNAs have been used throughout this manuscript. Differential gene expression analysis, with samr package (R BioConductor) and BRB-Array tools, revealed a total of 300 differentially expressed lncRNAs (adjusted $P<0.05$; FDR<0.10) between JMML patients and controls, of which 15 up-regulated and 285 down-regulated in JMML patients (Figure 1A and B and *Online Supplementary Table S2*). The 5 most up-regulated

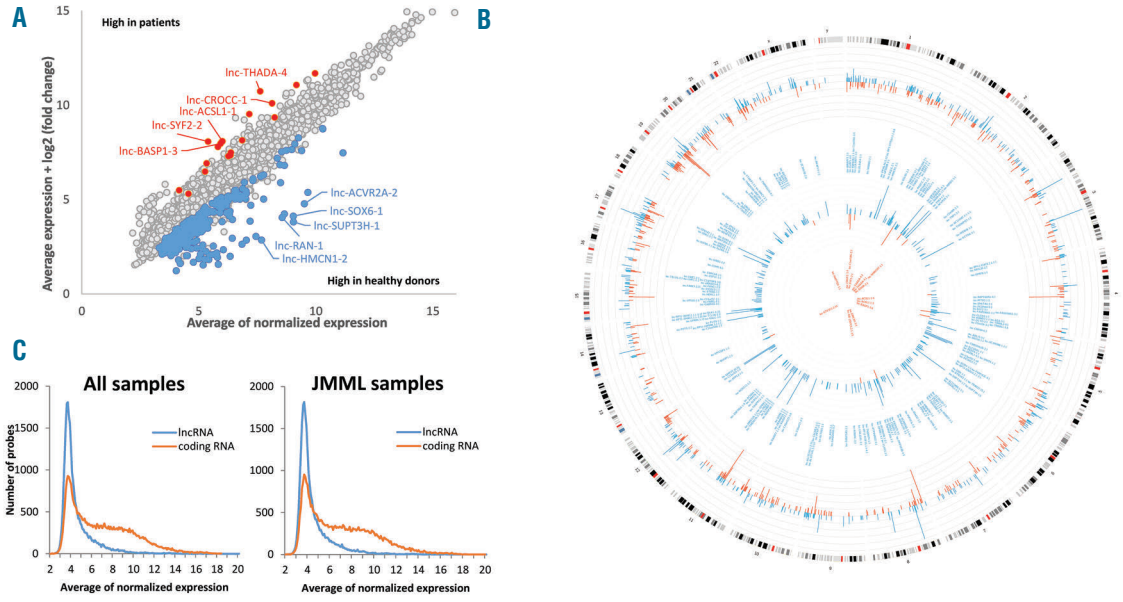

Figure 1. Long non-coding RNA transcriptome in juvenile myelomonocytic leukemia (JMML). (A) Diagonal plot representing the average expression *versus* the log2 fold change added to the average expression of lncRNAs when comparing JMML (n=44) to healthy donors (n=7). lncRNAs significantly up-regulated in patients are represented in red and lncRNAs significantly down-regulated in blue. The names of the top 5 up- and down-regulated lncRNAs based on log2 fold change are indicated. (B) Circos plot showing lncRNAs and mRNAs on human chromosomes. From the outside in, the first layer of the Circos plot represents a chromosome map of the human genome. Black and white bars are chromosome cytobands, and red bars represent centromeres. The second circle shows all differentially expressed coding genes when comparing JMML patients with healthy donors, and the inner circle shows all differentially expressed lncRNAs. Up-regulated lncRNA and mRNA genes are represented in red and down-regulated genes in blue. The height of the bars is representative for the fold change between JMML and healthy donors. Names of lncRNA genes with fold change of more than 2 between JMML and healthy donors are indicated between the circles with mRNA and lncRNA genes for the down-regulated lncRNAs and in the middle of the plot for the up-regulated lncRNAs. (C) Density plots representing the average of the normalized expression of all lncRNAs and mRNAs detected by the array in all samples (left) and in JMML samples only (right).

lncRNAs based on log2 fold change were *lnc-THADA-4, lnc-SYF2-2, lnc-BASP1-3, lnc-ACSL1-1* and *lnc-CROCC-1*, whereas *lnc-SUPT3H-1, lnc-SOX6-1, lnc-HMCN1-2, lnc-ACVR2A-2* and *lnc-RAN-1* were the top 5 most down-regulated lncRNAs. As shown in other disease entities, the mean expression level of lncRNAs was nearly 2-fold lower as compared to mRNA (P<0.0001) (Figure 1C). To exclude the impact of differences in cellular composition between JMML and healthy controls, we compared our dataset with an lncRNA expression atlas of flow cytometry sorted blood cell populations (*Online Supplementary Table S3*).[10] This analysis revealed 2 of 15 differentially up-regulated lncRNAs in JMML to be specific for granulocytic lineage (*lnc-ACSL1-1* and *lnc-BASP1-3*) and none for monocytic differentiation. Conversely, one monocytic specific lncRNA showed significant downregulation in JMML patients (*lnc-AL713998.1-2*).

To explore the functional characteristics of differentially expressed lncRNAs, we used a guilt-by-association approach, based on the hypothesis that non-coding and protein coding genes belonging to the same biological pathways are co-ordinately regulated. Spearman's rho values were calculated between lncRNAs of interest and all protein coding genes on the array. The resulting ranked gene lists served as inputs for pre-ranked gene-set enrichment analysis (GSEA),[11] using the MSigDB.v.6.0 C2 collection and FDR<0.1. This revealed an overlap between the 15 curated gene sets most positively correlated with the top 5 up-regulated lncRNAs (Figure 2 and *Online Supplementary Table S4*). Using the same approach for the top 5 down-regulated lncRNAs, a strong overlap both in gene sets positively correlated with these lncRNAs, as well as negatively correlated was observed (Figure 2 and *Online Supplementary Table S4*). Interestingly, several gene sets showed positive correlation with up-regulated lncRNAs as well as negative correlation with down-regulated lncRNAs, suggesting a common pathway. Notably, these gene sets are related to myeloid cell development, dendritic cell maturation, neutrophil function at skin wound, systemic lupus erythematosus, and different acute myeloid leukemia (AML) phenotypes.

Extensive cytogenetic and mutational data were available for our patient cohort and we evaluated whether the expression of lncRNAs was associated with specific JMML subgroups based on clinical, cytogenetic and mutational characteristics. First, patients with elevated fetal hemoglobin (HbF) levels for age were compared with patients with normal HbF levels. Only 8 lncRNAs were found differentially expressed between these two

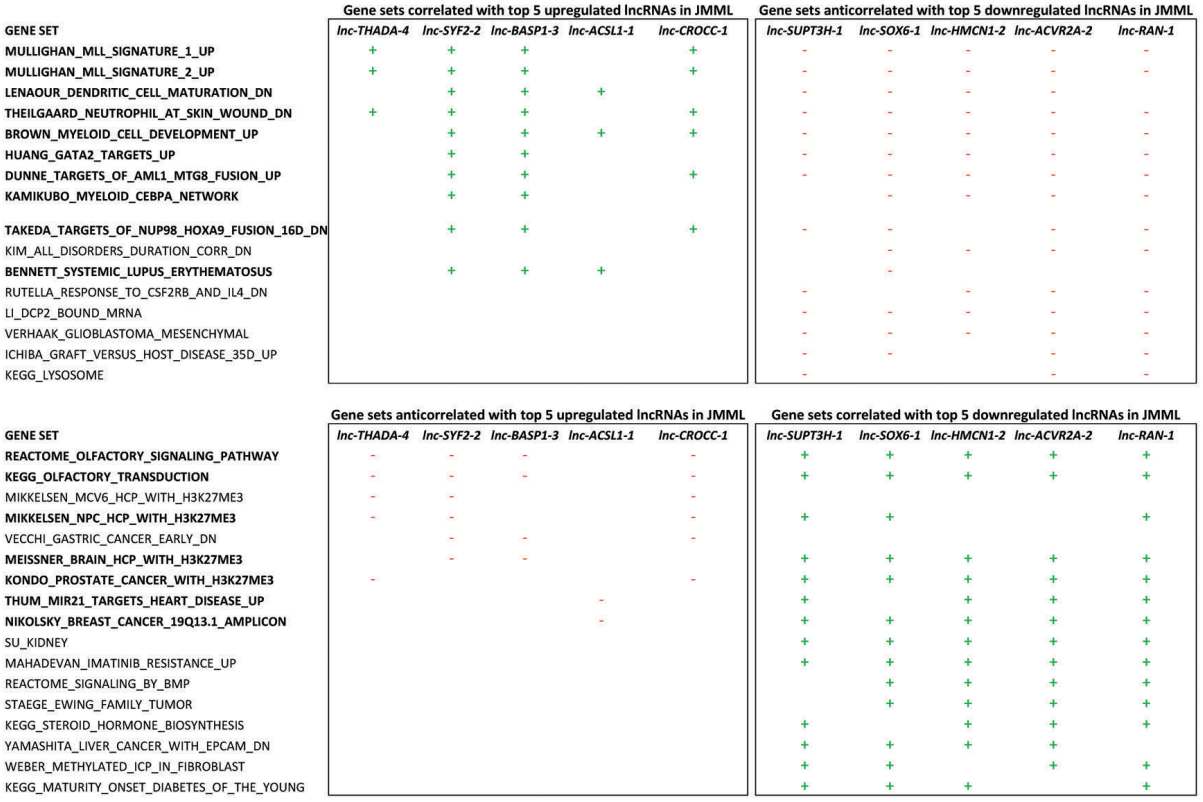

Figure 2. Schematic overview of the overlap between curated gene sets positively and negatively correlating with the top 5 up- and down-regulated lncRNAs. The overlap between curated gene sets positively and negatively correlating with the top 5 up-regulated lncRNAs (*lnc-THADA-4, lnc-SYF2-2, lnc-BASP1-3, lnc-ACSL1-1* and *lnc-CROCC-1*) and the top 5 down-regulated lncRNAs (*lnc-SUPT3H-1, lnc-SOX6-1, lnc-HMCN1-2, lnc-ACVR2A-2* and *lnc-RAN-1*) in juvenile myelomonocytic leukemia (JMML) compared to healthy controls. Only gene sets with overlap in 3 or more lncRNAs are presented. The green '+' symbol indicates a positive correlation between the gene set and the lncRNA, and the red '-' symbol indicates a negative correlation between the gene set and the lncRNA. Gene sets (anti-)correlating both with up-regulated and down-regulated lncRNAs are written in bold. Marked overlap between these gene sets can be observed suggesting a common pathological pathway.

groups, all up-regulated in the group with normal HbF levels for age (adjusted $P<0.05$) (*Online Supplementary Table S5*). Given the strong downregulation of *lnc-C2orf54-1* and *lnc-ATP6V0D2-2* in JMML patients with elevated HbF for age (log fold change -7.4 and -2.8, respectively), we predicted that these lncRNA would be anti-correlated with genes involved in HbF regulation. Indeed, *HBG1* (Spearman rho value (ρ): -0.63 and -0.61), *HBBP1* (Hemoglobin Subunit Beta Pseudogene 1; ρ: -0.60 and -0.41), *HBE1* (ρ: -0.37 and -0.45) and *LIN28B* (ρ: -0.58 and -0.70), belonged to the top anti-correlated protein coding genes, suggesting that both *lnc-C2orf54-1* and *lnc-ATP6V0D2-2* could be part of a molecular network controlling the formation of HbF. Moreover, *lnc-ATP6V0D2-2* is associated with event-free survival (EFS) in cytogenetically normal AML and is down-regulated in patients with mutations in *RUNX1*.[12] Subsequently, patients with monosomy 7 were compared to patients with a normal karyotype. To avoid confounding effects due to RAS gene mutations, a 2-class analysis (monosomy 7 *vs.* normal karyotype) blocking for *RAS* mutations through a randomized block design was applied.[13] As such, different *RAS* mutations are considered one group (block), eliminating variability not of interest in this specific analysis.[13] This analysis revealed a total of 36 lncRNAs significantly up-regulated in the monosomy 7 group and 5 lncRNAs down-regulated. Of note, 6 of 8 lncRNAs (*lnc-C2orf54-1, PRINS, lnc-ATP6V0D2-2, lnc-CCT8-1, lnc-SH3BGRL2-1, lnc-FRMD4A-1*) down-regulated in patients with elevated HbF levels for age were significantly up-regulated in JMML patients with monosomy 7. This observation corresponds with the established inverse relation between monosomy 7 and elevated HbF levels in JMML patients.[2]

Next, we performed pairwise analyses comparing patients with *PTPN11* (n=18), *KRAS* (n=11), *NRAS* (n=8), *NF1* (n=4) mutations or without any of the above mutations (quadruple negative; n=3) to all other patients (*Online Supplementary Table S5*). Patients harboring a *KRAS* mutation showed 26 differentially expressed lncRNAs (17 up-regulated and 9 down-regulated) compared to JMML patients without this mutation. *NF1* mutated patients were characterized by 62 down-regulated lncRNAs whereas none were significantly up-regulated. *PTPN11* mutated patients were marked by differential expression of 8 lncRNAs (6 up-regulated and 2 down-regulated) and 4 *lncRNAs* were differentially expressed in *NRAS* mutated JMML (all up-regulated). Remarkably, in quadruple negative patients, only one lncRNA, *lnc-ATP10A-2*, was differentially up-regulated with high fold change (FC=14.4). This lncRNA is anti-correlated with *BAP1*, a key regulator of many cancer-associated pathways.

Finally, we studied differentially expressed lncRNAs between JMML patients with and without *LIN28B* overexpression. Sixty-one lncRNAs were differentially expressed of which 60 were up-regulated in JMML patients with *LIN28B* overexpression (*Online*

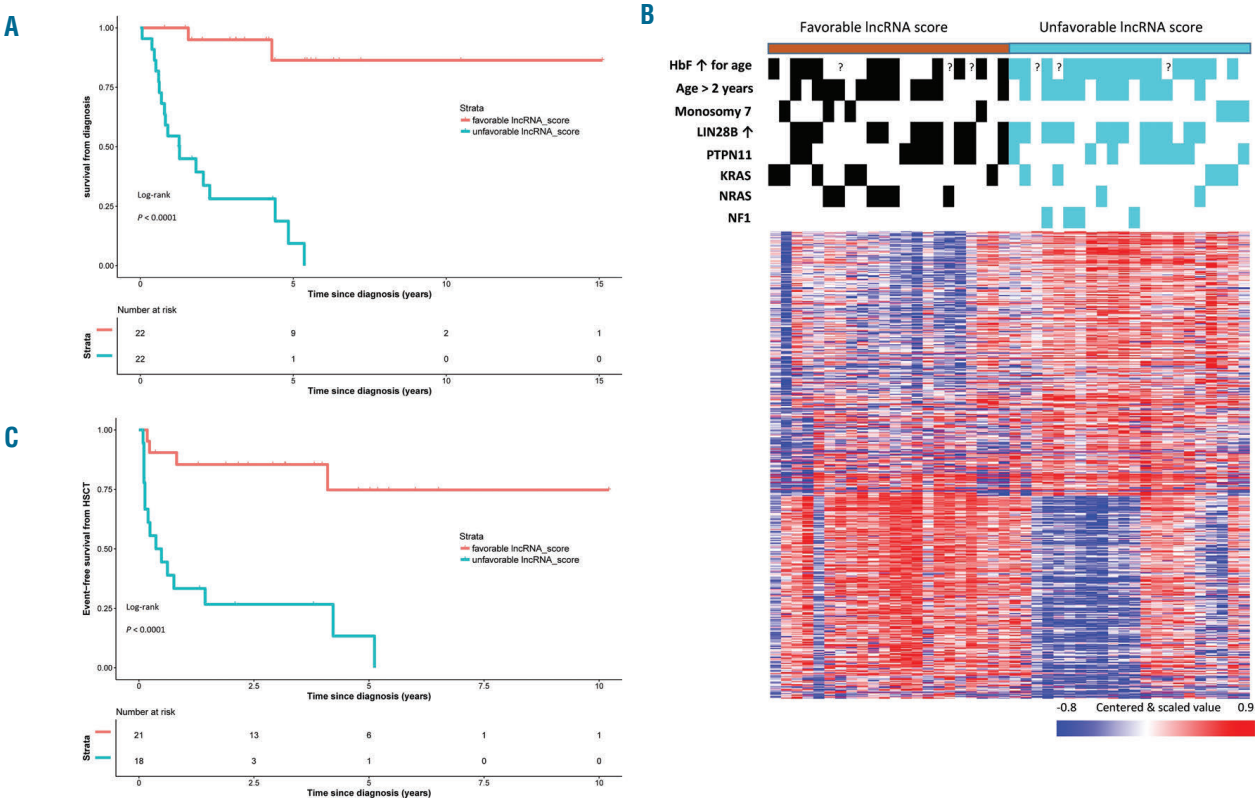

Figure 3. Stratification of juvenile myelomonocytic leukemia (JMML) patients based on lncRNA score. (A and C) Kaplan-Meier curve for overall survival (OS) and event-free survival (EFS) after hematopoietic stem cell transplantation (HSCT) from JMML diagnosis for patients with favorable and unfavorable lncRNA score. (B) Heatmap displaying differential gene expression (n=549 coding genes with an absolute fold change of more than 2; 238 genes up-regulated in patients with favorable lncRNA score and 311 genes up-regulated in patients with unfavorable lncRNA score) between JMML patients with favorable and unfavorable lncRNA score. Clinical annotation for "genotype" (somatic mutations in *PTPN11, KRAS, NRAS* and *NF1*; *LIN28B* overexpression and karyotype) and some clinical characteristics [age group and fetal hemoglobin (HbF) level for age] are depicted on top of the heatmap.

Supplementary Table S5). Interestingly, both *MEG3* and *MEG8* were up-regulated in *LIN28B* overexpression and are known to be involved in regulation of pluripotence and fetal development.[14]

As the prognostic value of coding gene expression signatures in JMML has previously been reported by Bresolin *et al.*,[15] we hypothesized that lncRNA expression might influence clinical outcome. Indeed, univariable Cox regression analysis identified 45 individual lncRNAs highly associated with OS (P<0.001) (*Online Supplementary Table S6*) of which 7 lncRNAs (*UCKL1-AS1, lnc-GPR63-5, lnc-ZSWIM2-3, lnc-RORB-1, lnc-NPFFR2-1, lnc-EMB-3* and *lnc-CCDC90A-2*) remained significantly associated with OS after multivariable Cox regression analysis with age category (< or > 2 years old), HbF status, karyotype and molecular status (P<0.001). Interestingly, these lncRNAs did not associate with known prognostic characteristics and pre-ranked GSEA could not identify a common (anti-)correlated gene set. We devised an lncRNA score, derived as a linear combination of the expression of the significant lncRNAs (method described in Garzon *et al.*[12]) and dichotomized patients based on lncRNA score (scores below the median were classified as favorable and scores above as unfavorable). Patients with unfavorable lncRNA score had a significantly lower OS and EFS after HSCT (log rank test P<0.001) (Figure 3A and C). Next, we compared the expression levels of 29,654 mRNA probes (21,659 unique genes) between patients with favorable and unfavorable lncRNA score. We identified 549 coding genes with differential expression between the two groups (absolute FC ≥ 2) (Figure 3B). Together, these results suggested that lncRNAs might be of prognostic significance in JMML. However, studies in independent patient cohorts are necessary to validate the prognostic value of lncRNA expression in JMML.

In conclusion, we documented for the first time the lncRNA landscape in 44 JMML patients and have associated lncRNA expression with clinical and molecular characteristics. We demonstrated that JMML patients exhibit a distinct lncRNA expression profile compared to healthy controls and different JMML subgroups have distinct lncRNA signatures contributing to disease heterogeneity. Follow-up studies are needed to confirm the prognostic value of lncRNAs and to perform an integrative analysis of lncRNA expression profiles with mutational patterns, DNA methylation profiles and clinical patients' characteristics. This study paves the way for further functional research on the role of lncRNAs in JMML biology, and their diagnostic and therapeutic application.

Mattias Hofmans,[1,2] Tim Lammens,[1,3] Hetty H. Helsmoortel,[4] Silvia Bresolin,[5] Hélène Cavé,[6] Christian Flotho,[7] Henrik Hasle,[8] Marry M. van den Heuvel-Eibrink,[9,10] Charlotte Niemeyer,[7] Jan Stary,[11] Nadine Van Roy,[3,4] Pieter Van Vlierberghe,[3,4] Jan Philippé[2,3] and Barbara De Moerloose[1,3]

[1]Department of Pediatric Hematology-Oncology and Stem Cell Transplantation, Ghent University Hospital, Belgium; [2]Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Belgium; [3]Cancer Research Institute Ghent, Belgium; [4]Center for Medical Genetics, Ghent University Hospital, Belgium; [5]Department of Women and Child Health, University of Padova, Italy; [6]Department of Genetics, University Hospital of Robert Debré and Paris-Diderot University, Paris, France; [7]Division of Pediatric Hematology and Oncology, Department of Pediatrics and Adolescent Medicine, University of Freiburg, Germany; [8]Department of Pediatrics, Aarhus University Hospital Skejby, Aarhus, Denmark; [9]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands; [10]Dutch Childhood Oncology Group, The Hague, the Netherlands and [11]Department of Pediatric Hematology/ Oncology, Charles University and University Hospital Motol, Prague, Czech Republic

Funding: the authors would like to thank the Foundation against Cancer (2016-113, BDM), the Cancer Plan, action 29 (KP_29_020, JP) and vzw Kinderkankerfonds – a non-profit childhood cancer foundation under Belgian law (grant to TL) for supporting this work. MH is a PhD candidate at Ghent University and this work is submitted in partial fulfilment of the requirement for the PhD.

Correspondence: mattias.hofmans@ugent.be
doi:10.3324/haematol.2018.189977

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Niemeyer CM. RAS diseases in children. Haematologica. 2014; 99(11):1653-1662.
2. Locatelli F, Niemeyer CM. How I treat juvenile myelomonocytic leukemia. Blood. 2015;125(7):1083-1090.
3. Stieglitz E, Taylor-Weiner AN, Chang TY, et al. The genomic landscape of juvenile myelomonocytic leukemia. Nat Genet. 2015; 47(11):1326-1333.
4. Caye A, Strullu M, Guidez F, et al. Juvenile myelomonocytic leukemia displays mutations in components of the RAS pathway and the PRC2 network. Nat Genet. 2015;47(11):1334-1340.
5. Helsmoortel HH, Bresolin S, Lammens T, et al. LIN28B overexpression defines a novel fetal-like subgroup of juvenile myelomonocytic leukemia. Blood. 2016;127(9):1163-1172.
6. Lipka DB, Witte T, Toth R, et al. RAS-pathway mutation patterns define epigenetic subclasses in juvenile myelomonocytic leukemia. Nature Commun. 2017;8(1):2126.
7. Stieglitz E, Mazor T, Olshen AB, et al. Genome-wide DNA methylation is predictive of outcome in juvenile myelomonocytic leukemia. Nature Commun. 2017;8(1):2127.
8. Volders PJ, Helsens K, Wang X, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res. 2013;41(Database issue):D246-251.
9. Bhan A, Soleimani M, Mandal SS. Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res. 2017;77(15):3965-3981.
10. Schwarzer A, Emmrich S, Schmidt F, et al. The non-coding RNA landscape of human hematopoiesis and leukemia. Nature Commun. 2017;8(1):218.
11. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545-15550.
12. Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci USA. 2014;111(52):18679-18684.
13. Addelman S. The Generalized Randomized Block Design. Am Stat. 1969;23(4):35-36.
14. Lee S, Seo HH, Lee CY, et al. Human Long Noncoding RNA Regulation of Stem Cell Potency and Differentiation. Stem Cells Int. 2017;2017:6374504.
15. Bresolin S, Zecca M, Flotho C, et al. Gene expression-based classification as an independent predictor of clinical outcome in juvenile myelomonocytic leukemia. J Clin Oncol. 2010;28(11):1919-1927.