Genomics Data 7 (2016) 57–59

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Transcriptome sequencing of hematopoietic stem cells and chronic myelgenous leukemia stem cells

Junil Kim [a], Seong-Jin Kim [a,b], Kazuhito Naka [c,d,*]

[a] CHA Cancer Institute, CHA University, CHA Bio Complex, 335 Pangyo-ro, Bundang-ku, Seongnam, Kyunggi-do 463-400, Republic of Korea
[b] Department of Biomedical Science, CHA University, CHA Bio Complex, 335 Pangyo-ro, Bundang-ku, Seongnam, Kyunggi-do 463-400, Republic of Korea
[c] Exploratory Project on Cancer Stem Cells, Cancer Research Institute, Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan
[d] Department of Stem Cell Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan

ARTICLE INFO

Article history:
Received 9 November 2015
Accepted 22 November 2015
Available online 23 November 2015

Keywords:
Chronic myelogenous leukemia
CML stem cell
Dipeptide transporter
RNA sequencing
Slc15a2

ABSTRACT

Dipeptide species are accumulated in the chronic myelogenous leukemia (CML) stem cells [1]. To investigate the molecular mechanisms of the accumulation of dipeptide species in CML stem cells, we performed transcriptome sequencing of long-term stem cells, short-term stem cells, progenitor cells from healthy control and CML-affected mice (GSE70031). The transcriptome data revealed that the expression of a dipeptide transporter (solute carrier family 15, member 2 (SLC15A2)) was elevated only in the CML stem cells. This result indicates that dipeptide species accumulates in CML stem cells through a dipeptide transporter SLC15A2.

Specifications	
Organism/cell line/tissue	*Mus musculus*/Bone marrow
Sex	
Sequencer or array type	Illumina HiSeq 2000
Data format	RNA sequencing: raw data (Fastq files) and processed data (tab-delimited text files include RPKM values)
Experimental factors	8 RNA samples for RNA sequencing as follows: 2 samples of normal long-term stem cell 1 sample of normal short-term stem cell 1 sample of KLS^- progenitor cell 2 samples of chronic myeloid leukemia long-term stem cell 1 sample of chronic myeloid leukemia short-term stem cell 1 sample of chronic myeloid leukemia KLS^- progenitor cell
Experimental features	Immature KLS^+ cells and KLS^- progenitor cells were obtained from healthy control and CML-affected mice by using FACS Aria III cell sorter.
Consent	
Sample source location	

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70031.

2. Experimental design, materials and methods

2.1. RNA sample preparation and transcriptome sequencing

We isolated the most primitive long-term (LT) stem cells ($CD150^+$-$CD48^-CD135^-KLS^+$ cells), short-term (ST) stem cells ($CD150^-$-$CD48^-CD135^-KLS^+$ cells), and KLS^- progenitor cells from healthy littermate control and CML-affected mice. Eight different RNA samples were extracted from two samples of normal LT stem cells, one sample of normal ST stem cells, one sample of normal KLS^- progenitor cells, two samples of CML LT stem cells, one sample of CML ST stem cells, and one sample of CML KLS^- progenitor cells. Paired-end reads RNA sequencing was performed using Illumina HiSeq2000 for all RNA samples. All sequenced reads were trimmed for adaptor sequence, then mapped to mm9 whole genome using DNAnexus. Reads Per Kilobase of exon per Megabase of library size (RPKM) were calculated using DNAnexus.

2.2. Differentially expressed genes (DEGs)

We identified DEGs by comparing expression levels of CML stem cells with those of three other types of cells (normal stem cells, normal KLS^- progenitor cells, and CML KLS^- progenitor cells). Genes were considered DEGs if their fold-change was more than 2-fold and p-value was less than 0.05. A one-sided two-sample t-test was used to calculate the p-values. From the analysis, we identified 528 up- and 238 down-regulated DEGs in CML stem cells (Fig. 1a). Among up-regulated DEGs,

* Corresponding author at: Department of Stem Cell Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan.
E-mail address: kanaka55@hiroshima-u.ac.jp (K. Naka).

http://dx.doi.org/10.1016/j.gdata.2015.11.017

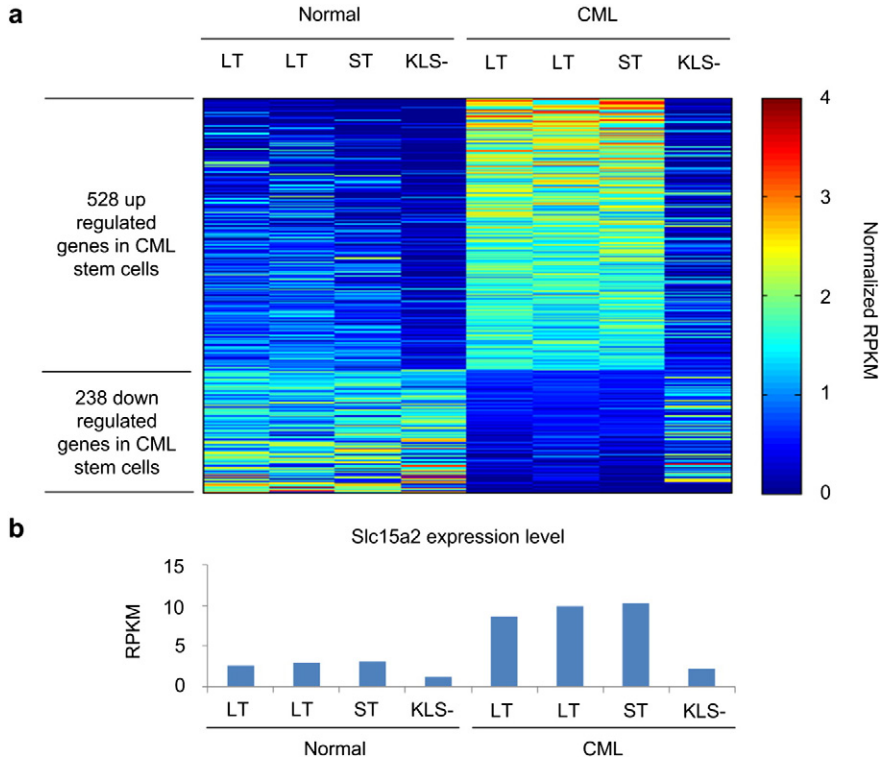

Fig. 1. Differentially expressed genes in chronic myelogenous leukemia (CML) cells. (a) Heat map of up- and down-regulated DEGs in CML stem cells. (b) Slc15a2 expression level. LT, ST, and KLS$^-$ represent long-term stem cell, short-term stem cell, and KLS$^-$ progenitor cell, respectively.

a dipeptide transporter Slc15a2 was highly expressed only in CML stem cells (Fig. 1b). This represents that high expressed Slc15a2 gene causes the accumulation of dipeptide species in CML stem cells.

2.3. Gene ontology (GO) analysis

We identified GO terms enriched in the up- and down-regulated DEGs of CML stem cells using DAVID functional annotation tool [1], respectively. GO analysis revealed that the up-regulated DEGs were associated with GO terms "antigen processing and presentation", "cell adhesion", "sensory perception of light stimulus", and "enzyme linked receptor protein signaling pathway" (Table 1). The down-regulated DEGs were associated with GO terms "nucleosome assembly", "actin cytoskeleton organization", "immune response", and "response to nutrient levels" (Table 2).

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development

Table 1
Gene ontology (GO) terms associated with the up-regulated differentially expressed genes in chronic myelogenous leukemia stem cells.

Term	p-value	Genes
GO:0019882~antigen processing and presentation	1.16E-06	H2-EA, H2-Q10, MILL2, GM8909, H2-TW3, H2-BL, H2-Q1, EG547347, FCGRT, H2-T10, H2-T24, 1500011B03RIK, H2-DMB2, H2-T3
GO:0007156~homophilic cell adhesion	5.45E-04	DSG4, CADM1, FAT2, PCDH9, ROBO2, ESAM, PCDHB12, PCDHGB8, PCDHB21, CDH23, PCDHGA1
GO:0007155~cell adhesion	5.48E-04	CADM1, PKHD1, CLDN5, PCDHB12, TGFB2, PCDHGA1, CGREF1, LAMB2, FAT2, ROBO2, ESAM, DPT, CDH23, CNTN5, INPPL1, PCDH9, PCDHGB8, EMILIN2, GPR98, PCDHB21, THY1, NCAM2, DSG4, LAMA3, OTOG, CNTN4, PERP, AOC3
GO:0050953~sensory perception of light stimulus	0.0033091	TULP1, PDE6B, EPAS1, ABCA4, DTNBP1, USH2A, GPR98, NYX, CDH23
GO:0007167~enzyme linked receptor protein signaling pathway	0.0067592	FGFR2, EGFR, EFNA1, LTBP4, ZFP128, EPHB4, EPHA2, TGFB2, IGSF10, EPHA4, EPHA6, DOK4, PDGFRB, TGFA, PDGFC
GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway	0.0071313	IGSF10, EGFR, FGFR2, EPHA4, EPHA6, DOK4, EFNA1, TGFA, PDGFRB, PDGFC, EPHB4, EPHA2
GO:0002474~antigen processing and presentation of peptide antigen via MHC class I	0.0074334	H2-Q10, GM8909, H2-TW3, H2-Q1, H2-T3

Table 2
Gene ontology (GO) terms associated with the down-regulated differentially expressed genes in chronic myelogenous leukemia stem cells.

Term	P-value	Genes
GO:0006334~nucleosome assembly	7.60E-08	HIST1H2AB, HIST1H2BB, HIST1H2BC, HIST1H2BG, A730008H23RIK, HJURP, HIST2H2AC, HIST1H2BJ, HIST3H2A, HIST1H4C, HIST1H3E, HIST1H3F, HIST1H4I, HIST3H2BA
GO:0030036~actin cytoskeleton organization	3.88E-04	CNN3, MYBPC3, GHRL, SH2B2, EVL, CSRP1, PROX1, DAAM2, CAPN3
GO:0006955~immune response	0.0045826	MASP2, IL1RN, MYO1F, RSAD2, TLR5, NLRP3, CXCL10, CFP, H2-T9, OASL1, CD300LG, LBP, CLEC4D
GO:0031667~response to nutrient levels	0.0077562	UGT1A2, PCSK9, GHRL, VARS, KLF4, LEFTY1

Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant numbers: HI14C3426 and HI14C2640), and by a grant from the Uehara Memorial Foundation, and a Grant-in-Aid for Scientific Research (B) from MEXT (26290038), Japan.

Reference

[1] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (1) (2009) 44–57.