

1 **Distinct clinical and biological characteristics of acute myeloid leukemia with higher**  
2 **expression of long noncoding RNA *KIAA0125***

3 Yu-Hung Wang<sup>1,2,\*</sup>, Chien-Chin Lin<sup>1,2,3,\*</sup>, Chia-Lang Hsu<sup>4</sup>, Sheng-Yu Hung<sup>5</sup>, Chi-Yuan Yao<sup>3,5</sup>,  
4 Sze-Hwei Lee<sup>1,6</sup>, Cheng-Hong Tsai<sup>2</sup>, Hsin-An Hou<sup>2</sup>, Wen-Chien Chou<sup>2,3,@</sup>, Hwei-Fang Tien<sup>2,@</sup>

5 <sup>1</sup>Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University,  
6 Taipei, Taiwan, <sup>2</sup>Division of Hematology, Department of Internal Medicine, <sup>3</sup>Department of  
7 Laboratory Medicine, <sup>4</sup>Department of Medical Research, National Taiwan University Hospital,  
8 Taipei, Taiwan, <sup>5</sup>Division of Hematology, Department of Internal Medicine, National Taiwan  
9 University Hospital Yunlin Branch, Yunlin, Taiwan, <sup>6</sup>Tai-Cheng Stem Cell Therapy Center,  
10 National Taiwan University, Taipei, Taiwan

11 \*Yu-Hung Wang and Chien-Chin Lin contribute equally to this study.

12 @**Correspondence:** Dr. Wen-Chien Chou, Department of Laboratory Medicine, National  
13 Taiwan University Hospital, No. 7, Chung-Shan S. Rd., Taipei City 10002, Taiwan; E-mail:  
14 wchou@ntu.edu.tw; Tel: +886-2-23123456; or Dr. Hwei-Fang Tien, Department of Internal  
15 Medicine, National Taiwan University Hospital, No. 7, Chung-Shan S. Rd., Taipei City 10002,  
16 Taiwan; E-mail: hftien@ntu.edu.tw; Tel: +886-2-23123456.

17

18 **Abstract**

19 Expression of long non-coding RNA *KIAA0125* has been incorporated in various gene  
20 expression signatures for prognostic prediction in acute myeloid leukemia (AML) patients,  
21 yet its functions and clinical significance remain unclear. This study was aimed to investigate

22 the clinical and biological characteristics of AML bearing different levels of *KIAA0125*. We  
23 profiled *KIAA0125* expression levels in bone marrow cells from 347 *de novo* AML patients  
24 and found higher *KIAA0125* expression was closely associated with *FLT3*-ITD, *RUNX1*, and  
25 *DNMT3A* mutations, and M1 subtype by the French-American-British classification, but  
26 inversely correlated with t(8;21) and t(15;17). Among the 227 patients who received  
27 standard chemotherapy, those with higher *KIAA0125* expression had a lower complete  
28 remission rate, shorter overall survival (OS) and disease-free survival (DFS) than those with  
29 lower expression. The prognostic significance was validated in both TCGA and GSE12417  
30 cohorts. Subgroup analyses showed that higher *KIAA0125* expression also predicted shorter  
31 DFS and OS in patients with normal karyotype or non-M3 AML. In multivariable analysis,  
32 higher *KIAA0125* expression remained an adverse risk factor independent of age, WBC  
33 counts, karyotypes, and mutation patterns. Bioinformatics analyses revealed that higher  
34 *KIAA0125* expression was associated with hematopoietic and leukemic stem cell signatures  
35 and ATP-binding cassette transporters, two predisposing factors for chemoresistance.  
36

## 37 **Introduction**

38 Long non-coding RNAs (lncRNAs) are non-protein coding RNAs that are longer than 200  
39 nucleotides. Comparing to other classes of ncRNAs, lncRNAs exhibit a wide range of  
40 structures and functions[1]. Recently, lncRNAs have emerged as important regulators for  
41 gene expression via remodeling nuclear architecture, modulating mRNA stability and  
42 translation, and post-translational modifications[1-4]. Besides, some lncRNAs are  
43 dysregulated and harbor prognostic relevance in several types of cancers[5-8]. However, the  
44 roles of lncRNAs in tumorigenesis are still largely unknown.

45 In recent years, several gene expression-based prognostic scores have been developed for  
46 better risk stratification of acute myeloid leukemia (AML) patients[9-14]. Among those  
47 high-risk genes, lncRNA gene *KIAA0125* (also named as *FAM30A*) is unique because it is the  
48 only non-coding gene and is expressed in humans but not in mice (From the UniProt  
49 database, <https://www.uniprot.org/uniprot/Q9NZY2>). Additionally, *KIAA0125* expression  
50 was integrated into a recently proposed 17-gene stemness score, which could predict  
51 outcomes in AML patients[9].

52 In this study, we aimed to investigate the association of *KIAA0125* expression with clinical  
53 and biological characteristics in AML patients. We first profiled the expression levels of  
54 *KIAA0125* in bone marrow (BM) cells from AML patients and normal controls and  
55 demonstrated that AML patients had higher *KIAA0125* expression than normal controls.

56 Higher expression of *KIAA0125* was associated with distinct clinical and biological  
57 characteristics and served as an independent poor prognostic biomarker for AML patients in

58 ours and two other publicly annotated cohorts. Further bioinformatics analyses showed that  
59 higher expression of *KIAA0125* in AML was closely associated with hematopoietic stem cell  
60 (HSC) and leukemic stem cell (LSC) signatures and several important ATP-binding cassette  
61 transporters (ABC transporters); these factors are regarded responsible for chemoresistance  
62 in AML. *KIAA0125* could be a potential target for novel treatment in AML patients with high  
63 *KIAA0125* expression.

64

## 65 **Materials and Methods**

### 66 **Patients**

67 We recruited 347 adult patients with *de novo* AML diagnosed in the National Taiwan  
68 University Hospital (NTUH) from 1996 to 2011 who had enough cryopreserved BM cells for  
69 tests. The diagnoses were based on the French–American–British (FAB) and the 2016 World  
70 Health Organization classifications[15, 16]. Among them, 227 patients received standard  
71 chemotherapy. Non M3 (acute promyelocytic leukemia, APL) patients received idarubicin 12  
72 mg/m<sup>2</sup> per day days 1-3 and cytarabine 100 mg/m<sup>2</sup> per day days 1-7, and then consolidation  
73 chemotherapy with 2-4 courses of high-dose cytarabine 2000 mg/m<sup>2</sup> q12h for total 8 doses,  
74 with or without an anthracycline (Idarubicin or Mitoxantrone), after achieving complete  
75 remission (CR) as described previously[17]. APL patients received concurrent all-trans  
76 retinoic acid and chemotherapy. The remaining 120 patients received supportive care and/or  
77 reduced-intensity anti-leukemia therapy due to underlying comorbidities or based on the  
78 decision of the physicians or patients. BM samples from 30 healthy donors of hematopoietic

79 stem cell transplantation (HSCT) were collected as normal controls. This study was approved  
80 by the Institutional Review Board of the NTUH.

### 81 **Microarray and genetic alteration analysis**

82 We profiled the global gene expression of BM mononuclear cells from 347 AML patients and  
83 30 healthy transplant donors by Affymetrix GeneChip Human Transcriptome Array 2.0 as  
84 described previously[11, 18, 19]. The raw and normalized microarray data reported in this  
85 article have been deposited in the Gene Expression Omnibus database (accession number  
86 GSE68469 and GSE71014)[11, 18, 19]. For external validation, we analyzed microarray  
87 datasets of GSE12662 (n=76) and GSE12417-GPL96 cohorts (n=163), and RNAseq dataset of  
88 the TCGA cohort (n=186)[20-22]. Cytogenetic analyses were performed and interpreted as  
89 described previously[23]. We also analyzed the mutation statuses of 17 myeloid-relevant  
90 genes including *ASXL1*, *IDH1*, *IDH2*, *TET2*, *DNMT3A*, *FLT3-ITD*, *FLT3-TKD*, *KIT*, *NRAS*, *KRAS*,  
91 *RUNX1*, *MLL/PTD*, *CEBPA*, *NPM1*, *PTPN11*, *TP53*, and *WT1* by Sanger sequencing as  
92 previously described[17, 18, 23-26].

### 93 **Analysis of gene expression in next-generation sequencing datasets**

94 We analyzed gene expression data of 141 AML samples profiled with Illumina Genome  
95 Analyzer RNA Sequencing in the TCGA database[22] to investigate the absolute gene  
96 expression levels.

### 97 **Gene Set Enrichment Analysis**

98 The preranked Gene Set Enrichment Analysis (GSEA) implemented by R package  
99 clusterProfiler was performed using the stem cell-related gene sets from the MSigDB

100 databases. The genes were ranked based on the Spearman's correlation coefficient between  
101 the given gene and *KIAA0125*.

## 102 **Statistical analysis**

103 We used the Mann-Whitney U test to compare continuous variables and medians of  
104 distributions. The Fisher exact test or the  $\chi^2$  test were performed to examine the difference  
105 in discrete variables, including gender, FAB classification, cytogenetic changes, and genetic  
106 alterations between patients with lower and higher *KIAA0125* expression. Overall survival  
107 (OS) was the duration from the date of initial diagnosis to the time of last follow-up or death  
108 from any cause, whichever occurred first. Disease free survival (DFS) was the duration from  
109 the date of attaining leukemia-free state until the date of AML relapse or death from any  
110 cause, whichever occurred first. The survival prediction power of *KIAA0125* expression was  
111 evaluated by both the log-rank test and the univariate Cox proportional hazards model. We  
112 plotted the survival curves with Kaplan-Meier analysis and calculated the statistical  
113 significance with the log-rank test. The Cox proportional hazards model was used in  
114 multivariable regression analysis. P values <0.05 were considered statistically significant. All  
115 statistical analyses were performed with BRB-ArrayTools (version 4.5.1; Biometric Research  
116 Branch, National Cancer Institute, Rockville, MD), and IBM SPSS Statistics 23 for Windows.

117

## 118 **Results**

119 The median age of the 347 AML patients was 57 years. Among the 331 patients who had  
120 cytogenetic data at diagnosis, 165 (49.8%) had clonal chromosomal abnormalities. Sixty

121 patients (18.1%) had favorable cytogenetics; 223 (67.2%), intermediate-risk cytogenetics;  
122 and 14.8%, unfavorable cytogenetics (Supplement Table 1) based on the refined British  
123 Medical Research Council (MRC) classification[27]. The clinical and laboratory characteristics  
124 of these patients at diagnosis are summarized in Table 1.

### 125 **Comparison of clinical characteristics and genetic alterations between patients with higher** 126 **and lower *KIAA0125* expression**

127 We first compared the BM *KIAA0125* expression between the 30 healthy controls and 347  
128 AML patients. The expression of *KIAA0125* was significantly higher in AML samples than  
129 healthy controls ( $p < 0.001$ , Figure 1a). Then, the 347 AML patients were divided into two  
130 groups by the median value of the *KIAA0125* expression. The comparison of clinical and  
131 laboratory features between the two groups is shown in Table 1. The higher-*KIAA0125* group  
132 had higher circulating blasts at diagnosis ( $p = 0.021$ ) and higher incidence of FAB M1 subtype  
133 ( $p < 0.001$ ), but lower incidence of M3 subtype ( $p < 0.001$ ), compared to the  
134 lower-*KIAA0125* group (Table 1). Patients with FAB immature subtype of M0 or M1 had  
135 higher *KIAA0125* expression while those with M3 subtype had lower expression of *KIAA0125*  
136 ( $p < 0.001$ , Figure 1c), and similar results were noted in the GSE12662 cohort ( $p < 0.001$ , Figure  
137 1d). Moreover, higher-*KIAA0125* patients had significantly lower frequencies of t(8;21) and  
138 t(15;17) in both NTUH cohort (both  $p < 0.001$ , Supplement Table 1) and TCGA cohort ( $p = 0.006$   
139 and  $p < 0.001$ , respectively, Supplement Table 1). The higher-*KIAA0125* patients more  
140 frequently had *FLT3*-ITD ( $p = 0.048$ ) and mutations in *DNMT3A* ( $p = 0.015$ ),  
141 and *RUNX1* ( $p = 0.034$ ) (Supplement Table 2). Compatible with this finding, patients with

142 *DNMT3A* or *RUNX1* mutation had higher *KIAA0125* expression than those without the  
143 mutation ( $p=0.019$  and  $0.045$ , respectively, Supplement Figure 1). Similarly, there was close  
144 association between higher *KIAA0125* expression and *DNMT3A* ( $p=0.001$ )  
145 and *RUNX1* mutations ( $p=0.017$ ) in the TCGA cohort (Supplement Table 3). Among the 227  
146 patients who received standard chemotherapy, 166 (73.1%) patients attained a complete  
147 remission (CR) while 43 (18.9%) patients had primary refractory diseases. Notably, the  
148 patients with higher *KIAA0125* expression had a lower CR rate (61.2% vs. 84.7%,  $p<0.001$ )  
149 than those with lower expression. In accordance with this finding, the patients who achieved  
150 CR after induction chemotherapy had lower expression of BM *KIAA0125* at diagnosis than  
151 those who did not ( $p<0.001$ , Figure 1b).

#### 152 **The impacts of the *KIAA0125* expression on OS and DFS**

153 Patients with higher *KIAA0125* expression had an inferior DFS and OS than those with lower  
154 expression, no matter whether the survival was censored on the day of hematopoietic stem  
155 cell transplantation (HSCT) (median, 11.7 months vs. 101.7 months,  $p<0.001$ ; and 20 months  
156 vs not reached (NR),  $p=0.001$ , respectively; Figure 2a and 2b) or not ( $p=0.001$  and  $p=0.001$ ,  
157 respectively; Supplement Figure 2a and 2b). Subgroup analyses showed that the prognostic  
158 significance of *KIAA0125* expression for DFS and OS remained valid in both non-APL and  
159 normal karyotype patients (Figure 2c and 2d).

160 In multivariable analysis, we included clinically relevant parameters and variables with  
161 a  $p$ -value  $<0.05$  in univariate Cox regression analysis (Supplement Table 4) as covariates,  
162 including age, white blood cell counts at diagnosis, karyotypes, mutation statuses



163 of *NPM1/FLT3-ITD*, *CEBPA*<sup>double mutations</sup>, *RUNX1*, *MLL-PTD*, and *TP53*, and *KIAA0125*  
164 expression. Higher *KIAA0125* expression, either divided by a median (Table 2) or calculated  
165 as continuous values (Supplement Table 5), was an independent adverse prognostic factor  
166 for DFS ( $p < 0.001$  and  $p < 0.001$ , respectively) and OS ( $p = 0.002$  and  $p = 0.001$ , respectively). To  
167 verify the prognostication power of the *KIAA0125* expression, we analyzed the expression  
168 of *KIAA0125* and its prognostic significance in the TCGA cohort and the GSE12417-GPL96  
169 cohort. Consistent with the findings in the NTUH cohort, patients with higher *KIAA0125*  
170 expressions had a significantly shorter OS (12.2 months vs. 27.4 months,  $p = 0.004$ , and 7.8  
171 months vs. 33.3 months,  $p < 0.001$ , respectively, Figure 2e and 2f) than those with lower  
172 *KIAA0125* expression in the two external validation cohorts.

### 173 **Biological impacts of *KIAA0125* in AML**

174 To gain biological insights into the underlying mechanism of unfavorable prognosis related  
175 to *KIAA0125* overexpression, we investigated the genes whose expression are strongly  
176 correlated with that of *KIAA0125*. Since *KIAA0125* was reported as an LSC marker, we  
177 curated several published HSC and LSC signatures from different studies[28-30]. GSEA  
178 showed HSC and LSC signatures were all significantly enriched in the patients with higher  
179 *KIAA0125* expression in both the NTUH and TCGA cohorts (both  $p < 0.001$ , Figure 3a). We  
180 next checked the leading-edge genes whose expression levels were most positively  
181 correlated to *KIAA0125* expression in both NTUH and TCGA cohort. Among them, *SPINK2*,  
182 *MAP7*, *HOPX*, *MMRN1*, *DNMT3B*, *TCF4*, *SLC38A1*, *DOCK1*, *ARHGAP22*, *MN1*, and 4 genes in  
183 the ATP-binding cassette (ABC) superfamily (*ABCG1*, *ABCA2*, *ABCB1*, and *ABCC1*) have been

184 reported to be associated with poor prognosis or chemoresistance in AML (Figure 3b and  
185 Table 3)[9, 31-50].

186

## 187 **Discussion**

188 AML cells have abnormal genetic background, either mutations or aberrant expression of  
189 specific genes. In recent years, several gene expression scores have been proposed for  
190 prognostic prediction of AML patients. We previously developed a 11-gene mRNA  
191 expression signature, including *AIF1L*, *CXCR7*, *DNTT*, *GPR56*, *H1FO*, *IFITM3*, *KIAA0125*, *MX1*,  
192 *STAB1*, *TM4SF1* and *TNS3*, for prognostication in AML patients[11]. Another group built a  
193 six-gene leukemia stem cell (LSC) score with the incorporation of *DNMT3B*, *GPR56*, *CD34*,  
194 *SOCS2*, *SPINK2*, and *KIAA0125* expressions for pediatric AML[32]. Recently, Ng et al.  
195 proposed a 17-gene LSC score that incorporated expressions of 17 stemness-related genes,  
196 including *KIAA0125*, and showed the scoring system was powerful to predict prognosis in  
197 AML patients[9]. Among these prognostic-relevant genes, *KIAA0125* is the only non-coding  
198 gene and expressed only in the homo sapiens, but not in mice.

199 *KIAA0125* is located on chromosome 14 of the human genome. It was reported to be  
200 upregulated in ameloblastoma but shown as a tumor suppressor gene in colorectal  
201 cancer[51, 52]. However, its role in tumorigenesis is still largely unknown. In this study, we  
202 found that the expression level of *KIAA0125* in BM was significantly higher in AML patients  
203 than normal HSC transplant donors. The expression of *KIAA0125* was lower in the mature  
204 subtype M3, but higher in more immature subtypes, M0 and M1, compatible with the

205 finding that it is an LSC-related gene[20]. The similar finding could be seen in the GSE 12662  
206 cohort. Further bioinformatics study also showed highly significant association of *KIAA0125*  
207 expression with stem cell signatures, either HSC or LSC. We found that expressions of *SPINK2*,  
208 *MAP7*, *HOPX*, *MMRN1*, *DNMT3B*, *TCF4*, *SLC38A1*, *DOCK1*, *ARHGAP22*, *MN1*, and 4 genes in  
209 the ATP-binding cassette (ABC) superfamily (*ABCG1*, *ABCA2*, *ABCB1*, and *ABCC1*), which have  
210 been reported to be associated with poor prognosis or chemoresistance in AML, were  
211 positively correlated to higher expression of *KIAA0125* (Figure 3b and Table 3). *HOPX*, *DOCK1*,  
212 *DNMT3B*, *MMRN1*, and *ARHGAP22* genes were reported as important leukemia stem cell  
213 markers[9, 34, 35, 37, 42]. Higher *SPINK2* expression was associated with poor prognosis in  
214 adult and pediatric AML[31, 32]. *TCF4* expression could predict outcome in *RUNX1*-mutated  
215 and translocated AML[39, 40]. *MN1* overexpression could induce AML in mice and predict  
216 ATRA resistance in human AML patients[43, 44]. Current knowledge about the association  
217 between these *KIAA0125*-correlated genes and AML is summarized in Table 3.  
218 Interestingly, the expression levels of several ABC transporter genes, including *ABCA2*,  
219 *ABCB1*, *ABCC1*, and *ABCG1*, were also significantly higher in AML patients with higher  
220 *KIAA0125* expression. The ABC transporter family consists of 48 proteins in subfamilies  
221 designated A to G and some of them are known to be associated with multidrug resistance  
222 via ATP-dependent drug efflux[45, 46, 49]. *ABCB1*, *ABCC1* and *ABCG1* were reported to be  
223 responsible for chemoresistance in AML[45, 48]. The translational expression of *ABCA2* was  
224 shown to be a prognostic marker for drug resistance in pediatric acute lymphoblastic

225 leukemia[47, 50]. The underlying mechanistic basis of the high correlation of these 4 genes  
226 to the expression of *KIAA0125* warrants further studies.

227 To the best of our knowledge, this is by far the first study specifically addressing the  
228 expression of lncRNA *KIAA0125* and its clinical and biological associations in AML patients.

229 We found that higher *KIAA0125* expression were closely associated with *RUNX1* and  
230 *DNMT3A1* mutations in both the NTUH and TCGA cohorts. Patients with higher *KIAA0125*  
231 expression were more refractory to chemotherapy with a lower CR rate and higher  
232 refractory rate (Table 1). They had shorter OS and DFS not only among total cohort, but also  
233 in subgroups of patients with non-APL and those with normal karyotype. Based on its  
234 important clinical significance, *KIAA0125* could be a potential therapeutic target for AML  
235 and might give some directives to future work. Further experimental studies are necessary  
236 to delineate how *KIAA0125* participates in the stem cell biology of hematopoietic lineages  
237 and its role in the pathogenesis in AML.

238

239 **Acknowledgements**

240 We would like to acknowledge the service provided by Department of Laboratory Medicine,  
241 Department of Medical Research, and Division of Hematology, Department of Internal  
242 Medicine, National Taiwan University Hospital.

243 **Authorship Contributions**

244 YHW and CCL contribute equally to this study. YHW and CCL were responsible for data  
245 collection and management, statistical analysis and interpretation, literature research, and  
246 manuscript writing; SYH and CYC were responsible for data management and statistical  
247 analysis; CLH assisted in statistical analysis; SHL, CHT and HAH was responsible for data  
248 collection and management; and WCC and HFT planned, designed, and coordinated the  
249 study over the entire period and wrote the manuscript.

250 **Disclosure of Conflicts of Interest**

251 The authors declare that they have no competing interests.

252 **Funding**

253 The study was supported by grants from Ministry of Health and Welfare, Taiwan (project  
254 number: MOHW107-TDU-B-211-114009) and Ministry of Science and Technology, Taiwan  
255 (project number: MOST 107-2314-B-002-013 and MOST 108-2314-B-002-011).

256

## 257 Reference

- 258 1. Fatica, A. and I. Bozzoni, *Long non-coding RNAs: new players in cell differentiation*  
259 *and development*. Nat Rev Genet, 2014. **15**(1): p. 7-21.
- 260 2. Wilusz, J.E., H. Sunwoo, and D.L. Spector, *Long noncoding RNAs: functional surprises*  
261 *from the RNA world*. Genes Dev, 2009. **23**(13): p. 1494-504.
- 262 3. Kung, J.T., D. Colognori, and J.T. Lee, *Long noncoding RNAs: past, present, and future*.  
263 *Genetics*, 2013. **193**(3): p. 651-69.
- 264 4. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into*  
265 *functions*. Nat Rev Genet, 2009. **10**(3): p. 155-9.
- 266 5. Lin, C. and L. Yang, *Long Noncoding RNA in Cancer: Wiring Signaling Circuitry*. Trends  
267 *Cell Biol*, 2018. **28**(4): p. 287-301.
- 268 6. Prensner, J.R. and A.M. Chinnaiyan, *The emergence of lncRNAs in cancer biology*.  
269 *Cancer Discov*, 2011. **1**(5): p. 391-407.
- 270 7. Huarte, M., *The emerging role of lncRNAs in cancer*. Nat Med, 2015. **21**(11): p.  
271 1253-61.
- 272 8. Cheetham, S.W., et al., *Long noncoding RNAs and the genetics of cancer*. Br J Cancer,  
273 2013. **108**(12): p. 2419-25.
- 274 9. Ng, S.W., et al., *A 17-gene stemness score for rapid determination of risk in acute*  
275 *leukaemia*. Nature, 2016. **540**(7633): p. 433-437.
- 276 10. Metzeler, K.H., et al., *An 86-probe-set gene-expression signature predicts survival in*  
277 *cytogenetically normal acute myeloid leukemia*. Blood, 2008. **112**(10): p. 4193-201.
- 278 11. Chuang, M.K., et al., *An mRNA expression signature for prognostication in de novo*  
279 *acute myeloid leukemia patients with normal karyotype*. Oncotarget, 2015. **6**(36): p.  
280 39098-110.
- 281 12. Wilop, S., et al., *A three-gene expression-based risk score can refine the European*  
282 *LeukemiaNet AML classification*. J Hematol Oncol, 2016. **9**(1): p. 78.
- 283 13. Marcucci, G., et al., *Epigenetics meets genetics in acute myeloid leukemia: clinical*  
284 *impact of a novel seven-gene score*. J Clin Oncol, 2014. **32**(6): p. 548-56.
- 285 14. Li, Z., et al., *Identification of a 24-gene prognostic signature that improves the*  
286 *European LeukemiaNet risk classification of acute myeloid leukemia: an international*  
287 *collaborative study*. J Clin Oncol, 2013. **31**(9): p. 1172-81.
- 288 15. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification*  
289 *of myeloid neoplasms and acute leukemia*. Blood, 2016. **127**(20): p. 2391-405.
- 290 16. Bennett, J.M., et al., *Proposals for the classification of the acute leukaemias.*  
291 *French-American-British (FAB) co-operative group*. Br J Haematol, 1976. **33**(4): p.  
292 451-8.
- 293 17. Tang, J.L., et al., *AML1/RUNX1 mutations in 470 adult patients with de novo acute*  
294 *myeloid leukemia: prognostic implication and interaction with other gene alterations*.  
295 *Blood*, 2009. **114**(26): p. 5352-61.

- 296 18. Chiu, Y.C., et al., *Prognostic significance of NPM1 mutation-modulated*  
297 *microRNA-mRNA regulation in acute myeloid leukemia*. *Leukemia*, 2016. **30**(2): p.  
298 274-84.
- 299 19. Chuang, M.K., et al., *A 3-microRNA scoring system for prognostication in de novo*  
300 *acute myeloid leukemia patients*. *Leukemia*, 2015. **29**(5): p. 1051-9.
- 301 20. Payton, J.E., et al., *High throughput digital quantification of mRNA abundance in*  
302 *primary human acute myeloid leukemia samples*. *J Clin Invest*, 2009. **119**(6): p.  
303 1714-26.
- 304 21. Metzeler, K.H., et al., *An 86-probe-set gene-expression signature predicts survival in*  
305 *cytogenetically normal acute myeloid leukemia*. *Blood*, 2008. **112**(10): p. 4193-201.
- 306 22. Ley, T.J., et al., *Genomic and epigenomic landscapes of adult de novo acute myeloid*  
307 *leukemia*. *N Engl J Med*, 2013. **368**(22): p. 2059-74.
- 308 23. Chou, W.C., et al., *TET2 mutation is an unfavorable prognostic factor in acute myeloid*  
309 *leukemia patients with intermediate-risk cytogenetics*. *Blood*, 2011. **118**(14): p.  
310 3803-10.
- 311 24. Chou, W.C., et al., *Distinct clinical and biologic characteristics in adult acute myeloid*  
312 *leukemia bearing the isocitrate dehydrogenase 1 mutation*. *Blood*, 2010. **115**(14): p.  
313 2749-54.
- 314 25. Chou, W.C., et al., *Distinct clinical and biological features of de novo acute myeloid*  
315 *leukemia with additional sex comb-like 1 (ASXL1) mutations*. *Blood*, 2010. **116**(20): p.  
316 4086-94.
- 317 26. Hou, H.A., et al., *Integration of cytogenetic and molecular alterations in risk*  
318 *stratification of 318 patients with de novo non-M3 acute myeloid leukemia*. *Leukemia*,  
319 2014. **28**(1): p. 50-8.
- 320 27. Grimwade, D., et al., *Refinement of cytogenetic classification in acute myeloid*  
321 *leukemia: determination of prognostic significance of rare recurring chromosomal*  
322 *abnormalities among 5876 younger adult patients treated in the United Kingdom*  
323 *Medical Research Council trials*. *Blood*, 2010. **116**(3): p. 354-65.
- 324 28. Eppert, K., et al., *Stem cell gene expression programs influence clinical outcome in*  
325 *human leukemia*. *Nat Med*, 2011. **17**(9): p. 1086-93.
- 326 29. Ivanova, N.B., et al., *A stem cell molecular signature*. *Science*, 2002. **298**(5593): p.  
327 601-4.
- 328 30. Jaatinen, T., et al., *Global gene expression profile of human cord blood-derived*  
329 *CD133+ cells*. *Stem Cells*, 2006. **24**(3): p. 631-41.
- 330 31. Xue, C., et al., *Elevated SPINK2 gene expression is a predictor of poor prognosis in*  
331 *acute myeloid leukemia*. *Oncol Lett*, 2019. **18**(3): p. 2877-2884.
- 332 32. Elsayed, A.H., et al., *A six-gene leukemic stem cell score identifies high risk pediatric*  
333 *acute myeloid leukemia*. *Leukemia*, 2020. **34**(3): p. 735-745.
- 334 33. Fu, L., et al., *High expression of MAP7 predicts adverse prognosis in young patients*  
335 *with cytogenetically normal acute myeloid leukemia*. *Sci Rep*, 2016. **6**: p. 34546.

- 336 34. Lin, C.C., et al., *Higher HOPX expression is associated with distinct clinical and*  
337 *biological features and predicts poor prognosis in de novo acute myeloid leukemia.*  
338 *Haematologica*, 2017. **102**(6): p. 1044-1053.
- 339 35. Laszlo, G.S., et al., *Multimerin-1 (MMRN1) as Novel Adverse Marker in Pediatric Acute*  
340 *Myeloid Leukemia: A Report from the Children's Oncology Group.* *Clin Cancer Res*,  
341 2015. **21**(14): p. 3187-95.
- 342 36. Hayette, S., et al., *High DNA methyltransferase DNMT3B levels: a poor prognostic*  
343 *marker in acute myeloid leukemia.* *PLoS One*, 2012. **7**(12): p. e51527.
- 344 37. Niederwieser, C., et al., *Prognostic and biologic significance of DNMT3B expression in*  
345 *older patients with cytogenetically normal primary acute myeloid leukemia.* *Leukemia*,  
346 2015. **29**(3): p. 567-75.
- 347 38. Wong, K.K., C.H. Lawrie, and T.M. Green, *Oncogenic Roles and Inhibitors of DNMT1,*  
348 *DNMT3A, and DNMT3B in Acute Myeloid Leukaemia.* *Biomark Insights*, 2019. **14**: p.  
349 1177271919846454.
- 350 39. Gerritsen, M., et al., *RUNX1 mutations enhance self-renewal and block granulocytic*  
351 *differentiation in human in vitro models and primary AMLs.* *Blood Adv*, 2019. **3**(3): p.  
352 320-332.
- 353 40. In 't Hout, F.E.M., et al., *Transcription factor 4 (TCF4) expression predicts clinical*  
354 *outcome in RUNX1 mutated and translocated acute myeloid leukemia.*  
355 *Haematologica*, 2019.
- 356 41. Li, Y., et al., *High expression of SLC38A1 predicts poor prognosis in patients with de*  
357 *novo acute myeloid leukemia.* *J Cell Physiol*, 2019. **234**(11): p. 20322-20328.
- 358 42. Lee, S.H., et al., *High expression of dedicator of cytokinesis 1 (DOCK1) confers poor*  
359 *prognosis in acute myeloid leukemia.* *Oncotarget*, 2017. **8**(42): p. 72250-72259.
- 360 43. Heuser, M., et al., *MN1 overexpression induces acute myeloid leukemia in mice and*  
361 *predicts ATRA resistance in patients with AML.* *Blood*, 2007. **110**(5): p. 1639-47.
- 362 44. Heuser, M., et al., *Cell of origin in AML: susceptibility to MN1-induced transformation*  
363 *is regulated by the MEIS1/AbdB-like HOX protein complex.* *Cancer Cell*, 2011. **20**(1): p.  
364 39-52.
- 365 45. Choi, Y.H. and A.M. Yu, *ABC transporters in multidrug resistance and*  
366 *pharmacokinetics, and strategies for drug development.* *Curr Pharm Des*, 2014. **20**(5):  
367 p. 793-807.
- 368 46. Dean, M., *The genetics of ATP-binding cassette transporters.* *Methods Enzymol*, 2005.  
369 **400**: p. 409-29.
- 370 47. Efferth, T., et al., *Expression profiling of ATP-binding cassette transporters in*  
371 *childhood T-cell acute lymphoblastic leukemia.* *Mol Cancer Ther*, 2006. **5**(8): p.  
372 1986-94.
- 373 48. Marzac, C., et al., *ATP Binding Cassette transporters associated with chemoresistance:*  
374 *transcriptional profiling in extreme cohorts and their prognostic impact in a cohort of*  
375 *281 acute myeloid leukemia patients.* *Haematologica*, 2011. **96**(9): p. 1293-301.



- 376 49. Vasiliou, V., K. Vasiliou, and D.W. Nebert, *Human ATP-binding cassette (ABC)*  
377 *transporter family*. Hum Genomics, 2009. **3**(3): p. 281-90.
- 378 50. Aberuyi, N., et al., *The translational expression of ABCA2 and ABCA3 is a strong*  
379 *prognostic biomarker for multidrug resistance in pediatric acute lymphoblastic*  
380 *leukemia*. Onco Targets Ther, 2017. **10**: p. 3373-3380.
- 381 51. Diniz, M.G., et al., *The long noncoding RNA KIAA0125 is upregulated in*  
382 *ameloblastomas*. Pathol Res Pract, 2019. **215**(3): p. 466-469.
- 383 52. Yang, Y., et al., *lncRNA KIAA0125 functions as a tumor suppressor modulating growth*  
384 *and metastasis of colorectal cancer via Wnt/beta-catenin pathway*. Cell Biol Int, 2019.
- 385
- 386

387 **Table 1. Comparison of clinical and laboratory features between AML patients with lower**  
 388 **and higher BM *KIAA0125* expression**

Clinical characters	Total (N=347)	High <i>KIAA0125</i> (n=174)	Low <i>KIAA0125</i> (n=173)	P value
<b>Sex</b>				0.174
<b>Male</b>	196	92	104	
<b>Female</b>	151	82	69	
<b>Age*</b>		57 (15-91)	58 (18-90)	0.830
<b>Laboratory data*</b>				
<b>WBC, X 10<sup>9</sup> /L</b>	21.9 (0.38-423)	21.4 (0.38-417.5)	22.38 (0.65-423.0)	0.872
<b>Hb, g/dL</b>	8.1 (3.3-16.2)	8.1 (3.3-13.2)	8.1 (3.7-16.2)	0.959
<b>Platelet, X 10<sup>9</sup> /L</b>	45 (2-655)	54 (6-455)	41 (2-655)	0.060
<b>Blast, X 10<sup>9</sup> /L</b>	9.1 (0-369.1)	12.3 (0-345.9)	5.7 (0-369.1)	<b>0.021</b>
<b>LDH (U/L)</b>	917 (202-13130)	892.5 (242-7734)	925 (202-13130)	0.787
<b>FAB classification, n (%)</b>				
<b>M0</b>	6	4 (2.4)	2 (1.2)	0.414
<b>M1</b>	67	48 (28.4)	19 (11.1)	<b>&lt;0.001</b>
<b>M2</b>	109	56 (33.1)	53 (30.8)	0.756
<b>M3</b>	28	3 (1.8)	25 (14.5)	<b>&lt;0.001</b>
<b>M4</b>	103	49 (28.9)	54 (31.4)	0.534
<b>M5</b>	20	6 (3.6)	14 (8.1)	0.063
<b>M6</b>	8	3 (1.8)	5 (2.9)	0.469
<b>Undetermined</b>	6	5	1	0.215
<b>Induction response, n (%)</b>	227	116	111	
<b>CR</b>	165 (72.7)	71 (61.2)	94 (84.7)	<b>&lt;0.001</b>
<b>PR</b>	5 (2.2)	4 (3.4)	1 (0.9)	0.191
<b>Refractory</b>	42 (18.5)	33 (28.4)	9 (8.1)	<b>&lt;0.001</b>
<b>Induction death</b>	15 (6.6)	8 (6.9)	7 (6.3)	0.858
<b>Relapse (%)</b>	72 (31.7)	42 (36.2)	30 (27.0)	0.137

389 Abbreviations: CR, complete remission; Hb, hemoglobin; HSCT, allogeneic hematopoietic stem cell transplantation; LDH,

390 lactate dehydrogenase; PR, partial remission.

391 \*Median (range).

392

393 **Table 2. Multivariable analysis for DFS and OS in 227 AML patients who received standard**  
 394 **intensive chemotherapy**

Variable	DFS				OS			
	HR	95% CI		P	HR	95% CI		P
		Lower	Upper			Lower	Upper	
Age*	1.009	0.997	1.021	0.129	1.031	1.014	1.047	<0.001
WBC*	1.004	1.001	1.007	0.002	1.005	1.001	1.008	0.009
Karyotype†	1.762	1.310	2.369	<0.001	1.910	1.288	2.833	0.001
<i>NPM1/FLT3-ITD</i> ‡	0.540	0.297	0.981	0.043	0.812	0.404	1.634	0.559
<i>CEBPA</i> <sup>double</sup>	0.423	0.216	0.830	0.012	0.253	0.077	0.833	0.024
<i>RUNX1</i>	1.463	0.841	2.546	0.178	1.492	0.763	2.914	0.242
<i>MLL-PTD</i>	2.864	1.346	6.094	0.006	3.017	1.136	8.017	0.027
<i>TP53</i>	1.691	0.607	4.705	0.315	2.798	0.875	8.950	0.083
Higher <i>KIAA0125</i> expression§	2.609	1.760	3.867	<0.001	2.226	1.335	3.711	0.002

395 Abbreviations: HR, hazard ratios; CI, confidence interval.

396 \*As continuous variable.

397 †Unfavorable cytogenetics versus others. The classification of favorable, intermediate and unfavorable

398 cytogenetics is based on the refined Medical Research Council (MRC) classification [27]. Favorable:

399 t(15;17)(q22;q21), t(8;21)(q22;q22), and inv(16)(p13q22)/t(16;16)(p13;q22); unfavorable: abn(3q) (excluding

400 t(3;5)(q25;q34)), inv(3)(q21q26)/t(3;3)(q21;q26), add(5q)/del(5q), -5, -7, add(7q)/del(7q), t(6;11)(q27;q23),

401 t(10;11)(p1113;q23), other t(11q23) (excluding t(9;11)(p21~22;q23) and t(11;19)(q23;p13)), t(9;22)(q34;q11),

402 -17, and abn(17p); and intermediate: entities not classified as favorable or adverse. Seven patients without

403 chromosome data were not included in the analysis.

404 ‡*NPM1*+/*FLT3*-ITD- versus other subtypes.

405 §High vs. low expression of *KIAA0125* (median as cutoff)

406

407 **Table 3. Summary of the biological functions of the *KIAA0125*-associated genes that have**  
 408 **been reported to be associated with prognosis or drug resistance in AML patients and**  
 409 **their correlation values with *KIAA0125* in ours and the TCGA cohorts**

Genes	Correlation coefficient ( <i>p</i> value)				Association with leukemia
	NTUH		TCGA		
<b><i>SPINK2</i></b>	0.661	(3.4E-45)	0.5798	(6.2E-15)	Serine Peptidase Inhibitor; upregulation is associated with poor outcomes in adult patients with AML[31]; integrated into a 6-gene LSC score to identifies high risk pediatric AML[32]
<b><i>MAP7</i></b>	0.653	1.0E-43	0.696	(<E-45)	Microtubule-associated proteins, overexpressed in cytogenetically normal AML patients with dismal outcomes[33]
<b><i>HOPX</i></b>	0.619	(2.6E-38)	0.643	(<E-45)	The smallest homeodomain protein; higher expression predicts poor prognosis in <i>de novo</i> AML[34]
<b><i>MMRN1</i></b>	0.609	(9.7E-37)	0.597	(<E-45)	A member of the elastin microfibrillar interface protein; an adverse marker in both pediatric and adult AML[35]
<b><i>DNMT3B</i></b>	0.599	(1.7E-35)	0.631	(<E-45)	DNA methyltransferases; an important LSC marker[36-38]
<b><i>TCF4</i></b>	0.556	(1.1E-29)	0.626	(<E-45)	A transcription factor; predict outcome in <i>RUNX1</i> mutated and translocated AML[39,40]
<b><i>SLC38A1</i></b>	0.536	(2.3E-27)	0.585	(<E-45)	A glutamine amino acid transporter, overexpressed in AML patients with adverse clinical outcomes[41]
<b><i>DOCK1</i></b>	0.530	(1.1E-26)	0.597	(5.9E-16)	A novel class of guanine nucleotide exchange factors; high expression confers poor prognosis in AML[42]
<b><i>ARHGAP22</i></b>	0.519	(1.5E-25)	0.518	(<E-45)	Rho GTPase activating protein, incorporated in the 17-gene LSC score which predicts treatment response in AML[9]
<b><i>MN1</i></b>	0.502	(1.1E-23)	0.565	(<E-45)	A transcriptional coactivator, overexpression could induce AML in mice and predict ATRA resistance in human AML patients[43,44]
<b><i>ABCG1</i></b>	0.504	(6.7E-24)	0.610	(<E-45)	Belongs to ATP-binding cassette (ABC) superfamily; responsible for important chemoresistance mechanism in AML[45-50]
<b><i>ABCA2</i></b>	0.367	(1.5E-12)	0.507	(2.3E-11)	Belongs to ATP-binding cassette (ABC) superfamily; a strong prognostic biomarker for multidrug resistance in pediatric acute lymphoblastic leukemia[45-50]
<b><i>ABCB1</i></b>	0.353	(1.2E-11)	0.364	(5.2E-6)	Belongs to ATP-binding cassette (ABC) superfamily; responsible for important chemoresistance mechanism in AML[45-50]
<b><i>ABCC1</i></b>	0.310	(3.2E-9)	0.458	(5.2E-9)	Belongs to ATP-binding cassette (ABC) superfamily; responsible for important chemoresistance mechanism in AML[45-50]

410

411

412 **Figure legends**

413 **Figure 1. Box plots depicting expression levels of *KIAA0125* in healthy controls and various**

414 **AML subgroups.**

415 (a) Patients with AML had significantly higher expression of *KIAA0125* than healthy controls;

416 (b) patients who achieved CR after induction chemotherapy had lower expression of BM

417 *KIAA0125* at diagnosis than those who did not; and (c, d) patients with AML, M3 had

418 significantly lower expression of *KIAA0125* while those with more immature subtype M0 or

419 M1 had significantly higher expression in both NTUH cohort (c) and GSE12662 cohort (d).

420 **Figure 2. Kaplan-Meier survival curves stratified by expression of *KIAA0125*.**

421 DFS (a) and OS (b) of the 227 AML patients receiving standard chemotherapy in the NTUH

422 cohort; OS of 201 non-APL patients (c) and 110 cytogenetically normal AML patients (d) who

423 received standard treatment in the NTUH cohort; and OS of 141 patients in the TCGA cohort

424 (e) and GSE12417-GPL96 cohort. Patients with higher *KIAA0125* expression had worse

425 clinical outcomes than those with lower expression.

426 **Figure 3. GSEA enrichment plots of HSC and LSC signatures and scatter plot of genes**

427 **positively associated with higher *KIAA0125* expression.**

428 (a) GSEA enrichment plots show positive association of higher *KIAA0125* expression with

429 HSC and LSC signatures curated from several published reports in both the NTUH and TCGA

430 cohorts; (b) the scatter plot reveals the genes strongly correlated to *KIAA0125* expression in

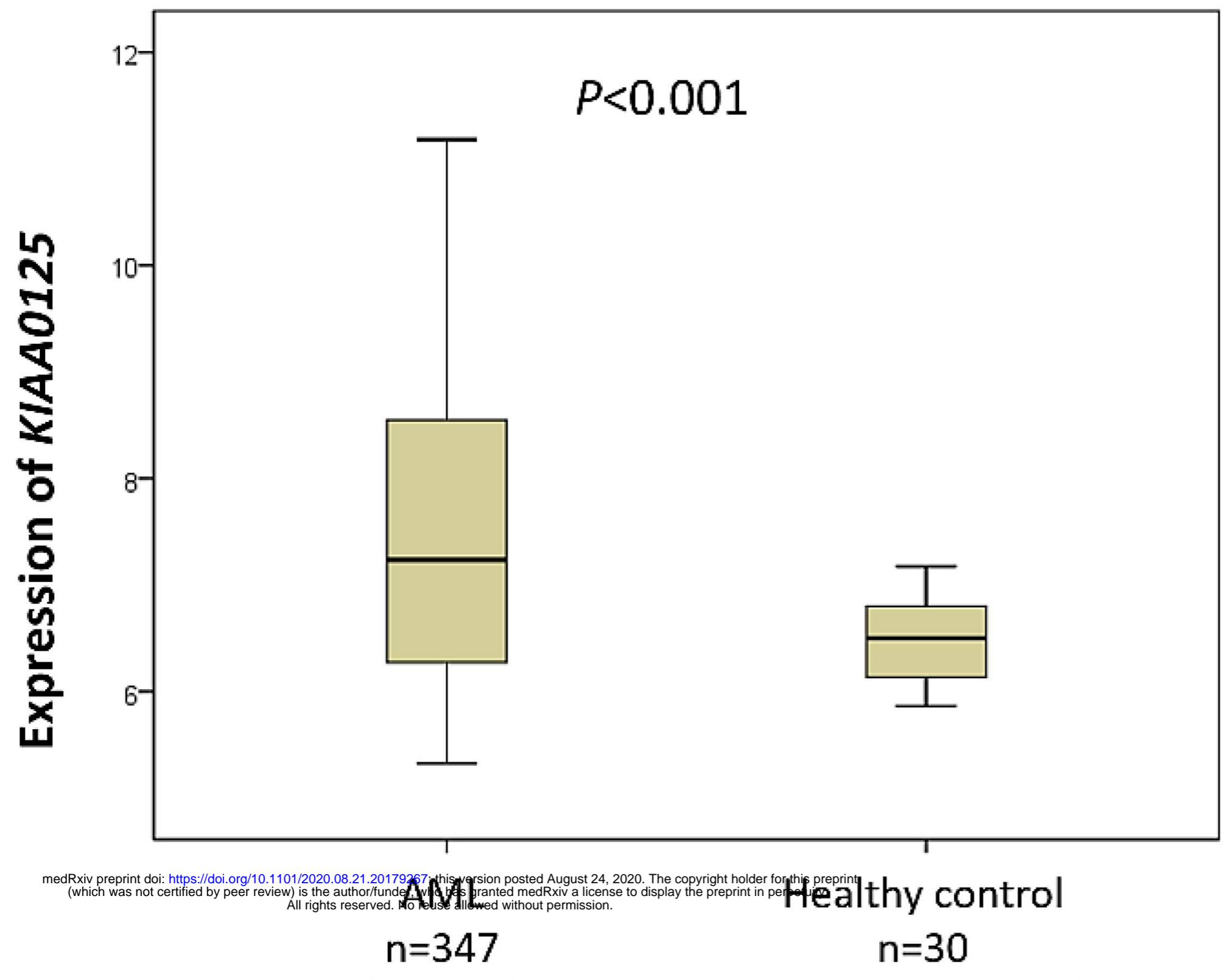
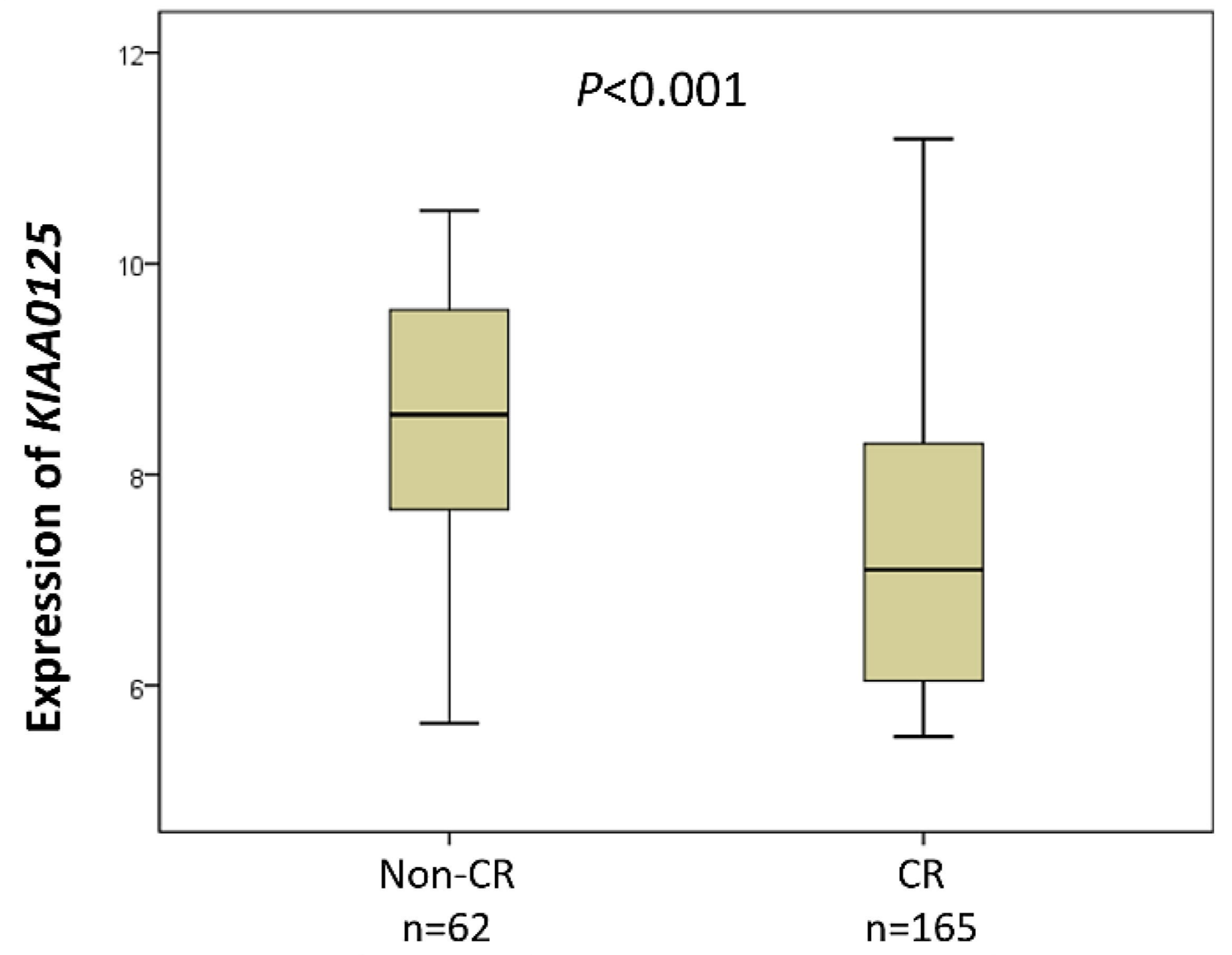
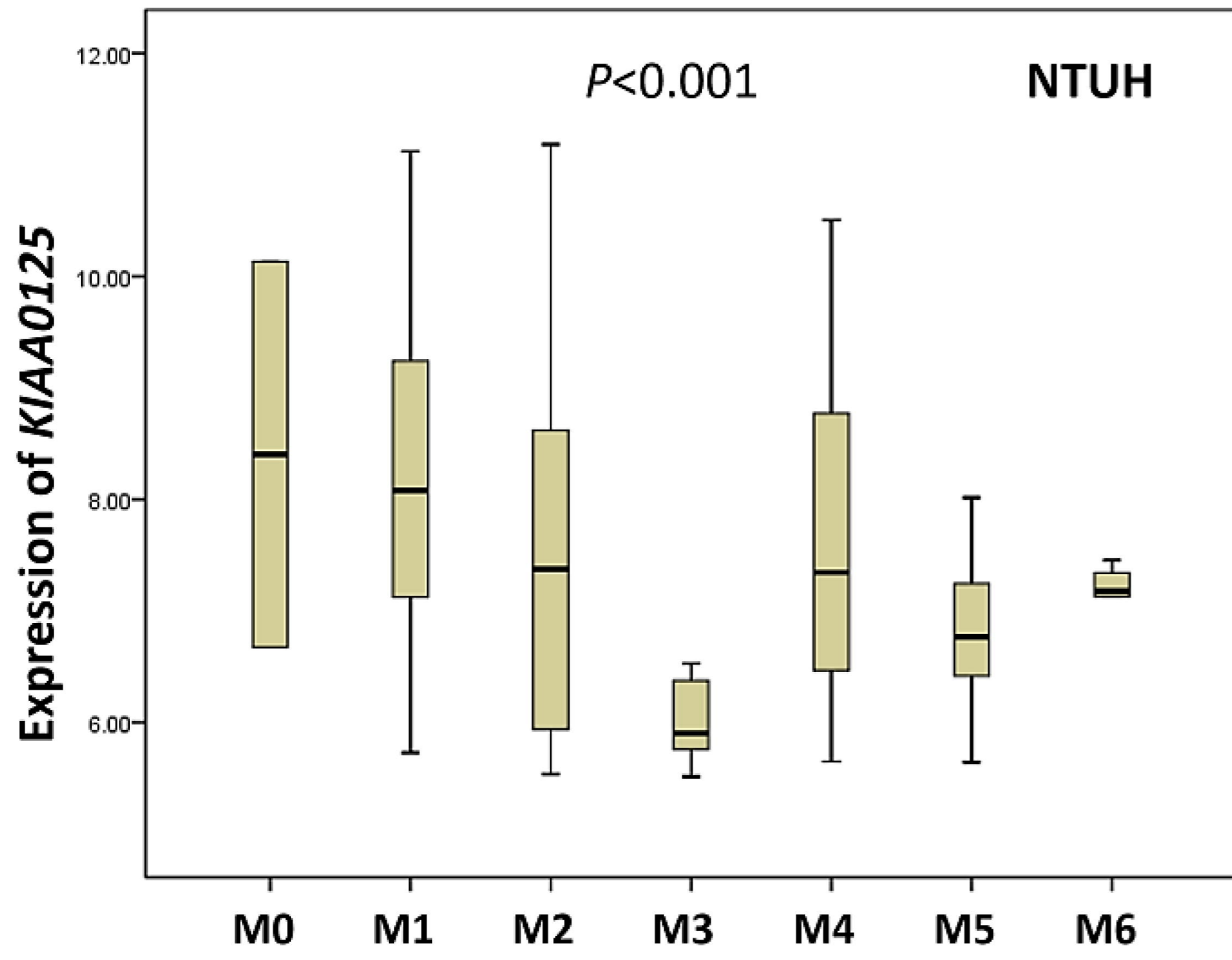
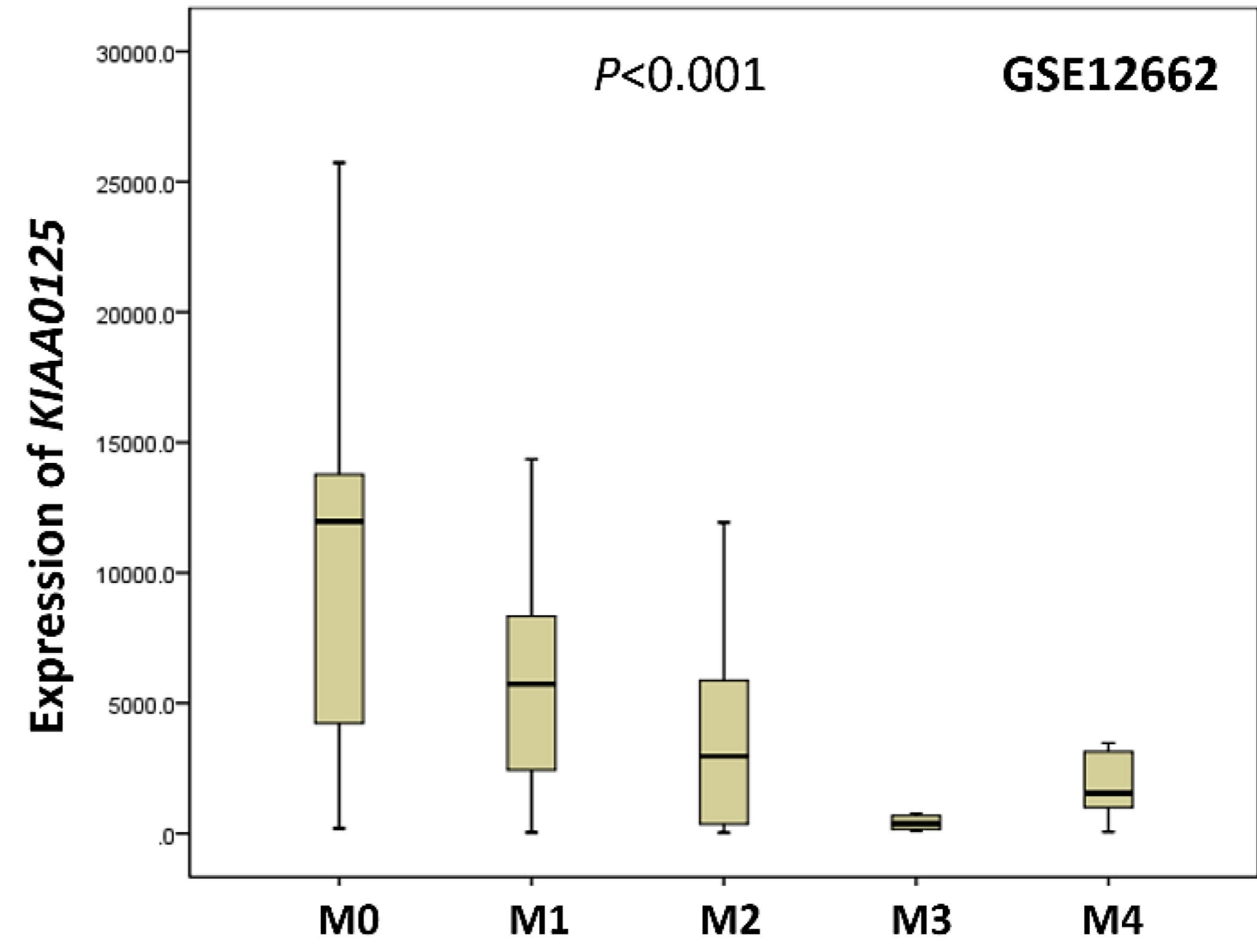
431 both the NTUH and TCGA cohorts (pink). The correlation measurement is based on the

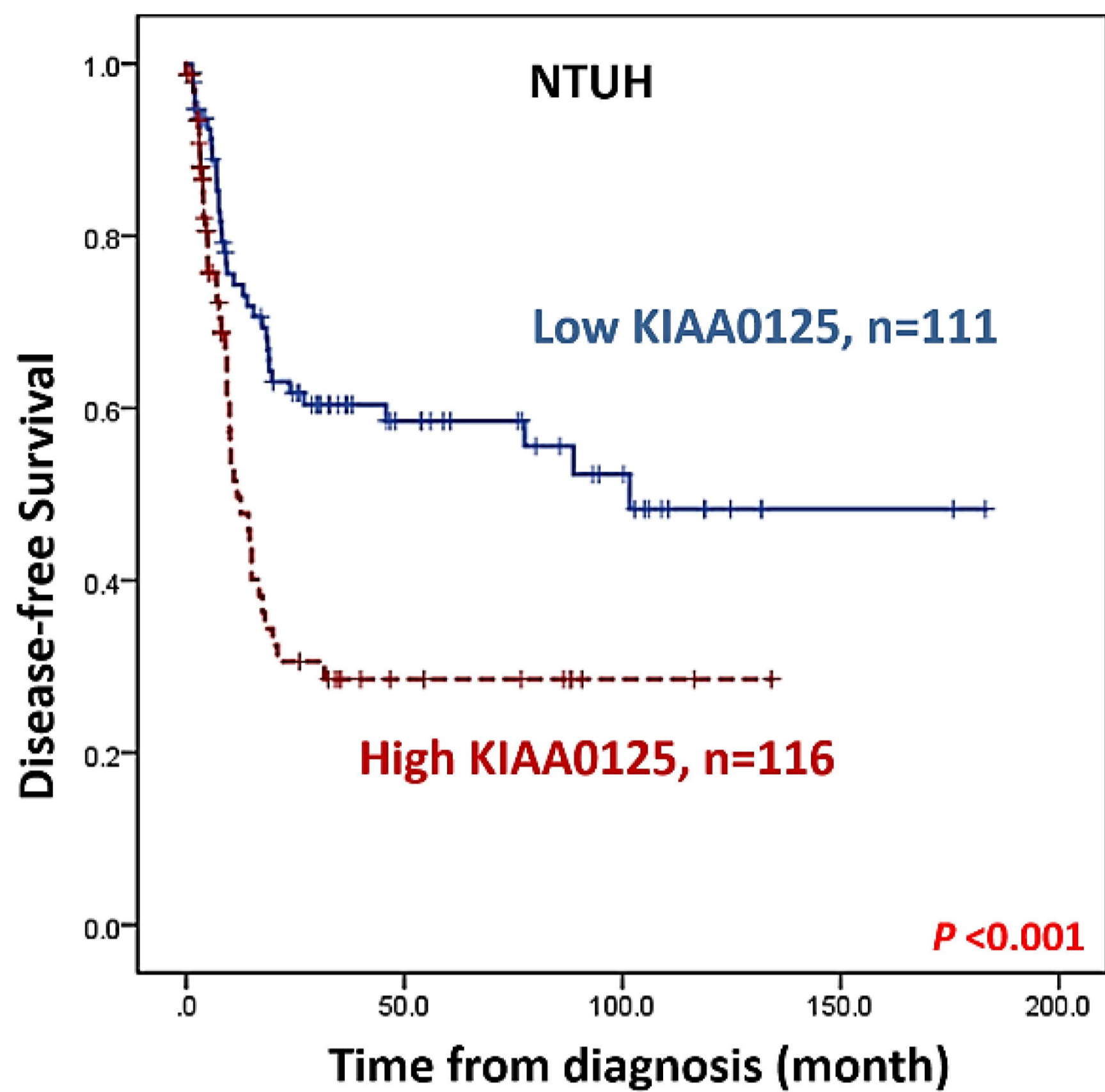
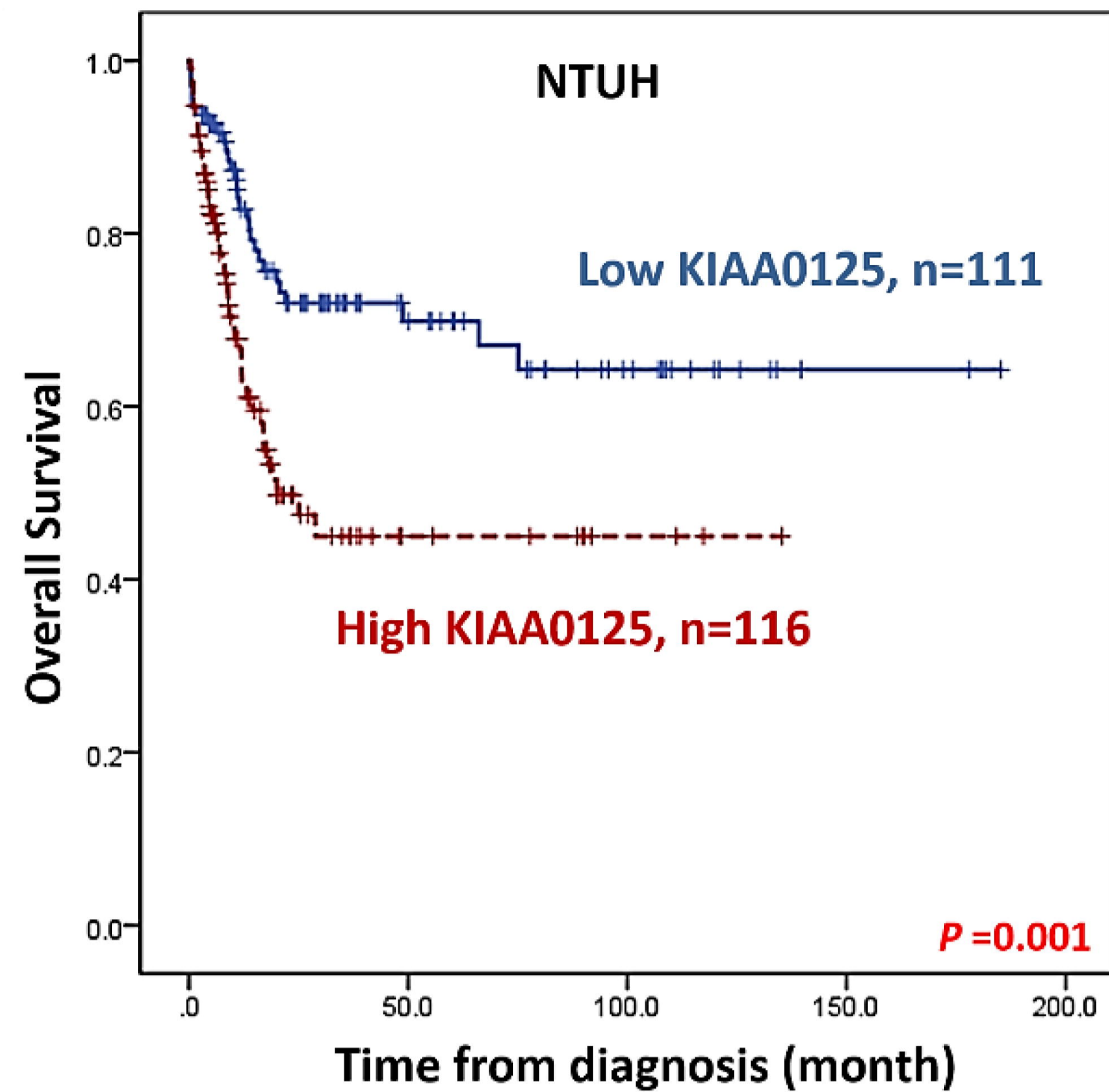
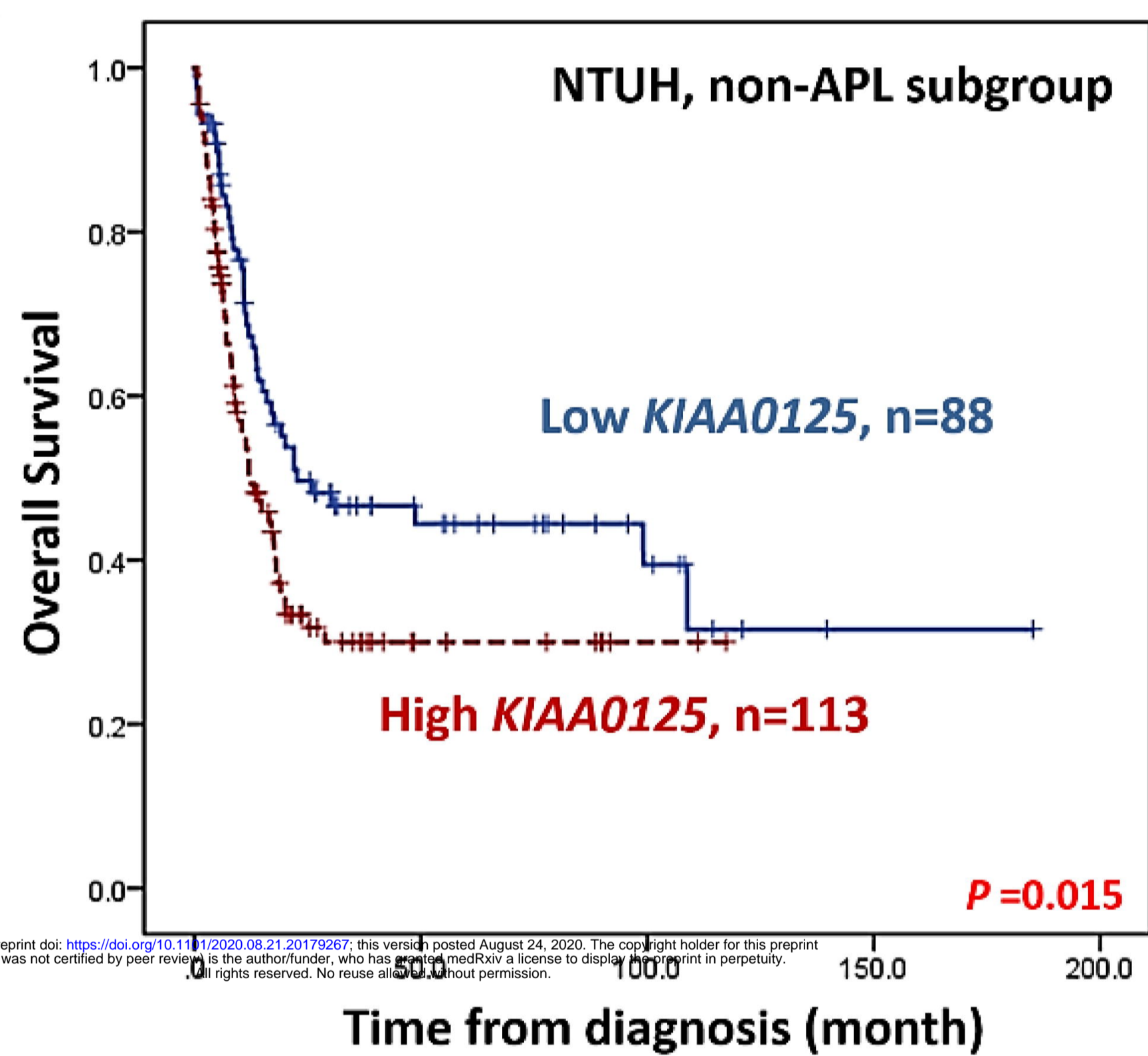
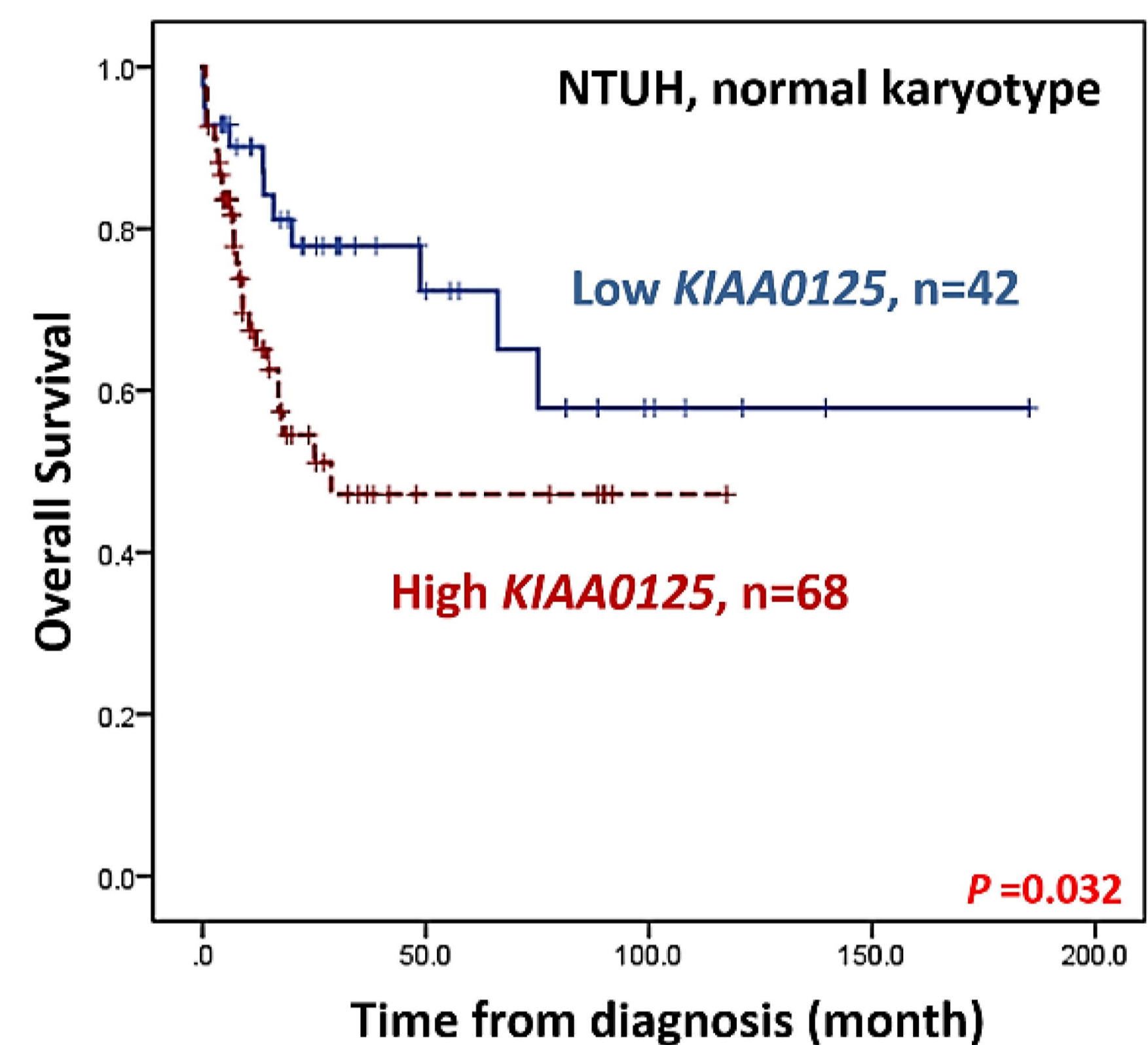
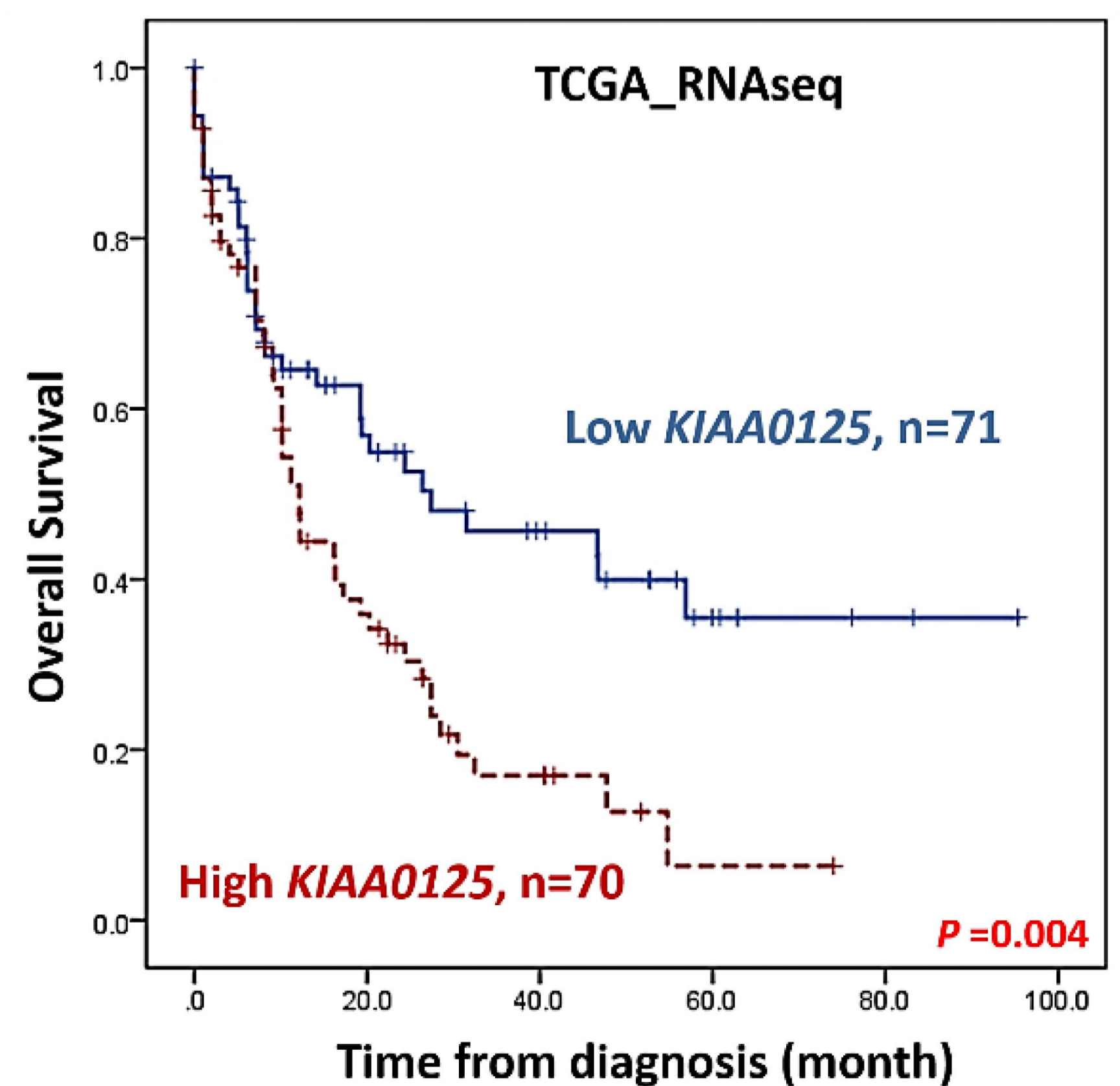
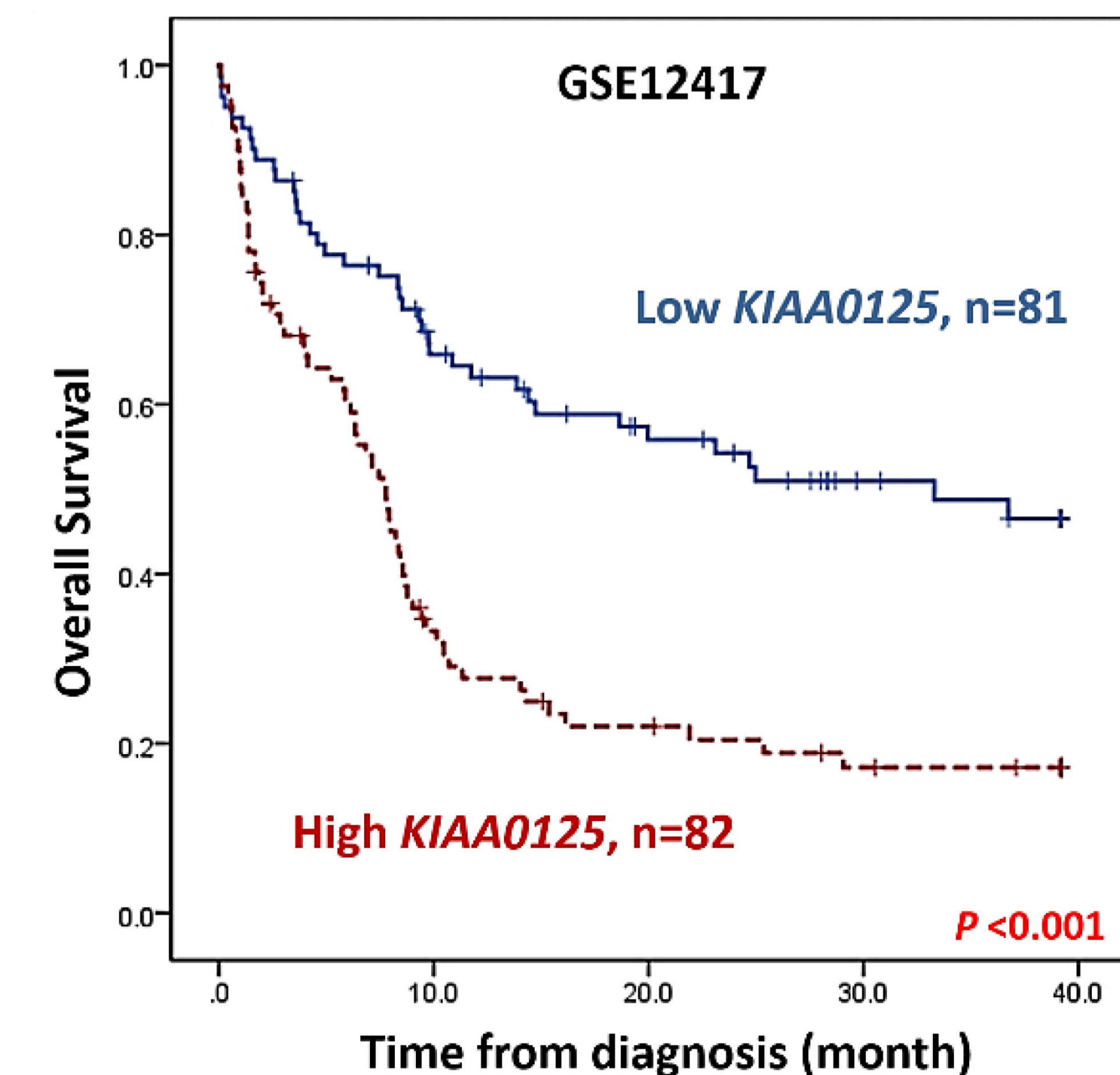
432 Spearman's correlation coefficient between the given gene and *KIAA0125*. The strongly

433 correlated genes are defined as their correlation values at top 5% of all genes in both

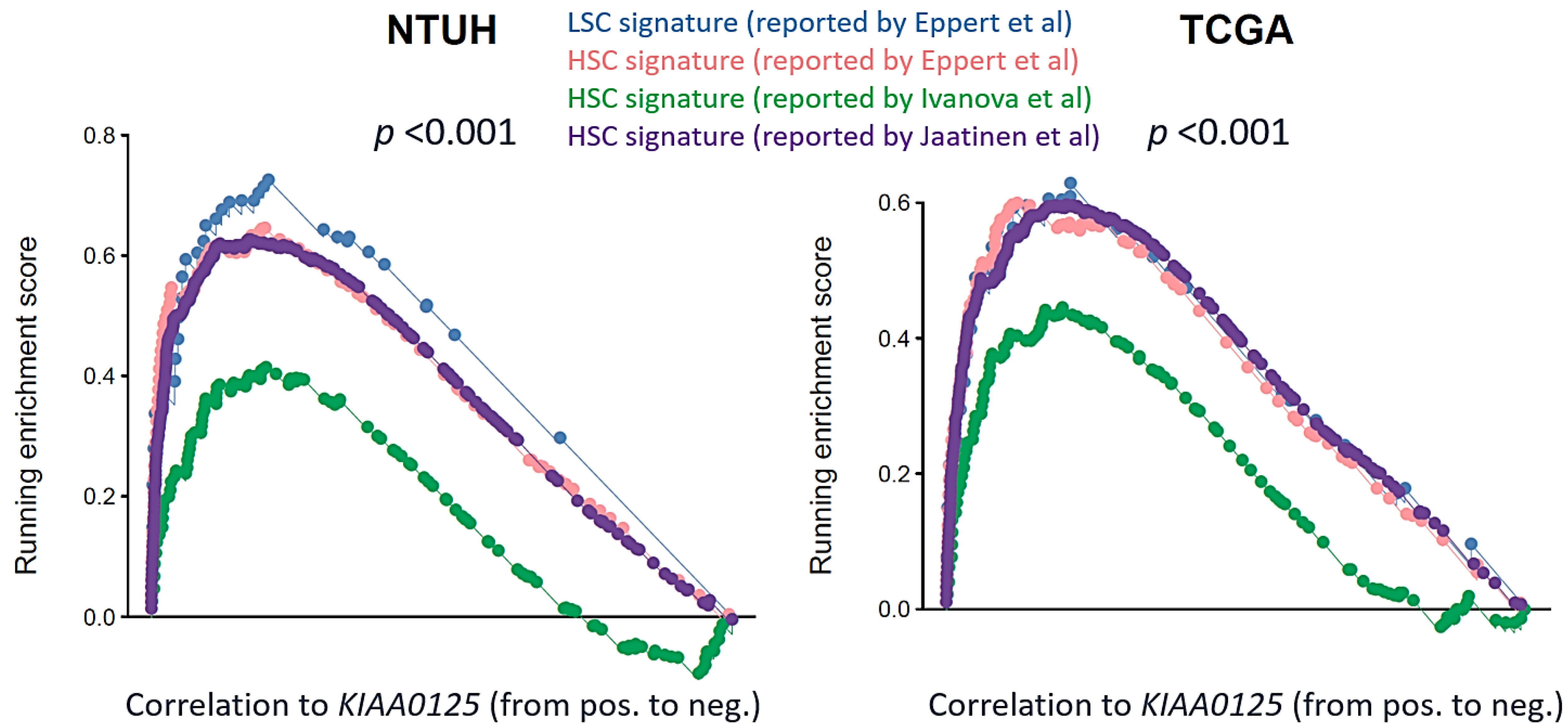
434 cohorts.

435

**a****b****c****d**

**a****b****c****d****e****f**



**a****b**