Research paper

Expression of ERV3-1 in Leukocytes of Acute Myelogenous Leukemia Patients

So Nakagawa^{a,b†}, Masaharu Kawashima^{c,d†}, Yuji Miyatake^{d†}, Kai Kudo^d, Ryutaro Kotaki^{d¶}, Kiyoshi Ando^{b,e}, Ai Kotani^{b,d*}

^a Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan

^b Institute of Medical Sciences, Tokai University, Isehara, Kanagawa 259-1193, Japan

^c Division of Clinical Oncology and Hematology, The Jikei University School of Medicine, Minato-ku, Tokyo 105-8471, Japan

^d Department of Hematological Malignancy, Institute of Medical Science, Tokai University, Isehara, Kanagawa 259-1193, Japan

^e Department of Hematology and Oncology, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan

[†] These authors contributed equally to this work.

[¶] Present address: Department of Immunology, Duke University, Durham, NC 27710, USA

* To whom correspondence should be addressed:

So Nakagawa

so@tokai.ac.jp

Tel: +81-463-93-1121 ext. 2661

Fax: +81-463-93-5418

Ai Kotani

aikotani@k-lab.jp

Tel. 81-463-93-1121 ext. 2781

35 **Abstract**

36 Acute myelogenous leukemia (AML) is one of the major hematological malignancies. In
 37 the human genome, several have been found to originate from retroviruses, and some
 38 of which are involved in progression of various cancers. Hence, to investigate whether
 39 retroviral-like genes are associated with the development of AML, we conducted a
 40 transcriptome sequencing analysis of 12 retroviral-like genes of 150 AML patients using
 41 The Cancer Genome Atlas database. We found high expression of ERV3-1, an
 42 envelope gene of endogenous retrovirus group 3 member 1. In particular, blood and
 43 bone marrow cells of the myeloid lineage in AML patients, exhibited higher expression
 44 of ERV3-1 than those of the monocytic AML lineage. We also examined the protein
 45 expression of ERV3-1 by immunohistochemical analysis and found expression of
 46 ERV3-1 protein in 7 out of 12 AML patients, with a particular concentration observed at
 47 the membrane of some leukemic cells. Transcriptome analysis further suggested that
 48 upregulated ERV3-1 expression may be associated with chromosome 8 trisomy as
 49 anomaly was found to be more common among the high expression group compared to
 50 the low expression group. However, this finding was not corroborated by the
 51 immunohistochemical data. This discrepancy may have been caused, in part, by the
 52 small number of samples analyzed in this study. Although the precise associated
 53 molecular mechanisms remain unclear, our results suggest that ERV3-1 may be
 54 involved in AML development.

55 **(226 words)**

56

57 **Keywords:** endogenous retrovirus, acute myelogenous leukemia, cancer development,
 58 immunosuppression

59

60 **Highlights**

- 61 ● Expression of 12 retroviral-like genes in the human genome were analyzed using
- 62 transcriptome data of 150 acute myelogenous leukemia (AML) patients.
- 63 ● ERV3-1, an envelope gene of endogenous retrovirus group 3 member 1, was found
- 64 to uniquely show high expression level.
- 65 ● Morphologic characteristics and chromosomal abnormalities are found to be related
- 66 with the expression of ERV3-1.

67

68 1. Introduction

69 Approximately 8% of the human genome corresponds to retroviral origins
70 (Lander et al. 2001). These areas of the genome are referred to as long terminal repeat
71 (LTR) retrotransposons, many of which correspond to human endogenous retroviruses
72 (HERVs). HERVs originally derived from retroviruses that infected germline cells of the
73 host species. Therefore, HERVs contain retroviral genetic elements including
74 cis-regulatory regions (LTRs) in their 5' and 3' terminals, as well as several coding
75 sequences: gag, protease, polymerase, and envelope. The structures of other LTR
76 retrotransposons are similar to that of HERV except for the absence of an envelope
77 gene. Generally, HERVs are incapable of generating infectious virions that can
78 competently replicate in human cells due to the accumulation of multiple mutations
79 during evolution (Tönjes et al. 1999). Therefore, such retroviral sequences are believed
80 to be “junk” DNA in the human genome. However, many recent studies showed that
81 certain sequences, similar to those of retroviruses, have obtained new functions in the
82 hosts.

83 In the human genome, at least 12 retroviral-like genes are annotated in the
84 GRCh38 assembly provided by National Center for Biotechnology Information (NCBI):
85 *ARC*, *ASPRV1/SASpase*, *ERV3-1*, *ERVK13-1*, *ERVH48-1/Suppressyn*,
86 *ERVMER34-1/HEMO*, *ERVV-1*, *ERVV-2*, *ERVW-1/Syncytin-1*, *ERVFRD-1/Syncytin-2*,
87 *PEG10/SIRH1*, *PEG11/RTL1/SIRH2*, *RTL4/ ZCCHC16/SIRH11* and *SIRH7/LDOC1*.
88 *ERVW-1/Syncytin-1* and *ERVFRD-1/Syncytin-2* are the most well studied retroviral-like
89 genes corresponding to retroviral envelope genes (Mi et al. 2000, Blaise et al. 2003),
90 both of which are involved in human placenta development. Specifically, these genes
91 are associated with cell-cell fusion and immunosuppression, both of which function are
92 quite similar to those operated by envelope proteins of retroviruses (Kim et al. 2004).

93 Those molecular functions may be also related to cancer progression. Indeed,
 94 Syncytin-1 and Syncytin-2 are reported to be involved in cancer development (Larsen et
 95 al. 2009). In addition, an LTR retrotransposon-derived PEG10/Sirh1 that is similar to a
 96 gag-pro-like gene is involved in placenta development (Ono et al. 2006), as well as in
 97 the progression of various cancers including pancreatic carcinoma, breast cancer,
 98 prostate cancer, gallbladder carcinoma, thyroid cancer, oral squamous cell carcinoma,
 99 colon cancer, enchondromas, and B-cell chronic lymphocytic leukemia (reviewed in Xie
 100 et al. 2018). Indeed, these retroviral-like genes originate from viruses making their
 101 unexpected expression potentially harmful to humans (Gonzalez-Cao et al. 2016).

102 Acute myelogenous leukemia (AML) is one of the major hematological
 103 malignancies, characterized by overproduction of myeloid progenitor cells in the bone
 104 marrow, which then rapidly migrates to the blood, and in some cases, can spread to
 105 other organs, such as liver and spleen. AML is associated with curative rates of 35 to
 106 40% in patients aged < 60 years (Döhner et al. 2010); however, the number of AML
 107 patients increase with age, and 70% of patients $\geq$ 65 years die of the disease within a
 108 year, despite treatment (Meyers et al. 2013). The French-American-British (FAB)
 109 classification system is a standard classification of AML patients that are divided into
 110 eight different subtypes (M0 through M7) based on morphologic characteristics (Bennett
 111 et al. 1976): undifferentiated acute myeloblastic leukemia (M0), acute myeloblastic
 112 leukemia with minimal maturation (M1), acute myeloblastic leukemia with maturation
 113 (M2), acute promyelocytic leukemia (M3), acute myelomonocytic leukemia (M4), acute
 114 monocytic leukemia (M5), acute erythroid leukemia (M6), and acute megakaryoblastic
 115 leukemia (M7).

116 Although numerous studies suggested relationships between HERVs and
 117 leukemia including AML (Depil et al. 2002; Chen et al. 2013; Bergallo et al. 2017;

118 Cuellar et al. 2017; Deniz et al. 2020), details regarding the roles of retroviral-like genes
 119 in AML remain unclear, particularly as they pertain to the different AML subtypes.
 120 Therefore, in this study, we evaluated expression of retroviral-like genes in leukocytes of
 121 AML patients that are potentially harmful to AML. To this end, we first examined
 122 RNA-seq data obtained from 150 AML patients that were downloaded from The Cancer
 123 Genome Atlas (TCGA) database (<https://www.cancer.gov/tcga>). We then screened the
 124 expression of the abovementioned 12 retroviral-like genes and statistically examined
 125 the relationship between the expression levels and FAB subtypes, with exception of M6
 126 and M7 cases, as they are relatively rare in AML (< 5%) (Bennett et al. 1976). We
 127 further validated the protein expression of the highly expressed retroviral-like gene in
 128 the leukemic cells obtained from AML patients by immunostaining and investigated
 129 whether the gene could be related to the progress of AML.

130

131

132 **2. Materials and Methods**

133 **2.1 Ethics**

134 This study was approved by the Institutional Review Board of Tokai University School of
135 Medicine, of which protocol numbers are 15-I-26, 18-I-08 and 19-R-323 for
136 immunohistochemistry and clinical sequencing data analyses of AML patients. Informed
137 consent was provided according to the Helsinki Declaration in the Tokai University
138 Hospital.

139

140 **2.2 Cancer genome data analysis**

141 Sequence and annotation data of the human genome GRCh38 was downloaded from
142 the Illumina iGenomes
143 (https://support.illumina.com/sequencing/sequencing_software/igenome.html). We also
144 obtained RNA-seq data and clinical record data for 150 AML patients from TCGA-LAML
145 database (<https://portal.gdc.cancer.gov/projects/TCGA-LAML>), which are summarized
146 in the Supplementary data (Table S1 – S3). In this study, we used the RNA-seq data of
147 which sequences are mapped to the human genome GRCh38 (BAM files) using STAR
148 2 (Dobin et al. 2013) provided by TCGA-LAML. We counted the mapped reads based
149 on the gene annotation, and computed expression scores of TPM (transcripts per
150 million) using StringTie2 version 2.0.6 (Kovaka et al. 2019). We extracted the TPM
151 scores of 12 retroviral-like genes: *ARC*, *ASPRV1/SASpase*, *ERV3-1*, *ERVK13-1*,
152 *ERVH48-1/Suppressyn*, *ERMER34-1/HEMO*, *ERVV-1*, *ERVV-2*, *ERVW-1/Syncytin-1*,
153 *ERVFRD-1/Syncytin-2*, *PEG10/SIRH1*, *PEG11/RTL1/SIRH2*, *RTL4/ZCCHC16/SIRH11*,
154 and *SIRH7/LDOC1*, which are also summarized in the Supplementary data (Table S1).
155 The TPM scores were log-transformed as follows: $\log_2(\text{TPM}+1)$. Using the
156 log-transformed TPM scores, we generated a heatmap of 12 retroviral-like genes using

157 the heatmap.2 program in the gplots package of R (<https://github.com/talgalili/gplots>).

158

159 **2.3 Statistical analysis**

160 Normal variables were assessed by Fisher's exact test. Continuous variables were
161 assessed by Mann-Whitney *U* test or Kruskal-Wallis test for two or multiple groups,
162 respectively. Data are presented as the mean $\pm$ standard deviation (SD). A *P* value <
163 0.05 was considered statistically significant.

164

165 **2.4 Immunohistochemistry**

166 To confirm ERV3-1 protein expression in AML, immunohistochemical (IHC) staining of
167 ERV3-1 was performed on 12 cases of AML patients using paraffin-embedded bone
168 marrow clot sections at Tokai University School of Medicine. Paraffin-embedded tissue
169 sections were stained with hematoxylin-eosin. For immunostaining, an anti-human
170 ERV3 antibody (rabbit polyclonal clone; Santa Cruz Biotechnology, CA), as a primary
171 antibody, and anti-rabbit peroxidase histofine simple stain kit (Nichirei, Tokyo), as a
172 secondary antibody, were used. The immunostaining tissue slides were observed by
173 Olympus BX 63 microscope and cellSens software.

174

175 3. Results

176 We first examined the expression level of ERV3-1 from 150 RNA-seq data of
177 blood and bone marrow of AML patients obtained from the TCGA database as
178 summarized in the Supplementary data (Table S1). All sequencing reads were mapped
179 to the human genome (GRCh38). Based on the mapped results, we measured the
180 expression levels of all genes using the human genome annotation. We then compared
181 the expression levels of 12 retroviral-like genes described in the Materials and Methods
182 section. Figure 1 shows a heatmap of the expression in 150 samples measured by
183 log-transformed TPM (transcripts per million) scores (see Materials and Methods).
184 ERV3-1 was found to exhibit higher expression compared to eleven of the other
185 retroviral-like genes. Indeed, the average and median TPM scores of ERV3-1 were 46.7
186 and 39.5, whereas those of the others were 2.4 and 0.2, respectively. We also
187 examined the expression level of ERV3-1 in the GTEx database
188 (<https://www.gtexportal.org/>), which collects various RNA-seq data from healthy people,
189 and found that the median TPM score of whole blood is 4.6, and the highest expression
190 score (27.3) was observed in adrenal gland. These results further indicate an
191 upregulated expression of ERV3-1 in blood-bone marrow of AML patients.

192 We then evaluated relationship between ERV3-1 expression and clinical data,
193 such as age, gender, cytogenetic risk, white cell count, and French-American-British
194 (FAB) classification, as summarized in the Supplementary data (Table S2). We selected
195 patients in the upper 20 and lower 20 percentiles of ERV3-1 expression (designated as
196 the ERV3-1 high and low groups, respectively). In total, 60 patients were analyzed, the
197 results for which are shown in Table 1 and the Supplementary data (Table S3). We
198 found that ERV3-1 expression was not associated with age, gender, or white blood cell
199 count, using the Mann-Whitney U test. Meanwhile, the cytogenetic risk is found to differ

200 between the ERV3-1 high and low groups ($P = 0.016$). Moreover, the expression of
 201 ERV3-1 in AML FAB M0-M3 (myeloid phenotype) was higher than that of FAB M4-M5
 202 (monocytic phenotype) ($P < 0.001$, Table 1 and Figure 2A). We then confirmed these
 203 observations using the whole 150 TCGA-LAML cases. All clinical data, excluding FAB
 204 classification, were not statistically associated with ERV3-1 expression (Figure S1).
 205 Hence, only FAB classification was statistically associated ($P < 0.001$, Figure 2B).
 206 Collectively, our transcriptome data analysis suggests that the blood and bone marrow
 207 of myeloid phenotype (FAB M0-M3) AML patients show higher expression levels of
 208 ERV3-1 than those of monocytic phenotype (FAB M4-M5).

209 To examine the protein expression of ERV3-1 in bone marrow from AML
 210 patients, we conducted an immunohistochemical analysis for 12 AML patients at the
 211 Tokai University School of Medicine in Japan. Patients' characteristics are summarized
 212 in Table 2. A previous study reported the expression of ERV3, including ERV3-1, in
 213 U-937 cells, which are one of AML cell lines classified as monocytic phenotype of AML
 214 (Larsson et al. 1996). Thus, we selected AML patients shown monocytic component
 215 classified as FAB M4-M5. In more than half of the cases (7/12), expression of ERV3-1
 216 was detected, and, in particular, ERV3-1 was expressed at some of the leukemic cell
 217 membrane (Figure 3). The results clearly suggest that ERV3-1 RNA in blood-bone
 218 marrow of AML patients was translated and expressed as protein. Moreover,
 219 considering that our transcriptome analysis revealed low expression of ERV3-1 in
 220 M4-M5 group compared to M0-M3 group in the TCGA-LAML data (Figure 2), most
 221 M0-M3 probably cases likely contain ERV3-1 protein in tumor cells as well.

222 We also evaluated association between ERV3-1 expression and chromosomal
 223 abnormalities and genetic mutations, which are considered to be involved in AML
 224 progression (Short et al. 2018). Specifically, chromosomal abnormalities, such as

225 translocation chromosomes t (15;17) and t (8;21), and trisomy of chromosome 8, are
 226 reportedly associated with AML (Vickers et al. 2000). We, therefore, focused on these
 227 anomalies in our analysis. Results show that trisomy 8 was more common in the
 228 ERV3-1 high group compared to those of the low group (Table S1, $P = 0.0232$). In our
 229 immunostaining analysis, however, the prevalence of trisomy 8 has not been a clear
 230 difference in both ERV3-1 positive (only Patient 9) and negative (Patient 10) cases
 231 (Table 2). We also evaluated three major mutations that related with AML: fms-related
 232 tyrosine kinase 3 (FLT3), isocitrate dehydrogenase 1 (IDH1), and nucleophosmin 1
 233 (NPM1); however, no significant associations were detected between these mutations
 234 and ERV3-1 expression, as shown in the Supplementary data (Table S4).

235

236

237 **4. Discussion**

238 Although many retroviral-like genes have been shown to be related to cancer
239 development (Gonzalez-Cao et al. 2016), here we specifically found that ERV3-1 shows
240 exclusively a high expression level in blood and bone marrow of all of AML patients
241 using TCGA database (Figure 1). We also confirmed that ERV3-1 protein was detected
242 in more than half of AML M4-M5 patients (7/12) (Figure 3). Although we have not
243 examined the protein expression of ERV3-1 in blood-bone marrow of AML M0-M3, we
244 found that mRNA expression level is higher in AML M0-M3 than in AML M4-M5 (Figure
245 2) suggesting that patients of AML M0-M3 may express the ERV3-1 protein as well.
246 Those results indicate that ERV3-1 protein as well as mRNAs may be expressed in
247 blood-bone marrow of most of AML patients.

248 ERV3-1 is an envelope gene of the endogenous retrovirus group 3 member 1,
249 which belongs to the HERV-R family. It is known that retroviral envelope gene is
250 involved in various biological processes, including infection and immunosuppression.
251 Indeed, ERV3-1 was reportedly expressed in placenta (Venables et al. 1995; Lin et al.
252 2000; Blaise et al. 2007) and in colorectal cancers (Lee et al. 2014). Although ERV3-1
253 lost its fusogenic activity (Blaise et al. 2007), it contains an immunosuppressive region
254 in the transmembrane domain, termed p15E, of C-type retroviruses, suggesting that
255 ERV3-1 may serve to suppress immune response (Venables et al. 1995). Indeed,
256 immunosuppressive region of another retroviral envelope-derived gene, syncytin-2,
257 supports the injection of MCA205 mouse fibrosarcoma cell line in mice (Mangeney et al.
258 2007). Therefore, immunosuppressive activity of ERV3-1 could potentially be related to
259 the progress of AML.

260 AML forms an immunosuppressive microenvironment by increasing the
261 number of myeloid-derived suppressor cells in the peripheral blood, as well as

262 regulatory T cells in both the peripheral blood and bone marrow (Beyar-Katz et al. 2018).
 263 In fact, allogeneic hematopoietic cell transplantation, one of the T-cell based
 264 immunotherapy, is the most effective in post-remission therapy, and is commonly used
 265 for AML treatment (Koreth et al. 2009). AML cell spontaneously fused with murine
 266 macrophages, endothelial, and dendritic cells, which may lead to dissemination of the
 267 disease (Martin-Padura et al. 2012). This observation suggests that
 268 immunosuppressive function of ERV3-1 might be involved in AML progression.

269 Although 150 cases show high ERV3-1 mRNA expression levels (Figure 1), we
 270 were unable to confirm the protein expression of ERV3-1 in 5 of 12 cases (Figure 3). We
 271 were also unable to identify an association of ERV3-1 expression with chromosomal
 272 abnormalities and genetic mutations (Table S1). These results might suggest that
 273 ERV3-1 is not an essential factor in AML development, but rather plays a supportive role.
 274 Therefore, the factor that affects ERV3-1 expression of AML, as well as the role of
 275 ERV3-1 in AML, should be further investigated. Moreover, considering that many
 276 viral-derived sequences have been described in eukaryote genomes that have not yet
 277 been annotated in the genome database (Nakagawa and Takahashi 2016; Pertea et al.
 278 2018; Kryukov et al. 2019) and that these viral-derived genes are dynamically altered
 279 during evolution (Imakawa et al. 2015; Imakawa and Nakagawa 2017; Pastuzyn et al.
 280 2018). Therefore, not only ERV3-1 but also other unknown viral-derived genes could be
 281 also involved in the progress of AML.

282

283

284 **Abbreviations**

285 AML: acute myelogenous leukemia
 286 ERV: endogenous retrovirus
 287 FAB: French-American-British
 288 LTR: long terminal repeat
 289 FAB: French-American-British Classification
 290 TCGA: The Cancer Genome Atlas
 291 TPM: transcripts per million

293 **Declaration of Interest**

294 None.

296 **Author's contributions**

297 S.N. and A.K. conceived the study idea. S.N. and M.K. conducted the data analysis.
 298 Y.M., K.K., R.K. and A.K. conducted experiments. M.K., K.A. and A.K. interpreted the
 299 data. S.N., M.K. and A.K. wrote the manuscript. All authors read and approved the final
 300 manuscript.

302 **Funding**

303 This study was funded by JSPS KAKENHI Grants-in-Aid for Scientific Research on
 304 Innovative Areas (16H06429, 16K21723, 17H05823, 19H04843 to SN), Challenging
 305 Exploratory Research (19K22365 to SN), Scientific Research (C) (20K06775 to SN), by
 306 Japan Agency for Medical Research and Development (AMED) of Research Program on
 307 Hepatitis (20fk0210054s0202 to AK) and Project for Cancer Research and Therapeutic
 308 Evolution (20cm0106275h0001, 20cm0106274h0001 to AK), and by research fund of
 309 Medical Research Institute, Tokai University.

311 **Acknowledgement**

312 The results shown here are in part based upon data generated by the TCGA Research
 313 Network: <https://www.cancer.gov/tcga>. Computations in this work were performed in
 314 part on the NIG supercomputer at ROIS National Institute of Genetics and SHIROKANE
 315 at Human Genome Center (the Univ. of Tokyo).

316

317

318 **References**

- 319 Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R.,
320 Sultan, C., 1976. Proposals for the classification of the acute leukaemias.
321 French-American-British (FAB) co-operative group. *Br. J. Haematol.* 33, 451-458.
- 322 Bergallo, M., Montanari, P., Mareschi, K., Merlino, C., Berger, M., Bini, I., Dapra, V.,
323 Galliano, I., Fagioli, F., 2017. Expression of the pol gene of human endogenous
324 retroviruses HERV-K and -W in leukemia patients. *Arch. Virol.* 162, 3639-3644.
- 325 Beyar-Katz, O., Gill, S., 2018. Novel approaches to acute myeloid leukemia
326 immunotherapy. *Clin. Cancer Res.* 24, 5502-5515.
- 327 Blaise, S., de Parseval, N., Bnit, L., Heidmann, T., 2003. Genomewide screening for
328 fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene
329 conserved on primate evolution. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13013-13018.
- 330 Chen, T., Meng, Z., Gan, Y., Wang, X., Xu, F., Gu, Y., Xu, X., Tang, J., Zhou, H., Zhang,
331 X., Gan, X., Van Ness, C., Xu, G., Huang, L., Zhang, X., Fang, Y., Wu, J., Zheng, S.,
332 Jin, J., Huang, W., Xu, R., 2013. The viral oncogene Np9 acts as a critical molecular
333 switch for co-activating beta-catenin, ERK, Akt and Notch1 and promoting the growth
334 of human leukemia stem/progenitor cells. *Leukemia.* 27, 1469-1478.
- 335 Cuellar, T.L., Herzner, A.M., Zhang, X., Goyal, Y., Watanabe, C., Friedman, B.A.,
336 Janakiraman, V., Durinck, S., Stinson, J., Arnott, D., Cheung, T.K., Chaudhuri, S.,
337 Modrusan, Z., Doerr, J.M., Classon, M., Haley, B., 2017. Silencing of
338 retrotransposons by SETDB1 inhibits the interferon response in acute myeloid
339 leukemia. *J. Cell Biol.* 216, 3535-3549.
- 340 Deniz, ., Ahmed, M., Todd, C.D., Rio-Machin, A., Dawson, M.A., Branco, M.R., 2020.
341 Endogenous retroviruses are a source of enhancers with oncogenic potential in acute
342 myeloid leukaemia. *Nat. Commun.* 11, 3506.
- 343 Depil, S., Roche, C., Dussart, P., Prin, L., 2002. Expression of a human endogenous
344 retrovirus, HERV-K, in the blood cells of leukemia patients. *Leukemia.* 16, 254-259.
- 345 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
346 Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner.
347 *Bioinformatics.* 29, 15-21.
- 348 Dhner, H., Estey, E.H., Amadori, S., Appelbaum, F.R., Bchner, T., Burnett, A.K.,
349 Dombret, H., Fenaux, P., Grimwade, D., Larson, R.A., Lo-Coco, F., Naoe, T.,
350 Niederwieser, D., Ossenkoppele, G.J., Sanz, M.A., Sierra, J., Tallman, M.S.,
351 Lwenberg, B., Bloomfield, C.D., 2010. Diagnosis and management of acute myeloid
352 leukemia in adults: recommendations from an international expert panel, on behalf of
353 the European LeukemiaNet. *Blood.* 115, 453-474.
- 354 Gonzalez-Cao, M., Iduma, P., Karachaliou, N., Santarpia, M., Blanco, J., Rosell, R.,

2016. Human endogenous retroviruses and cancer. *Cancer Biol. Med.* 13, 483–488.

Imakawa, K., Nakagawa, S., Miyazawa, T., 2015. Baton pass hypothesis: Successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. *Genes Cells.* 20, 771-788.

Imakawa, K., Nakagawa, S., 2017. The phylogeny of placental evolution through dynamic integrations of retrotransposons. *Prog. Mol. Biol. Transl. Sci.* 145, 89-109.

Kim, F.J., Battini, J.L., Manel, N., Sitbon, M., 2004. Emergence of vertebrate retroviruses and envelope capture. *Virology.* 318, 183-191.

Koreth, J., Schlenk, R., Kopecky, K.J., Honda, S., Sierra, J., Djulbegovic, B.J., Wadleigh, M., DeAngelo, D.J., Stone, R.M., Sakamaki, H., Appelbaum, F.R., Döhner, H., Antin, J.H., Soiffer, R.J., Cutler, C., 2009. Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: systematic review and meta-analysis of prospective clinical trials. *JAMA.* 301, 2349-2361.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., Pertea, M., 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 278.

Kryukov, K., Ueda, M.T., Imanishi, T., Nakagawa, S., 2019. Systematic survey of non-retroviral virus-like elements in eukaryotic genomes. *Virus Res.* 262, 30-36.

Lander, E.S. et al. 2001. Initial sequencing and analysis of the human genome. *Nature.* 409, 860–921.

Larsen, J.M., Christensen, I.J., Nielsen, H.J., Hansen, U., Bjerregaard, B., Talts, J.F., Larsson, L.I., 2009. Syncytin immunoreactivity in colorectal cancer: Potential prognostic impact. *Cancer Lett.* 280, 44-49.

Larsson, E., Venables, P.J., Andersson, A.C., Fan, W., Rigby, S., Botling, J., Oberg, F., Cohen, M., Nilsson, K., 1996. Expression of the endogenous retrovirus ERV3 (HERV-R) during induced monocytic differentiation in the U-937 cell line. *Int. J. Cancer.* 67, 451-456.

Lee S.H., Kang Y.J., Jo J.O., Ock, M.S., Baek, K.W., Eo, J., Lee, W.J., Choi, Y.H. Kim, W.J., Leem, S.H., Kim, H.S., Cha, H.J., 2014. Elevation of human ERV3-1 env protein expression in colorectal cancer. *J. Clin. Pathol.* 67, 840–844.

Lin, L., Xu, B., Rote, N.S., 2000. The cellular mechanism by which the human endogenous retrovirus ERV-3 env gene affects proliferation and differentiation in a human placental trophoblast model, BeWo. *Placenta.* 21, 73-78.

Mangeney, M., Renard, M., Schlecht-Louf, G., Bouallaga, I., Heidmann, O., Letzelter, C., Richaud, A., Ducos, B., Heidmann, T., 2007. Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20534-20539.

392 Martin-Padura, I., Marighetti, P., Gregato, G., Agliano, A., Malazzi, O., Mancuso, P.,
393 Pruneri, G., Viale, A., Bertolini, F., 2012. Spontaneous cell fusion of acute leukemia
394 cells and macrophages observed in cells with leukemic potential. *Neoplasia*. 14,
395 1057-1066.

396 Meyers, J., Yu, Y., Kaye, J.A., Davis, K.L., 2013. Medicare fee-for-service enrollees with
397 primary acute myeloid leukemia: an analysis of treatment patterns, survival, and
398 healthcare resource utilization and costs. *Appl. Health Econ. Health Policy*. 11,
399 275-286.

400 Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y.,
401 Edouard, P., Howes, S., Keith, J.C., McCoy, J.M., 2000. Syncytin is a captive
402 retroviral envelope protein involved in human placental morphogenesis. *Nature*. 403,
403 785-789.

404 Nakagawa, S., Takahashi, M.U., 2016. gEVE: a genome-based endogenous viral
405 element database provides comprehensive viral protein-coding sequences in
406 mammalian genomes. *Database*. 2016, baw087.

407 Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T.,
408 Suzuki-Migishima, R., Ogonuki, N., Miki, H., Kohda, T., Ogura, A., Yokoyama, M.,
409 Kaneko-Ishino, T., Ishino, F., 2006. Deletion of Peg10, an imprinted gene acquired
410 from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* 38, 101-106.

411 Pastuzyn, E.D., Day, C.E., Kearns, R.B., Kyrke-Smith, M., Taibi, A.V., McCormick, J.,
412 Yoder, N., Belnap, D.M., Erlendsson, S., Morado, D.R., Briggs, J.A.G., Feschotte, C.,
413 Shepherd, J.D., 2018. The neuronal gene Arc encodes a repurposed retrotransposon
414 Gag protein that mediates intercellular RNA transfer. *Cell*. 172, 275-288.

415 Perte, M., Shumate, A., Perte, G., Varabyou, A., Breitwieser, F.P., Chang, Y.,
416 Madugundu, A.K., Pandey, A., Salzberg, S.L., 2018. CHESS: a new human gene
417 catalog curated from thousands of large-scale RNA sequencing experiments reveals
418 extensive transcriptional noise. *Genome Biol.* 19, 208.

419 Short, N.J., Rytting, M.E., Cortes, J.E., 2018. Acute myeloid leukaemia. *Lancet*. 392,
420 593-606.

421 Tönjes, R.R., Czauderna, F., Kurth, R., 1999. Genome-wide screening, cloning,
422 chromosomal assignment, and expression of full-length human endogenous
423 retrovirus type K. *J. Virol.* 73, 9187-9195.

424 Venables, P.J., Brookes, S.M., Griffiths, D., Weiss, R.A., Boyd, M.T., 1995. Abundance
425 of an endogenous retroviral envelope protein in placental trophoblasts suggests a
426 biological function. *Virology*. 211, 589-592.

427 Vickers, M., Jackson, G., Taylor, P., 2000. The incidence of acute promyelocytic
428 leukemia appears constant over most of a human lifespan, implying only one rate

429 limiting mutation. Leukemia. 14, 722-726.
 430 Xie, T., Pan, S., Zheng, H., Luo, Z., Tembo, K.M., Jamal, M., Yu, Z., Yu, Y., Xia, J., Yin,
 431 Q., Wang, M., Yuan, W., Zhang, Q., Xiong, J., 2018. PEG10 as an oncogene:
 432 expression regulatory mechanisms and role in tumor progression. Cancer Cell Int. 18,
 433 112.
 434

435 **Figure legends**

436

437 **Figure 1. Transcriptome analysis of retroviral-like gene expression**

438 Heatmap of 12 retroviral-like genes 150 RNA-seq data is shown. Log-transformed TPM
439 scores are used to compare the mRNA expression. Red or blue indicates the high or
440 low expression levels, respectively.

441

442 **Figure 2. ERV3-1 expression in AML M0-M3 is higher than that of M4-M5**

443 (A) Case distribution of FAB M0-M3 or M4-M5 in both ERV3-1 high ($n = 30$) and low
444 groups ($n = 30$). (B) Comparison of ERV3-1 expression in both AML M0-M3 ($n = 102$)
445 and M4-M5 ($n = 44$) in all cases of TCGA data. Statistical analysis was assessed by
446 Mann-Whitney U test. Boxes denote the median, and the first and third quartile. The
447 upper and lower whiskers represent the 90th and 10th percentile, respectively. $*P <$
448 0.001.

449

450 **Figure 3. More than half of AML M4-M5 patients express ERV3-1 protein in tumor** 451 **cells**

452 Immunohistochemical staining of ERV3-1 was performed using patient bone marrow
453 samples. Tumor cells occupy the majority of bone marrow tissue. Representative of
454 ERV3-1 (A) negative and (B) positive patients (left: low power field, right: high power
455 field) corresponding to Table 2 are shown.

456

457

Figure 1. Transcriptome analysis of expression levels of retroviral-like genes

Heatmap of 11 retroviral-like genes 150 RNA-seq data is shown. Log-transformed TPM scores are used to compare the mRNA expression. Color in red or blue indicates the high or low expression levels, respectively.

Figure 2. ERV3-1 expression of AML M0-M3 is higher than that of M4-M5

(A) Case distribution of FAB M0-M3 or M4-M5 in both ERV3-1 High (n= 30) and low groups (n=30) are shown. (B) Comparison of ERV3-1 expression of both AML M0-M3 (n = 102) and M4-M5 (n = 44) in all cases of the TCGA data. Statistical analysis was assessed by Mann-Whitney U test. The boxes denote the median, and the first and third quartile. The upper and lower whiskers represent the 90% and 10%, respectively. *P < 0.001.

a) ERV3-1 negative (Patient No3)

b) ERV3-1 positive (Patient No2)

Figure 3. More than half of AML M4-M5 patients express ERV3-1 protein in tumor cells.

Immunohistochemical staining of ERV3-1 was performed using bone marrow of patient samples. Tumor cells were occupied in the majority of bone marrow tissue. Representative of ERV3-1 (A) negative and (B) positive patients (left: low power field, right: high power field) corresponding to Table 2 are shown.

Table 1. Characteristics of the TCGA-LAML patients in both ERV High (>20%) and low (<20%) groups.

Patients	Total (n=60)	ERV3 High (n=30)	ERV3 low (n=30)	<i>P</i>
Age: range (median)	21-81 (56.5)	21-81 (56.5)	31-81 (56.5)	0.706
Gender				
Male	32	17	15	0.796
Female	28	13	15	
FAB classification				
M0	7	5	2	<0.001
M1	18	11	7	
M2	15	9	6	
M3	4	4	0	
M4	10	1	9	
M5	6	0	6	
Cytogenetic risk				
Favorable	12	8	4	0.016
Intermediate	31	10	21	
Poor	15	11	4	
Unknown	2	1	1	
FAB classification				
M0-M3	44	29	15	<0.001
M4-M5	16	1	15	
Diagnostic WBC: range (median)	1-58.5 (18)	1-46 (12)	1-58.5 (28.5)	0.265
ERV3-1 expression: range (median)	2.50-7.38 (5.36)	6.13-7.38 (6.44)	2.50-4.58 (4.02)	<0.001

WBC: white blood cell count

Table 2. Characteristics of AMLM4-M5 patients in in our immunostaining analysis

Patients	Age at diagnosis	gender	FAB	Biopsy Status	Diagnostic White cell count	Cytogenetic	Cytogenetic risk	ERV3-1 expression
1	79	male	M4	Primary	2000	45, X, -Y [20/20]	Intermediate	+
2	78	male	M4	Primary	27700	46,XY [20/20]	Intermediate	++
3	60	female	M4	Primary	17900	45, XX, -7 [20/20]	poor	-
4	68	female	M4	Primary	34700	46, XX, t(7;11)(p15;p15) [20/20]	Intermediate	-
5	69	male	M4	Primary	1500	46,XY,der(7)(q31), Inv(16)(p13.1q22) [20/20]	favorable	+
6	53	male	M5	Relapse	48600	46,XY [20/20]	Intermediate	++
7	70	male	M5	Primary	33800	46,XY, inv(9)(p12q13) [10/20]	Intermediate	+
8	65	male	M5	Primary	61000	47, XY, -8, +i(8)(q10) ×2 [20/20]	Intermediate	-
9	64	male	M5	Primary	50500	49,XY,+5,add(7)(q32), +8,+mar [8/20] /50,idem,+20[3/20]	Poor	++
10	58	male	M5	Primary	17400	47,XY,+8 [18/20]	Intermediate	-
11	56	male	M5	Primary	169800	46,XY [20/20]	Intermediate	+
12	71	male	M5	Primary	5700	46,XY [20/20]	Intermediate	-