

1 **Decoding mitochondrial genes in pediatric AML and development of a**
2 **novel prognostic mitochondrial gene signature**

3 Shilpi Chaudhary¹, Shuvadeep Ganguly¹, Jayanth Kumar Palanichamy², Archana Singh²,
4 Dibyabhaba Pradhan³, Radhika Bakhshi⁴, Anita Chopra⁵, Sameer Bakhshi¹

5 ¹Department of Medical Oncology, All India Institute of Medical Sciences, New Delhi, India.

6 ²Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India

7 ³ICMR Computational Genomics Centre, Indian Council of Medical Research, New Delhi,
8 India

9 ⁴Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi, India

10 ⁵Department of Laboratory Oncology, All India Institute of Medical Sciences, New Delhi,
11 India.

12
13 **Corresponding author:**

14 Prof. Sameer Bakhshi

15 Department of Medical Oncology

16 Dr. B.R.A. Institute Rotary Cancer Hospital

17 All India Institute of Medical Sciences,

18 New Delhi – 110029, INDIA

19 Email: sambakh@hotmail.com

20 Office Phone Number: +91-11-29575237

21 Fax Number: +91-11-26588153

22
23 **Abstract Word Count:** 250

24 **Text Word Count:** 5063

25 **No. of Tables:** 8

26 **No. of Figures:** 12

27 **Supplementary files:** 1 (Supplementary Tables: 6)

28 **No. of References:** 52

29 **Running title:** Mitochondria-related prognostic gene signature in AML

30

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

31 **Abstract**

32 **Background**

33 Gene expression profile of mitochondrial-related genes is not well deciphered in pediatric
34 acute myeloid leukaemia (AML). We aimed to identify mitochondria-related differentially
35 expressed genes (DEGs) in pediatric AML with their prognostic significance.

36 **Methods**

37 Children with *de novo* AML were included prospectively between July 2016-December 2019.
38 Transcriptomic profiling was done for a subset of samples, stratified by mtDNA copy
39 number. Top mitochondria-related DEGs were identified and validated by real-time PCR. A
40 prognostic gene signature risk score was formulated using DEGs independently predictive of
41 overall survival (OS) in multivariable analysis. Predictive ability of the risk score was
42 estimated along with external validation in The Tumor Genome Atlas (TCGA) AML dataset.

43 **Results**

44 In 143 children with AML, twenty mitochondria-related DEGs were selected for validation,
45 of which 16 were found to be significantly dysregulated. Upregulation of *SDHC* ($p<0.001$),
46 *CLIC1* ($p=0.013$) and downregulation of *SLC25A29* ($p<0.001$) were independently predictive
47 of inferior OS, and included for developing prognostic risk score. The risk score model was
48 independently predictive of survival over and above ELN risk categorization (Harrell's c-
49 index: 0.675). High-risk patients (risk score above median) had significantly inferior OS
50 ($p<0.001$) and event free survival ($p<0.001$); they were associated with poor-risk cytogenetics
51 ($p=0.021$), ELN intermediate/poor risk group ($p=0.016$), absence of RUNX1-RUNX1T1

52 (p=0.027), and not attaining remission (p=0.016). On external validation, the risk score also
53 predicted OS (p=0.019) in TCGA dataset.

54 **Conclusion**

55 We identified and validated mitochondria-related DEGs with prognostic impact in pediatric
56 AML and also developed a novel 3-gene based externally validated gene signature predictive
57 of survival.

58 **Keywords:** Acute Myeloid Leukaemia, Mitochondria, Gene signature, RNA sequencing,
59 Child

60

61

62

63

64

65

66

67

68

69

70

71

72 **Introduction**

73 Despite recent advancements, the survival in pediatric acute myeloid leukaemia (AML)
74 continues to remain dismal(1). Various molecular and genetic alterations are frequently used
75 for risk stratification, identification of therapeutic targets as well as predicting disease
76 prognosis in AML(2). Whole genome and transcriptome sequencing have been extensively
77 used in AML to identify potential novel molecular targets and developing prognostic gene
78 signatures to predict survival, relapse and risk stratification(3–5). However, data on potential
79 mitochondrial genes with impact on AML are limited.

80 Dysregulation of mitochondrial pathways have been implicated in pathogenesis and
81 progression of various malignancies(6). Multiple studies have reported the role of
82 mitochondrial DNA (mtDNA) mutations, metabolic pathways and oxidative phosphorylation,
83 on disease biology and prognosis of AML(7,8). We have previously reported the relationship
84 of mutations in mtDNA regulatory region with mitochondrial gene expression, and their
85 impact on survival in children with AML(9–11). Considering the impact of mitochondrial
86 pathways in outcome of AML, it is important to explore tumor cell heterogeneity in AML
87 with respect to mitochondrial transcriptome and identify potential therapeutic or prognostic
88 molecular targets.

89 Recently, we have reported that high mtDNA copy number is associated with poor outcome
90 in paediatric AML and also identified its potential regulation through *PGC1A*(12). In the
91 current study, among children with AML stratified according to mtDNA copy number, we
92 identified mitochondria-related differentially expressed genes (DEGs) through whole
93 transcriptome sequencing. We further validated the topmost identified mitochondria-related
94 DEGs in a cohort of paediatric AML patients and formulated a prognostic mitochondrial gene
95 signature for predicting survival outcome. We then validated this gene signature in an

96 external cohort of adult AML patients from The Cancer Genome Atlas (TCGA) dataset along
97 with estimation of predictive ability of the developed prognostic gene signature.

98 **Methodology**

99 **Study design, patient population, treatment and clinical follow up**

100 This was a prospective observational cohort study that included consecutive *de novo*
101 paediatric (≤ 18 years) patients with AML registered from July 2016 to December 2019 at
102 medical oncology outpatient clinic of our cancer centre. The workflow of the study is
103 depicted in figure 1. Study was ethically approved by institute ethics committee (IEC/NP-
104 336/2012, IEC/PG-79/22.03.2017) and informed consent was taken from care givers and
105 assent was obtained from all participants (≥ 8 years). Patients with granulocytic sarcoma
106 without marrow involvement, acute promyelocytic leukaemia, and mixed phenotypic acute
107 leukaemia were excluded. Fifty age-matched patients of solid malignancies without marrow
108 involvement were also enrolled as controls. Remission status and survival outcomes were
109 noted.

110 **Risk Assessment by karyotyping and mutation analysis**

111 Conventional cytogenetics were done at baseline to identify translocations, inversions,
112 deletion as well as other chromosomal abnormalities for risk stratification of pediatric AML.
113 Mutation profiling of *RUNX1-RUNX1T1* (Runt-related transcription factor 1-RUNX1 partner
114 transcriptional co-repressor 1 fusion transcript), *CBFB-MYH11* (Core binding factor beta-
115 myosin heavy chain 11 fusion transcript), *FLT3-ITD* (Fms like tyrosine kinase 3-internal
116 tandem duplication), and *NPM1* (Nucleophosmin 1) by reverse transcriptase polymerase
117 chain reaction (PCR) were performed at baseline for risk assessment as per European
118 LeukemiaNet (ELN) recommendation (13).

119 **Treatment protocol**

120 All patients were treated with uniform induction protocol, i.e., 3+7 regimen including
121 daunorubicin 60 mg/m² day 1-3 and cytarabine 100mg/m² continuous infusion day 1-7.
122 Consolidation therapy with three cycles of high dose cytarabine at 18g/m² were given to
123 patients after achieving complete remission (CR) whereas repeated induction with ADE
124 regimen (cytarabine: 100 mg/m² twice daily, day 1-10; daunorubicin: 50 mg/m², day 1-3;
125 and etoposide: 100 mg/m², day 1-5) were used for refractory or relapse cases. Patients at
126 CR2 underwent allogeneic hematopoietic stem cell transplantation with matched sibling
127 donor if available (13,14).

128 **Isolation of bone marrow mononuclear cells (BMMCs), DNA and RNA extraction and** 129 **quantification**

130 BMMCs were isolated by Ficoll-Hypaque (Sigma diagnostics, USA) density gradient
131 centrifugation followed by isolation of total RNA and DNA. Briefly, BM samples collected
132 from all the patients and controls were layered onto the histopaque in 15 ml falcon tube
133 followed by centrifugation at 400g for 30 min at room temperature. The mononuclear cells
134 layer was carefully taken out after centrifugation and washed twice with phosphate buffer
135 saline (PBS). BMMC were then stored for DNA isolation using isopropanol precipitation
136 method and total RNA was isolated using TRIZol method as per manufacturer's protocols.
137 Quality of DNA and RNA was checked by Nanodrop 1000 (Thermo Fisher) and integrity of
138 RNA was checked by TapeStation (Agilent).

139 **Estimation of mitochondrial DNA copy number**

140 Relative mtDNA copy number was assessed by fluorescent DNA binding dye SYBR based
141 quantitative real time PCR using, with Roche Light Cycler 480 II. The relative mitochondrial
142 DNA copy number was normalized to expression of nuclear gene *ACTB*, which was chosen
143 as a nuclear housekeeping gene. The mitochondrial DNA copy number, normalized to copies

144 per cell, in each subject and control sample was calculated using the following formula:
145 $2^{[Ct(\beta\text{-actin}) - Ct(\text{minor arc})]}$ (Ct being respective cycle thresholds). Relative
146 mitochondrial DNA copy numbers of patients were then compared with controls.

147 **Whole transcriptome sequencing, identification of differentially expressed genes (DEG)** 148 **and selection of DEG for analysis**

149 All the samples were classified into three separate groups based on relative mtDNA copy
150 number(12). Patients were categorized into: AMLCN_H (mtDNA copy number \geq 75th
151 percentile), AMLCN_I (mtDNA copy number 50th to 75th percentile) and AMLCN_L
152 (mtDNA copy number < 50th percentile) groups. A subset of samples was randomly selected
153 from each of the three sub-groups and controls with RNA integrity score above 7 and a total
154 of 15 samples (12 patients including 3 from AMLCN_H group, 4 from AMLCN_I group, 5
155 from AMLCN_L group and 3 controls) were sent for whole transcriptome profiling for the
156 identification of DEGs compared to controls.

157 **Library Construction and sequencing**

158 The sequencing library was prepared by random fragmentation of the cDNA sample,
159 followed by 5' and 3' adapter ligation, after end-repair and the addition of an 'A' base and
160 SPRI clean-up. The prepared cDNA library was amplified using PCR for the enrichment of
161 the adapter-ligated fragments. The individual libraries were quantified using a NanoDrop
162 spectrophotometer (Thermo Scientific) and validated for quality with a Bioanalyzer (Agilent
163 Technologies). Adapter-ligated fragments were then PCR amplified and gel purified. cDNA
164 library was used for sequencing which was carried out on Illumina Hiseq 4000 NGS
165 platform.

166 **Processing of the reads and differential expression analysis**

167 The quality of raw reads was first checked by FastQC version v0.11.8. Trimmomatic was
168 performed to remove adapter sequences and low-quality reads for further analysis. The
169 trimmed reads were then aligned to reference Human genome (hg38) using HISAT2 tool.
170 SAM tool was used to convert SAM files into BAM files which was used for quantification
171 and estimation of aligned reads by StringTie (v2.0.6). Lastly differential expression analysis
172 was performed by limma Bioconductor package (version 3.48.1). Absolute fold change value
173 ≥ 2 (a \geq two-fold change in expression, either upregulated or downregulated) and adjusted p
174 value ($q \leq 0.05$) threshold was used for the identification of differentially expressed genes
175 (DEGs). Volcano plots were made with log fold change; p value and adjusted p value of the
176 transcripts were calculated using devtools in R package (version 3.6.1). Gene name of the
177 corresponding transcripts with significant p value, adjusted p value and log 2-fold difference
178 compared to controls, were fetched using gProfilerR library in R package. Exclusive
179 differentially expressed genes were identified among the groups by making Venn diagram
180 using Venny (Venny 2.1) online tool. (Figure 2).

181 **Construction of protein-protein interaction network, determination of HUB genes and** 182 **MCODE analysis**

183 A interactive network of proteins encoded DEGs present in the respective groups were
184 constructed using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)
185 database (15) in Cytoscape. Cytohubba plugin was used for identifying Hub genes among the
186 whole interactome network. Those genes which got enriched using at least six different
187 topological algorithms among Degree, Edge Percolated Component (EPC), Maximum
188 Neighbourhood Component (MNC), Density of Maximum Neighbourhood Component
189 (DMNC), Maximal Clique Centrality (MCC) and centralities based on shortest paths, such as
190 Bottleneck (BN), Eccentricity, Closeness, Radiality, Betweenness, and Stress were
191 considered as Hub genes(16). Cytoscape plug-in MCODE clustering algorithm was used for

192 identifying the maximum ranked cluster which has a highly interconnected region , also
193 known as seed nodes as well as their neighbour nodes in the whole network (17). Subcellular
194 location of the genes was identified from the gene ontology table. Compartment mitochondria
195 score was used for selecting mitochondrial related genes among all the AML subgroups.

196 **Selection and validation of mitochondria-related DEGs**

197 Out of all identified DEGs from transcriptome sequencing, mitochondria-related genes were
198 filtered using Cytoscape compartment mitochondrion score (0 being minimum and 5 being
199 highest)(18). DEGs with topmost mitochondrial compartment score were selected for
200 validation in a cohort of paediatric patients with AML. The genes of MCODE cluster 1 and
201 Hub genes were assessed for their mitochondrial localization as above and genes in each
202 group with highest mitochondrial compartment score were selected for validation. Based on
203 these selection strategies, a total of 20 mitochondria-related DEGs were selected for
204 validation. Real time PCR was performed to validate the selected mitochondria-related genes
205 using specific primers (Table S1) and the gene expressions were quantified per previously
206 described protocol(12). All experiments were replicated in triplicates.

207 **Comparison of validated mitochondria-related DEGs in TCGA data set**

208 For external validation of mitochondrial related DEGs, the RNA-sequencing data (Illumina
209 HiSeq 2000) of TCGA adult AML(LAML) dataset was chosen, which is one of the largest
210 datasets of transcriptomic profile in AML with recorded clinical
211 outcome(https://www.cbioportal.org/study/summary?id=laml_tcga). The adult dataset was
212 specifically chosen to see the impact of prognostic impact of the validated mitochondria-
213 related age group in a different age group as well. The expression of validated DEGs was
214 compared with LAML data set using online available GEPIA2 (Gene Expression Profiling
215 Interactive Analysis) web server (<http://gepia2.cancer-pku.cn/#index>)(19).

216 **Statistical Methods**

217 **Prognostic impact of mitochondria-related DEGs and development of mitochondrial**
218 **gene signature**

219 Statistical analysis was carried out in SPSS (v23, IBM, NY, USA). Descriptive statistics were
220 used to summarize baseline characteristics. Gene expression was reported as median values
221 with interquartile ranges. Gene expression values and clinical continuous variables with non-
222 parametric distribution were compared by Mann Whitney test. Clinical categorical variables
223 were compared by Chi-square test/Fisher's exact test as applicable. Alpha error was adjusted
224 for multiple comparisons by Bonferroni correction. Kaplan Meier method was used to
225 analyse time to event outcomes. Duration from enrolment to relapse or death due to any cause
226 was considered as event free survival (EFS). Time from enrolment to death due to any cause
227 was defined as overall survival (OS). Survival data was censored till 31st Dec 2020. The
228 follow-up estimation was done by reverse Kaplan Meier method.

229 Prognostic impact of all validated DEGs on OS of the whole validation cohort was performed
230 by multivariable Cox regression analysis in a forward stepwise manner based on log
231 likelihood change. Validated DEGs with significant ($p < 0.05$) predictive impact on OS in
232 multivariable analysis were included for the prognostic gene signature model. The
233 proportional hazard assumption was assessed by Schoenfeld global test. Internal validation of
234 the multivariable prognostic model was carried out by bootstrapping method (10000
235 resampling) and genes that did not satisfy bootstrapping validation were excluded. A
236 prognostic risk score was generated using cox regression coefficient Beta (β) values of
237 included genes, of the final multivariable model as below:

$$\text{Risk score} = \left[\sum_{k=0}^n \binom{n}{k} \{ \beta_k \times (\text{Expression } mRNA_k) \} \right] \times 100$$

238 The area under the time-dependent receiver operating characteristic (ROC) curve (Timed
239 AUC) for 12-months and 18-months survival was estimated and Harrel's C-index of the
240 prognostic model was calculated using the R package "survminor" in R (version 4.0.3).
241 Patients were classified into two groups based on their risk score above (High-risk) and
242 below (Low-risk) the median. The survival outcomes of the patients were compared between
243 high-risk score vs low-risk score patients using Kaplan Meier analysis to evaluate the
244 prognostic significance of the gene signature model.

245 **Impact of clinical features and independent prognostic value of the gene signature**

246 The role of demographic and clinical features, including gender, age, haemoglobin,
247 hyperleukocytosis ($\geq 50000/\mu\text{l}$), platelet count, presence of chloroma and ELN risk
248 stratification(2) on survival outcome was analysed using the Cox regression. Factors with
249 $p < 0.1$ in univariable analysis were included for multivariable Cox regression in a forward
250 stepwise manner using log likelihood change. Clinico-demographic factors which were
251 significant in multivariable analysis were included in a multivariable Cox regression model
252 along with gene signature risk score to explore the independent predictive value of gene
253 signature. The timed AUC using 12-months survival and 18-months survival as the outcome
254 and Harrel's C-index of the clinical prognostic model and combined clinical and gene
255 signature prognostic model were compared for identifying the additional prognostic benefits
256 of gene signature over clinical parameters. The impact of mtDNA copy number on survival
257 outcome was also analysed similarly.

258 **External validation of mitochondrial prognostic gene signature in TCGA dataset**

259 The prognostic impact of our gene signature risk score on OS was done in TCGA LAML
260 (n=179) dataset by Cox regression analysis. Patients were similarly sub-grouped into high-
261 risk and low-risk category based on median value of the gene signature; the survival
262 outcomes of the patients were compared between high-risk score vs low-risk score patients

263 using Kaplan Meier analysis and timed AUC of 12-month and 18-month survival was
264 evaluated. Based on available karyotyping data, patients of the TCGA dataset were grouped
265 into poor-risk karyotype and others (including good and intermediate-risk karyotype). The
266 karyotype category and mitochondrial gene risk score were assessed for their impact on OS
267 by a multivariable Cox regression model to explore the independent predictive value of the
268 gene signature in the external cohort as well.

269 **Results**

270 **Patients' recruitment and baseline clinical features**

271 Total 170 patients were enrolled, out of which 27 patients (5 patients were AML M3, 4 had
272 granulocytic sarcoma without marrow involvement, and 18 patients had insufficient samples)
273 were excluded. The baseline demographic and clinical characteristics of final 143 patients are
274 summarized in Table 1. Median age was 10 years (range: 0.8-18 years) and 50% of the
275 patients were classified as ELN good risk. Total 104 patients (72.7%) achieved complete
276 remission (CR) after induction therapy. At median follow-up of 36 months (32.67-39.33
277 months), the median OS was 21.93 months (13.54–30.31months). The clinical characteristics
278 of the TCGA LAML dataset are summarized in Table S2.

279 **Identification of DEGs in paediatric AML based on mtDNA copy number**

280 We identified 898,769, and 953 significantly dysregulated transcripts in AMLCN_H,
281 AMLCN_I and AMLCN_L groups respectively by whole transcriptome sequencing as
282 represented in volcano plots (Figure3). Majority of genes were found significantly
283 downregulated in all three groups whereas the number of dysregulated genes were higher in
284 AMLCN_H group compared to other two groups. A total of 351 DEGs (59 upregulated and
285 292 downregulated) were identified in AMLCN_H. Similarly, AMLCN_I and AMLCN_L
286 groups had 290 (66 upregulated and 224 downregulated) and 332 (47 upregulated and 285
287 downregulated) DEGs respectively as compared to controls.

288 **Identification of mitochondria-related DEGs, hub genes and selection of genes for**
289 **validation**

290 Out of all DEGs, 78, 58, and 71 mitochondria-related DEGs were identified in AMLCN_H,
291 AMLCN_I and AMLCN_L groups respectively. Among them, 35 genes were common in all
292 three subgroups, whereas 18, 12 and 14 mitochondria-related DEGs were exclusively present
293 in AMLCN_H, AMLCN_I and AMLCN_L groups respectively (Figure 4). In AMLCN_H,
294 AMLCN_I and AMLCN_L groups, we identified 17, 18 and 17 hub genes respectively using
295 CytoHubba analysis, of which eight were common among all subgroups (Table S3, Figure 5).
296 Furthermore, using MCODE analysis, clusters with maximum scores were generated and
297 seed gene was determined in the three groups (Figure 6). *MMP9* was identified as seed node
298 with maximum MCODE score in both AMLCN_H and AMLCN_I group (Table S4). Based
299 on the mitochondrial compartment score, CytoHubba and MCODE analyses, a total of 20
300 DEGs were selected for further validation (Table 2).

301 **Validation of selected DEGs, comparison with TCGA database**

302 In the validation cohort of 143 AML patients, the expression of *SLC25A3*, *SDHC*,
303 *RACK1/GNB2L1*, *FASTKD1*, *ATP5J*, *CLIC1*, *GLUD1*, and *SLC25A29* were found to be
304 significantly upregulated (Figure 7, Table 3) while *FASLG*, *HRK*, *ALAS2*, *SLC25A21*,
305 *CYP11B1*, *SNCA*, *MMP9*, and *OLFM4* were significantly downregulated (Figure 8, Table 3)
306 compared to controls. Two selected genes, *LIG1* and *MRPL51* did not show significant
307 dysregulation while *LONPI* had a reverse expression in the validation compared to
308 transcriptomic expression profile. Upon comparison with TCGA dataset of adult AML
309 patients, similar dysregulation was observed for *ALAS2*, *SLC25A21* and *SLC25A29* genes
310 while a reverse expression pattern was observed for *ATP5J* and *CLIC1* genes; none of the
311 other genes showed significant dysregulation in the TCGA dataset (Table 3).

312 **Mitochondria-related DEGs and mtDNA copy number**

313 On univariable analysis, increased mtDNA copy number was significantly associated with
314 poor event free survival (HR= 2.14; 95%CI (1.39-3.29); p=0.001) and overall survival (HR=
315 2.77; 95% CI (1.70-4.59); p<0.001) (Figure 9). In patients with increased mtDNA copy
316 number, expression of *SLC25A3*, *SDHC*, *RACK1/GNB2L1* and *FASTKD1*, were significantly
317 higher compared to those with low mtDNA copy number (Figure 10). Exclusive elevated
318 expression of these 4 genes were also observed in transcriptome of samples with
319 high/intermediate mtDNA copy number (AMLCN_H/AMLCN_I) compared to low mtDNA
320 copy number (AMLCN_L). On correlation analysis, these 4 genes along with 2 other genes
321 *CLIC1* and *ATP5J* showed significant positive correlation with mtDNA copy number (Table
322 4).

323 **Predictive ability of expression of validated DEGs on survival outcome and** 324 **establishment of the prognostic gene signature**

325 On multivariable analysis, upregulated expression of 2 genes, *SDHC* (HR 1.29; 95% CI
326 (1.14-1.41); p<0.001) and *CLIC1* (HR 1.20; 95% CI (1.04-1.38); p=0.013), and
327 downregulation of *SLC25A29* (HR 0.88; 95% CI (0.83-0.93); p<0.001) were found to be
328 independently predictive of worse OS (Table 5) and they were included for the development
329 of a prognostic gene signature model. All these 3 genes (*SDHC*, *CLIC1*, *SLC25A29*) satisfied
330 internal validation by bootstrapping (Table S5), and were finally selected for prognostic
331 model building. Beta coefficient of each of the variables were used for calculation of risk
332 score as follows:

$$\text{Risk score} = [(0.237 \times \text{Expression } SDHC) + (0.179 \times \text{Expression } CLIC1) + ((-0.131) \\ \times \text{Expression } SLC25A29)] \times 100$$

333 The formula was used to calculate risk score of all the patients. Risk score median value
334 (10.382) was taken as the cut-off for subgrouping patients into high-risk and low risk group.

335 Patients with high-risk scores (≥ 10.382) had inferior OS (HR 1.010; 95% CI (1.007-1.014):
336 $p < 0.001$) compared to those with low-risk score (< 10.382) (Figure 11A). Harrel's C-index of
337 the prognostic model was 0.675. The timed AUC of the risk score for 12 months and 18
338 months survival was 0.747 and 0.736 respectively (Figure 11(B, C)).

339 **Association of gene signature-based risk score with event free survival (EFS)**

340 On multivariable Cox regression analysis, upregulation of *SDHC* (HR 1.225; 95% CI (1.100-
341 1.363); $p < 0.001$) and downregulation of *SLC25A29* (HR 0.905; 95% CI (0.860-0.952);
342 $p < 0.001$) were also predictive of worse EFS. We also found that patients with high-risk score
343 had significantly lower EFS as compared to low-risk score patients (HR 1.008; 95% CI
344 (1.001-1.012); $p < 0.001$) (Table 5). Harrel's C-index of prognostic model was 0.626. The
345 timed AUC of the risk score for 12 months and 18 months EFS was 0.617 and 0.612
346 respectively

347 **Impact of baseline clinical features on survival outcome and association with gene** 348 **signature model**

349 On univariable Cox regression analysis, ELN intermediate/poor risk and absence of chloroma
350 were significantly associated with inferior OS and only ELN category came out to be an
351 independent prognostic factor in multivariable analysis (Table 6, Figure 11D). Furthermore,
352 on multivariable analysis, both the ELN risk category ($p = 0.040$) and risk score ($p < 0.001$)
353 were found to be independent prognostic factors for OS. We also performed multivariable
354 analysis including mtDNA copy number and observed that all three factors i.e. risk score
355 ($p < 0.001$), ENA risk categories ($p = 0.012$) and mtDNA copy number ($p = 0.012$) were
356 independent prognostic factors for OS (Table S6).

357 **Impact of combined clinical and gene signature model on survival outcome of the cohort**

358 To compare the predictive ability of our gene signature risk score and ELN risk stratification
359 on OS of AML patients, a time dependent AUC was constructed. Harrel's C-index of the

360 model was 0.59 and the timed AUC of ELN risk category on 12 months and 18 months
361 survival was 0.60 and 0.64 respectively (Figure 11E, F). We combined the ELN risk strategy
362 with our risk score and calculated the predictive ability of the model. The Harrel's C-index of
363 the model was 0.688 and the timed AUC of combining ELN risk strategy with gene signature
364 risk score for 12 months and 18 months was 0.761 and 0.765 respectively (Figure 11H and
365 2I).

366 **Association of gene signature risk score on disease characteristics**

367 We found that a high-risk score was significantly associated with poor risk
368 cytogenetics ($p=0.021$), absence of RUNX1-RUNX1T1 translocation ($p=0.027$) and ELN
369 intermediate/poor risk group ($p=0.016$). Furthermore, the proportion of patients achieving CR
370 was significantly higher in the low-risk group as compared to the high-risk group ($p=0.017$)
371 (Table 7). On subgroup analysis, it was observed that the mitochondria-related gene signature
372 risk score category was significantly predictive of survival outcome across all clinically
373 relevant subgroups except in those with intermediate/poor-risk karyotype (Figure 12).

374 **Predictive ability of combined gene signature and ELN category model on survival** 375 **outcome**

376 The predictive ability of gene signature score along with ELN risk stratification on survival
377 outcome of paediatric AML patients was also assessed. Patients with low gene signature
378 score (low risk) belonging to ELN good risk category had significantly better survival
379 outcome (Median OS: Not reached) and predicted 12-months ($80\pm 6\%$), as well as 18-
380 months ($75\pm 7\%$) survival. Similarly, patients with high gene signature score (high risk)
381 belonging to ELN intermediate/poor risk category had significantly inferior outcome (4.67
382 months (0-3.71)) with 12-months and 18-months predicted survival of $33\pm 8\%$ and $25\pm$
383 7% respectively. On the other hand, patients belonging to other groups (ELN
384 intermediate/poor risk and low-risk; ELN good risk and high-risk score) had intermediate

385 survival outcome (median survival of 27.77-22.90 months respectively) between the two
386 other groups (Figure 11G; Table 8).

387 **External validation of gene signature risk score in TCGA database**

388 Using our risk calculation model, we calculated the risk score in TCGA dataset (n=179) and
389 similarly, patients were further sub-grouped as high-risk score (higher than median) and low
390 risk score (lower than median) based on the median value (43.434). Kaplan Meier analysis
391 showed that patients with a high-risk score (≥ 43.434) had inferior OS (HR 1.01; 95% CI
392 (1.00-1.02); $p < 0.019$) compared to those with a low-risk score (< 43.434) (Figure 11J; Table
393 5). Along with this, poor risk karyotype patients had worse overall survival (HR 1.89; 95%
394 CI (2.95-1.20); $p = 0.004$) compared to patients with good risk or intermediate risk karyotype
395 (7.03 vs 18.96 months). On multivariable analysis karyotype (poor vs good risk/intermediate
396 risk) and risk score were found to be independently predictive of ($p = 0.002$; $p = 0.025$
397 respectively) for worse OS. The timed AUC of risk score for 12-months and 18-months
398 survival in the TCGA dataset were 0.64 and 0.63 respectively (Figure 11K and 11L) and
399 Harrel's C-index of the prognostic model was 0.600.

400 **Discussion**

401 Our study is the first one to identify and validate mitochondria-related DEGs in paediatric
402 AML along with determining their prognostic significance. In paediatric AML patients, we
403 identified and validated 16 mitochondrial DEGs including 8 upregulated and 8
404 downregulated genes compared to controls. The dysregulated expression of these genes has
405 been previously reported in the pathogenesis of various malignancies (20–23). However, they
406 have not been studied in paediatric AML. Comparison with LAML dataset of TCGA cohort
407 suggests that the mitochondria-related gene expression profile in paediatric AML is likely
408 distinct. Elevated expression of genes like *SLC25A3*, *FASTKD1*, *SDHC*, *ATP5J*, which were
409 observed for the first time in our cohort, are involved in mitochondrial energy

410 metabolism(24–26). Genes like *FASLG*, *HRK* and *SNCA*, which were observed to be
411 downregulated, also play role in prevention of mitochondrial damage and apoptosis inhibition
412 in melanoma/medulloblastoma cell lines(27–29). Preliminary data suggests that
413 downregulation of genes like *MMP9* and *OLFM4*, as observed in our cohort, may aid in
414 AML progression(30,31). The expression of *CYP11B1* is reported to be elevated in various
415 malignancies, however, its expression is downregulated in early age leukaemia, as seen in our
416 cohort(32). These findings suggest that the observed mitochondria-related DEGs likely play
417 crucial role in disease progression in paediatric AML, which needs to be studied further
418 mechanistically.

419 Enhanced mtDNA copy number has been previously reported to be play role on AML
420 initiation, progression as well as predictive of survival outcomes(12,33). Similar to our
421 previous finding, we found that mtDNA copy number were significantly higher and
422 independently predictive of worse survival outcome in this cohort of pediatric AML patients.
423 These findings emphasise the importance of mtDNA copy number as a driver of disease
424 biology. Among the 16 validated mitochondria-related DEGs analysed, we observed that the
425 patients with higher mtDNA copy number had significantly higher expression of *SLC25A3*,
426 *SDHC*, *RACK1*, and *FASTKD1* compared to patients with low mtDNA copy number. While,
427 only a small percentage of mitochondrial proteins are coded by the mitochondrial genome,
428 variations in mtDNA may modulate molecular signals through nuclear-mitochondrial
429 crosstalk, which may promote tumorigenesis by upregulating oncogenes(34,35). This
430 suggests that in paediatric AML, cells with high mtDNA copy number are possibly driven
431 through unique gene expression alterations, influencing disease biology and therapeutic
432 response.

433 Comprehensive gene expression profiling has been extensively used to identify potential
434 prognostic genes in adult AML; however, dysregulation of mitochondria-related gene

435 expression, especially in children has not been well explored(36–38). Transcriptomic
436 profiling of cytogenetically normal paediatric AML has identified complex genomic
437 rearrangements and/or driver mutations in seemingly normal AML genomes and may even
438 aid risk stratification(39,40). Cai *et al.* developed a 3-gene prognostic risk model for children
439 with AML using NCI TARGET dataset, although it was not externally validated(36).
440 Similarly, Duployez *et al.* and Jiang *et al.* developed leukaemia stem cell score gene
441 signature and immune checkpoint related gene signature respectively in paediatric AML
442 predictive of survival outcomes(4,41). None of the above studies evaluated alterations in
443 mitochondrial gene expressions. Mitochondrial gene expression has been evaluated in other
444 malignancies like ovarian cancer, where a mitochondria-related gene signature, consisting of
445 8 metabolic genes, has been identified with independent prognostic impact(42).

446 In this study, we identified exclusive mitochondria-related DEGs in paediatric AML and
447 developed a prognostic gene signature including 3 genes (*SDHC*, *CLIC1*, and *SLC25A29*).
448 The gene signature risk score was additionally found to be independently predictive of
449 survival along with established ELN risk stratification with improved predictive ability over
450 clinical risk categorization. The risk score was also found to be associated with poor clinical
451 features of AML like the absence of RUNX1-RUNX1T1 translocation or poor-risk
452 cytogenetics. Hence, the gene signature model is able to categorize the heterogenous
453 molecular landscape of AML into clinically meaningful categories along with identification
454 of adverse disease biology. The developed prognostic score also has the potential to identify
455 high-risk subgroup even among those belonging to ELN good risk and vice-versa allowing
456 better upfront risk stratification and personalized treatment decisions.

457 TCGA LAML dataset has been extensively used for identifying as well as validating
458 prognostic gene signatures in various AML studies(43,44). We used the LAML dataset of
459 TCGA for external validation of our gene signature model and observed that the prognostic

460 gene signature score was also independently predictive of survival outcome in a large adult
461 cohort as well with predictive ability over and above known clinical predictors. This suggests
462 that the identified DEGs have a prognostic impact in AML across age group.

463 Our gene signature included 3 mitochondria-related genes i.e., *SDHC*, *CLIC1*, and
464 *SLC25A29*. *SDH* mutations lead to decreased activity of *SDH* with accumulation of succinate
465 and increase in oxidative stress resulting in DNA damage and tumorigenesis(45). In contrast
466 to previous findings, which suggests that the *SDH* gene is inactivated in solid tumors(46), we
467 observed an increased expression of *SDHC* gene in AML which was predictive of worse
468 survival. This is likely because, in contrast to solid malignancies, aggressive leukemias like
469 AML depend on cellular oxidative phosphorylation for proliferation which is supported by
470 upregulation of respiratory complex genes(47). Various studies also suggest dysregulation of
471 chloride ion channels such as the *CLIC1* gene which plays a role in drug resistance and
472 progression of various malignancies(22,48). Although, the role of *CLIC1* in AML is still
473 unexplored, we observed significant upregulation of *CLIC1* in paediatric AML with adverse
474 prognostic impact. The downstream effects of upregulation of *CLIC1* on disease biology of
475 AML need to be further deciphered.

476 In the current study, we observed an upregulation of *SLC25A29* in our cohort of paediatric
477 AML patients, which is in line with previous studies where it was found to be significantly
478 elevated in multiple malignancies(23). Similar upregulation was also been observed in adult
479 AML patients of TCGA LAML dataset. However, on survival analysis, downregulation of
480 *SLC25A29* was independently predictive of worse OS in our cohort. This finding was
481 consistent even in the external cohort of TCGA LAML dataset, where even though the
482 expression of *SLC25A29* was upregulated, a downregulated expression was predictive of
483 worse survival outcomes. This finding was intriguing and the mechanism by which
484 downregulation of *SLC25A29* drives a worse survival outcome remains unclear. *SLC25A29* is

485 the main arginine transporter in the mitochondrial membrane(49). Aberrant upregulation of
486 *SLC25A29* may result in transportation of more arginine into mitochondria, promoting
487 synthesis of metabolites like nitric oxide, polyamines, proline and creatine, which are
488 essential for cell survival and proliferation(50). Mitochondria-derived nitric oxide is known
489 to have a dichotomous role in regulation of cancer progression which is influenced by
490 expression of *SLC25A29* likely affecting disease outcome(51). The *SLC25* family of genes
491 which encodes for a set of mitochondrial inner membrane carrier proteins, have been
492 identified as a potential biomarker as well as novel therapeutic targets in various
493 malignancies(52). The implications of altered expression of *SLC25A29* on disease biology of
494 AML and its assessment as a therapeutic target is an exciting area of further research.

495 Our study has certain limitations. Transcriptomic profile and further validation by RTPCR
496 were done in whole isolated mononuclear cell and not in sorted blasts. Initial selection of
497 DEGs were also done from whole RNA sequencing of a limited number of samples, which
498 may lead to a bias in selection, however, external validation of the validated genes confirmed
499 their prognostic impact in an independent cohort.

500 In conclusion, this is the first study to report a validated set of mitochondria-related DEGs in
501 paediatric AML. We observed that patients with high mtDNA copy number have a unique
502 gene expression pattern possibly affecting disease biology. We developed a 3-gene based
503 mitochondrial gene signature model with ability to predict prognosis in paediatric AML
504 patients over and above established clinical prognostic parameters. The gene signature was
505 also externally validated in a cohort of adult AML patients demonstrating its predictive
506 ability in adult AML as well. Further directions for research include *in vitro* studies for
507 elucidating the role of prognostic genes in leukemogenesis and their evaluation as potential
508 targets for the treatment of paediatric AML.

509 **Data availability statement**

510 The sequencing data is publicly available in repository (NCBI-SRA at PRJNA778747) and
511 rest of data is available from the principal investigator on reasonable request.

512 **Acknowledgements**

513 The authors acknowledge the funding support from DST-SERB (Department of Science and
514 Technology - Science and Engineering Research Board), Government of India for this work.
515 (Extramural Research grant: EMR/2016/006376 and Core Research grant:
516 CRG/2021/001887). The authors acknowledge the funding support from ICMR (Indian
517 Council of Medical Research), Government of India for this work (ICMR SRF: 2019-
518 6059/CMB/BMS). The authors also acknowledge every member of paediatric oncology team
519 of our centre including research staff, nurses and dietician for their exemplary clinical
520 services.

521 **Authorship Contribution Statement**

522 Shilpi Chaudhary conceptualized the study, conducted the research, performed data analysis
523 and interpreted results and wrote the manuscript. Shuvadeep Ganguly analysed data,
524 interpreted results and wrote the manuscript. Jayanth Kumar Palanichamy, Archana Singh,
525 Radhika Bakhshi and Anita Chopra conceptualized the study, provided intellectual inputs,
526 administrative support and edited the manuscript. Dibyabhaba Pradhan conducted
527 transcriptome data analysis. Sameer Bakhshi conceptualized the study, provided
528 administrative support, intellectual inputs, interpreted results, wrote and edited the
529 manuscript. All authors have reviewed and approved the final version of the manuscript.

530 **Conflict of interest declaration statement**

531 All authors declare no potential conflicts of interest.

532

533 **References**

- 534 1. Lonetti A, Pession A, Masetti R. Targeted Therapies for Pediatric AML: Gaps and
535 Perspective [Internet]. Vol. 7, *Frontiers in Pediatrics*. Frontiers Media S.A.; 2019
536 [cited 2020 Nov 27]. p. 463. Available from: www.frontiersin.org
- 537 2. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al.
538 Diagnosis and management of AML in adults: 2017 ELN recommendations from an
539 international expert panel [Internet]. Vol. 129, *Blood*. American Society of
540 Hematology; 2017 [cited 2021 Apr 10]. p. 424–47. Available from:
541 <http://ashpublications.org/blood/article-pdf/129/4/424/1402046/blood733196.pdf>
- 542 3. Elsayed AH, Rafiee R, Cao X, Raimondi S, Downing JR, Ribeiro R, et al. A six-gene
543 leukemic stem cell score identifies high risk pediatric acute myeloid leukemia.
544 *Leukemia* [Internet]. 2020 Mar 1 [cited 2021 Feb 1];34(3):735–45. Available from:
545 <https://doi.org/10.1038/s41375-019-0604-8>
- 546 4. Duployez N, Marceau-Renaut A, Villenet C, Petit A, Rousseau A, Ng SWK, et al.
547 Evaluation of gene expression signatures predictive of cytogenetic and molecular
548 subtypes of pediatric acute myeloid leukemia. *Leukemia* [Internet]. 2019 Feb 1 [cited
549 2021 Aug 3];33(2):348–57. Available from:
550 <https://haematologica.org/article/view/5869>
- 551 5. Wagner S, Vadakekolathu J, Tasian SK, Altmann H, Bornhäuser M, Pockley AG, et al.
552 A parsimonious 3-gene signature predicts clinical outcomes in an acute myeloid
553 leukemia multicohort study. *Blood Adv* [Internet]. 2019 Apr 23 [cited 2021 Aug
554 5];3(8):1330–46. Available from: <http://www.vizome.org/aml/>
- 555 6. Weinberg SE, Chandel NS. Targeting mitochondria metabolism for cancer therapy
556 [Internet]. Vol. 11, *Nature Chemical Biology*. Nature Publishing Group; 2015 [cited
557 2020 Aug 27]. p. 9–15. Available from: [/pmc/articles/PMC4340667/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/3440667/)

- 558 7. Panina SB, Baran N, Brasil da Costa FH, Konopleva M, Kirienko N V. A mechanism
559 for increased sensitivity of acute myeloid leukemia to mitotoxic drugs. *Cell Death Dis*
560 [Internet]. 2019 Aug 1 [cited 2020 Aug 13];10(8):1–15. Available from:
561 <https://doi.org/10.1038/s41419-019-1851-3>
- 562 8. Al Ageeli E. Alterations of mitochondria and related metabolic pathways in leukemia:
563 A narrative review. *Saudi J Med Med Sci* [Internet]. 2020 [cited 2020 Aug 16];8(1):3.
564 Available from: </pmc/articles/PMC6945320/?report=abstract>
- 565 9. Tyagi A, Pramanik R, Bakhshi R, Singh A, Vishnubhatla S, Bakhshi S. Expression of
566 mitochondrial genes predicts survival in pediatric acute myeloid leukemia. *Int J*
567 *Hematol* [Internet]. 2019 Aug 5 [cited 2020 Jul 21];110(2):205–12. Available from:
568 <https://doi.org/10.1007/s12185-019-02666-2>
- 569 10. Tyagi A, Pramanik R, Vishnubhatla S, Bakhshi R, Bakhshi S. Prognostic impact of
570 mitochondrial DNA D-loop variations in pediatric acute myeloid leukemia. *Oncotarget*
571 [Internet]. 2019 Feb 1 [cited 2020 Aug 14];10(13):1334–43. Available from:
572 </pmc/articles/PMC6407682/?report=abstract>
- 573 11. Sharawat SK, Bakhshi R, Vishnubhatla S, Bakhshi S. Mitochondrial D-loop variations
574 in paediatric acute myeloid leukaemia: A potential prognostic marker. *Br J Haematol*
575 [Internet]. 2010 May 1 [cited 2021 Apr 12];149(3):391–8. Available from:
576 <http://doi.wiley.com/10.1111/j.1365-2141.2010.08084.x>
- 577 12. Chaudhary S, Ganguly S, Palanichamy JK, Singh A, Bakhshi R, Jain A, et al. PGC1A
578 driven enhanced mitochondrial DNA copy number predicts outcome in pediatric acute
579 myeloid leukemia. *Mitochondrion* [Internet]. 2021 May 1 [cited 2021 Apr 14];58:246–
580 54. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1567724921000453>
- 581 13. Arora S, Pushpam D, Tiwari A, Choudhary P, Chopra A, Gupta R, et al. Allogeneic
582 hematopoietic stem cell transplant in pediatric acute myeloid leukemia: Lessons learnt

- 583 from a tertiary care center in India. *Pediatr Transplant* [Internet]. 2020 [cited 2021 Apr
584 12]; Available from: <https://pubmed.ncbi.nlm.nih.gov/33142026/>
- 585 14. Tyagi A, Pramanik R, Chaudhary S, Chopra A, Bakhshi S. Cytogenetic Profiles of 472
586 Indian Children with Acute Myeloid Leukemia. *Indian Pediatr*. 2018;55(6):469–73.
- 587 15. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al.
588 STRING v9.1: protein-protein interaction networks, with increased coverage and
589 integration. *Nucleic Acids Res*. 2012 Nov;41(D1):D808–15.
- 590 16. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: Identifying hub
591 objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8(4):S11.
- 592 17. Bader GD, Hogue CWV. An automated method for finding molecular complexes in
593 large protein interaction networks. *BMC Bioinformatics*. 2003 Jan;4:2.
- 594 18. Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O’Donoghue SI, Schneider R, et
595 al. COMPARTMENTS: Unification and visualization of protein subcellular
596 localization evidence. *Database* [Internet]. 2014 [cited 2021 Jun 3];2014. Available
597 from: [/pmc/articles/PMC3935310/](https://pubmed.ncbi.nlm.nih.gov/24811110/)
- 598 19. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A web server for cancer and
599 normal gene expression profiling and interactive analyses. *Nucleic Acids Res*
600 [Internet]. 2017 Jul 3 [cited 2021 Jun 4];45(W1):W98–102. Available from:
601 <https://pubmed.ncbi.nlm.nih.gov/28407145/>
- 602 20. Li J-J, Xie D. RACK1, a versatile hub in cancer. *Oncogene* 2015 3415 [Internet]. 2014
603 Jun 2 [cited 2021 Jul 12];34(15):1890–8. Available from:
604 <https://www.nature.com/articles/onc2014127>
- 605 21. Oehler VG, Ka YY, Choi YE, Bumgarner RE, Raftery AE, Radich JP. The derivation
606 of diagnostic markers of chronic myeloid leukemia progression from microarray data.
607 *Blood*. 2009;114(15):3292–8.

- 608 22. Lee JR, Lee JY, Kim HJ, Hahn MJ, Kang JS, Cho H. The inhibition of chloride
609 intracellular channel 1 enhances Ca²⁺ and reactive oxygen species signaling in A549
610 human lung cancer cells. *Exp Mol Med* [Internet]. 2019 Jul 1 [cited 2020 Sep
611 25];51(7):1–11. Available from: <https://doi.org/10.1038/s12276-019-0279-2>
- 612 23. Zhang H, Wang Q, Gu J, Yin L, Liang S, Wu L, et al. Elevated mitochondrial
613 SLC25A29 in cancer modulates metabolic status by increasing mitochondria-derived
614 nitric oxide. *Oncogene*. 2018 May;37(19):2545–58.
- 615 24. Kolbe HVJ, Costello D, Wong A. Mitochondrial phosphate transport. Large scale
616 isolation and characterization of the phosphate transport protein from beef heart
617 mitochondria. *J Biol Chem* [Internet]. 1984 [cited 2021 Jul 9];259(14):9115–20.
618 Available from: [https://www.jbc.org/article/S0021-9258\(17\)47273-5/abstract](https://www.jbc.org/article/S0021-9258(17)47273-5/abstract)
- 619 25. Nazar E, Khatami F, Saffar H, Tavangar SM. The Emerging Role of Succinate
620 Dehydrogenase Genes (SDHx) in Tumorigenesis. *Int J Hematol Stem Cell Res*
621 [Internet]. 2019 [cited 2021 Jul 12];13(2):72. Available from:
622 </pmc/articles/PMC6660475/>
- 623 26. AA J, J P, MA de la F, JC M, P A, M S. The FASTK family of proteins: emerging
624 regulators of mitochondrial RNA biology. *Nucleic Acids Res* [Internet]. 2017 Nov 2
625 [cited 2021 Jul 12];45(19):10941–7. Available from:
626 <https://pubmed.ncbi.nlm.nih.gov/29036396/>
- 627 27. M C, W W, J M, P Y, K W. High glucose induces mitochondrial dysfunction and
628 apoptosis in human retinal pigment epithelium cells via promoting SOCS1 and
629 Fas/FasL signaling. *Cytokine* [Internet]. 2016 Feb 1 [cited 2021 Jul 13];78:94–102.
630 Available from: <https://pubmed.ncbi.nlm.nih.gov/26700587/>
- 631 28. Li Y, Yu Z, Jiang T, Shao L, Liu Y, Li N, et al. SNCA, a novel biomarker for Group 4
632 medulloblastomas, can inhibit tumor invasion and induce apoptosis. *Cancer Sci*

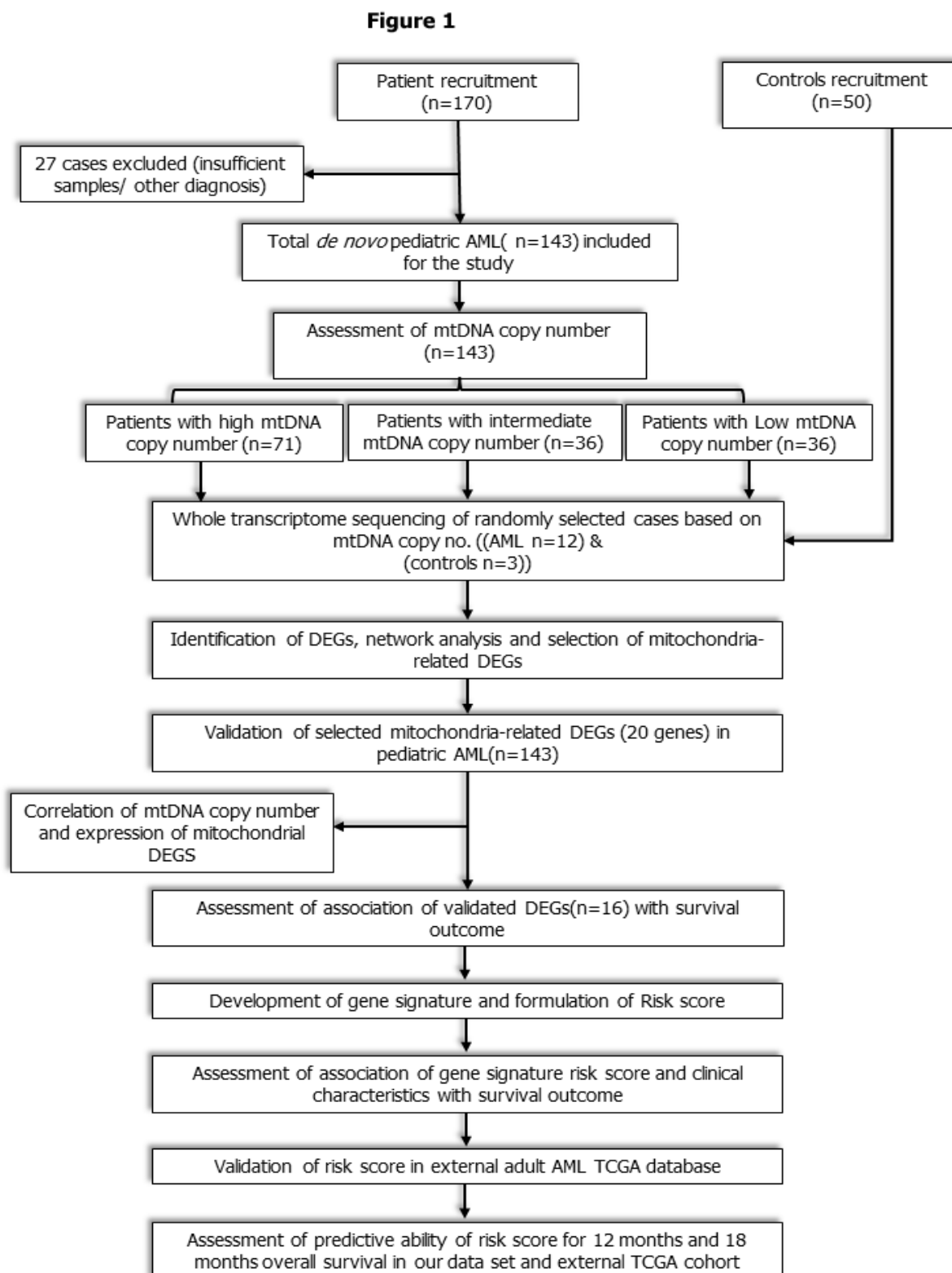
- 633 [Internet]. 2018 Apr 1 [cited 2021 Jul 13];109(4):1263. Available from:
634 /pmc/articles/PMC5891175/
- 635 29. Rizvi F, Heimann T, Herrnreiter A, O'Brien WJ. Mitochondrial Dysfunction Links
636 Ceramide Activated HRK Expression and Cell Death. PLoS One [Internet]. 2011
637 [cited 2021 Jul 13];6(3):e18137. Available from:
638 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018137>
- 639 30. Lin LI, Lin DT, Chang CJ, Lee CY, Tang JL, Tien HF. Marrow matrix
640 metalloproteinases (MMPS) and tissue inhibitors of MMP in acute leukaemia:
641 Potential role of MMP-9 as a surrogate marker to monitor leukaemic status in patients
642 with acute myelogenous leukaemia. Br J Haematol [Internet]. 2002 Jun 1 [cited 2021
643 Jan 25];117(4):835–41. Available from: [http://doi.wiley.com/10.1046/j.1365-](http://doi.wiley.com/10.1046/j.1365-2141.2002.03510.x)
644 [2141.2002.03510.x](http://doi.wiley.com/10.1046/j.1365-2141.2002.03510.x)
- 645 31. Liu W, Lee HW, Liu Y, Wang R, Rodgers GP. Olfactomedin 4 is a novel target gene
646 of retinoic acids and 5-aza-2'-deoxycytidine involved in human myeloid leukemia cell
647 growth, differentiation, and apoptosis. Blood [Internet]. 2010 Dec 2 [cited 2021 Jul
648 13];116(23):4938–47. Available from: [https://www.meta.org/papers/olfactomedin-4-](https://www.meta.org/papers/olfactomedin-4-is-a-novel-target-gene-of-retinoic/20724538)
649 [is-a-novel-target-gene-of-retinoic/20724538](https://www.meta.org/papers/olfactomedin-4-is-a-novel-target-gene-of-retinoic/20724538)
- 650 32. Lopes BA, Emerenciano M, Gonçalves BAA, Vieira TM, Rossini A, Pombo-de-
651 Oliveira MS. Polymorphisms in CYP1B1, CYP3A5, GSTT1, and SULT1A1 are
652 associated with early age acute leukemia. Kanungo J, editor. PLoS One [Internet].
653 2015 May 18 [cited 2021 Feb 1];10(5):e0127308. Available from:
654 <https://dx.plos.org/10.1371/journal.pone.0127308>
- 655 33. Kang MG, Kim YN, Lee JH, Szardenings M, Baek HJ, Kook H, et al.
656 Clinicopathological implications of mitochondrial genome alterations in pediatric
657 acute myeloid leukemia. Ann Lab Med [Internet]. 2016 Mar 1 [cited 2020 Jul

- 658 23];36(2):101–10. Available from: <https://pubmed.ncbi.nlm.nih.gov/26709256/>
- 659 34. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, et al. A
660 Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell*.
661 2008 Jul 11;134(1):112–23.
- 662 35. Horan MP, Cooper DN. The emergence of the mitochondrial genome as a partial
663 regulator of nuclear function is providing new insights into the genetic mechanisms
664 underlying age-related complex disease [Internet]. Vol. 133, *Human Genetics. Hum*
665 *Genet*; 2014 [cited 2021 Jul 9]. p. 435–58. Available from:
666 <https://pubmed.ncbi.nlm.nih.gov/24305784/>
- 667 36. Cai Z, Wu Y, Zhang F, Wu H. A three-gene signature and clinical outcome in pediatric
668 acute myeloid leukemia. *Clin Transl Oncol* [Internet]. 2020 Aug 29 [cited 2021 Feb 1];
669 Available from: <http://link.springer.com/10.1007/s12094-020-02480-x>
- 670 37. Duployez N, Marceau-Renaut A, Villenet C, Petit A, Rousseau A, Ng SWK, et al. The
671 stem cell-associated gene expression signature allows risk stratification in pediatric
672 acute myeloid leukemia. *Leukemia* [Internet]. 2019 Feb 1 [cited 2021 Feb
673 1];33(2):348–57. Available from: <https://doi.org/10.1038/s41375-018-0227-5>
- 674 38. Nguyen CH, Glüxam T, Schlerka A, Bauer K, Grandits AM, Hackl H, et al. SOCS2 is
675 part of a highly prognostic 4-gene signature in AML and promotes disease
676 aggressiveness. *Sci Rep* [Internet]. 2019 Dec 1 [cited 2021 Feb 1];9(1):1–13. Available
677 from: <https://doi.org/10.1038/s41598-019-45579-0>
- 678 39. Shiba N, Yoshida K, Hara Y, Yamato G, Shiraishi Y, Matsuo H, et al. Transcriptome
679 analysis offers a comprehensive illustration of the genetic background of pediatric
680 acute myeloid leukemia. *Blood Adv* [Internet]. 2019 Oct 22 [cited 2021 Aug
681 16];3(20):3157–69. Available from: [/pmc/articles/PMC6849955/](https://pubmed.ncbi.nlm.nih.gov/34449955/)
- 682 40. Balgobind B V., van den Heuvel-Eibrink MM, De Menezes RX, Reinhardt D, Hollink

- 683 IHI, Arentsen-Peters STJCM, et al. Evaluation of gene expression signatures
684 predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia.
685 *Haematologica*. 2011 Feb;96(2):221–30.
- 686 41. Jiang F, Wang XY, Wang MY, Mao Y, Miao XL, Wu CY, et al. An Immune
687 Checkpoint-Related Gene Signature for Predicting Survival of Pediatric Acute
688 Myeloid Leukemia. *J Oncol*. 2021;2021.
- 689 42. Wang L, Li X. Identification of an energy metabolism-related gene signature in
690 ovarian cancer prognosis. *Oncol Rep [Internet]*. 2020 Jun 1 [cited 2021 Feb
691 2];43(6):1755–70. Available from: <http://www.affymetrix>.
- 692 43. Zhang Y, Ma S, Wang M, Shi W, Hu Y. Comprehensive Analysis of Prognostic
693 Markers for Acute Myeloid Leukemia Based on Four Metabolic Genes. *Front Oncol*.
694 2020 Sep 29;0:1990.
- 695 44. Huang R, Liao X, Li Q. Identification and validation of potential prognostic gene
696 biomarkers for predicting survival in patients with acute myeloid leukemia. *Onco*
697 *Targets Ther [Internet]*. 2017 Nov 2 [cited 2021 Aug 3];10:5243. Available from:
698 </pmc/articles/PMC5679677/>
- 699 45. Zhao T, Mu X, You Q. Succinate: An initiator in tumorigenesis and progression
700 [Internet]. Vol. 8, *Oncotarget*. Impact Journals, LLC; 2017 [cited 2021 Aug 16]. p.
701 53819–28. Available from: </pmc/articles/PMC5581152/>
- 702 46. Li J, Liang N, Long X, Zhao J, Yang J, Du X, et al. SDHC-related deficiency of SDH
703 complex activity promotes growth and metastasis of hepatocellular carcinoma via
704 ROS/NFκB signaling. *Cancer Lett*. 2019 Oct;461:44–55.
- 705 47. Panina SB, Pei J, Kirienko N V. Mitochondrial metabolism as a target for acute
706 myeloid leukemia treatment [Internet]. Vol. 9, *Cancer & Metabolism*. Springer Science
707 and Business Media LLC; 2021 [cited 2021 May 26]. p. 1–25. Available from:

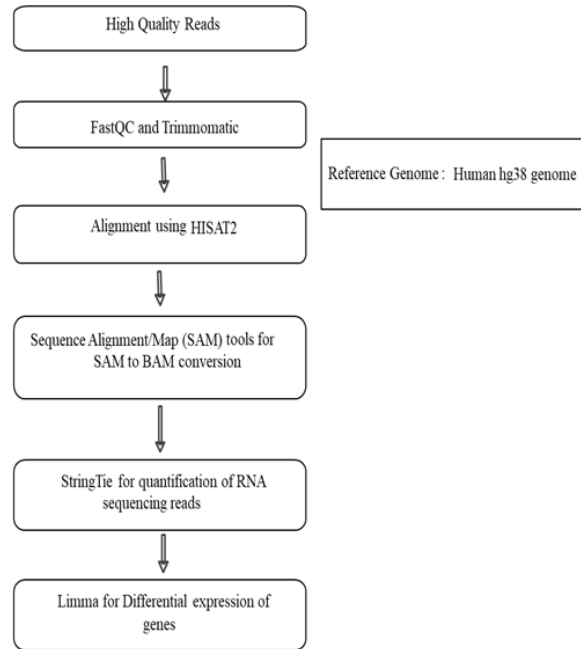
- 708 <https://doi.org/10.1186/s40170-021-00253-w>
- 709 48. He YM, Zhang ZL, Liu QY, Xiao YS, Wei L, Xi C, et al. Effect of CLIC1 gene
710 silencing on proliferation, migration, invasion and apoptosis of human gallbladder
711 cancer cells. *J Cell Mol Med*. 2018 May;22(5):2569–79.
- 712 49. Porcelli V, Fiermonte G, Longo A, Palmieri F. The human gene SLC25A29, of solute
713 carrier family 25, encodes a mitochondrial transporter of basic amino acids. *J Biol*
714 *Chem*. 2014 May;289(19):13374–84.
- 715 50. Keshet R, Erez A. Arginine and the metabolic regulation of nitric oxide synthesis in
716 cancer [Internet]. Vol. 11, *DMM Disease Models and Mechanisms*. Company of
717 Biologists; 2018 [cited 2021 Aug 19]. Available from: [/pmc/articles/PMC6124554/](https://pubmed.ncbi.nlm.nih.gov/3124554/)
- 718 51. Ghafourifar P, Cadenas E. Mitochondrial nitric oxide synthase. Vol. 26, *Trends in*
719 *Pharmacological Sciences*. Elsevier Ltd; 2005. p. 190–5.
- 720 52. Rochette L, Meloux A, Zeller M, Malka G, Cottin Y, Vergely C. Mitochondrial
721 SLC25 carriers: Novel targets for cancer therapy. Vol. 25, *Molecules*. MDPI AG;
722 2020.
- 723
- 724

725 **Figures**



726
727 **Figure 1. Workflow of the study:** Study workflow showing flow of patients from enrolment to
728 RNA sequencing, identification of mitochondria- related DEGs, development and validation of novel
729 3-gene based Risk score.

Figure 2



730

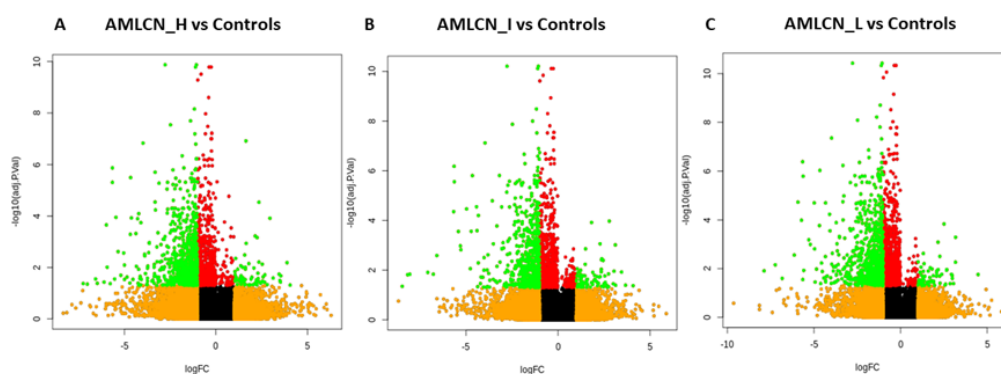
731 **Figure 2:** Workflow of RNA sequencing data analysis using various bioinformatic tools

732

733

734

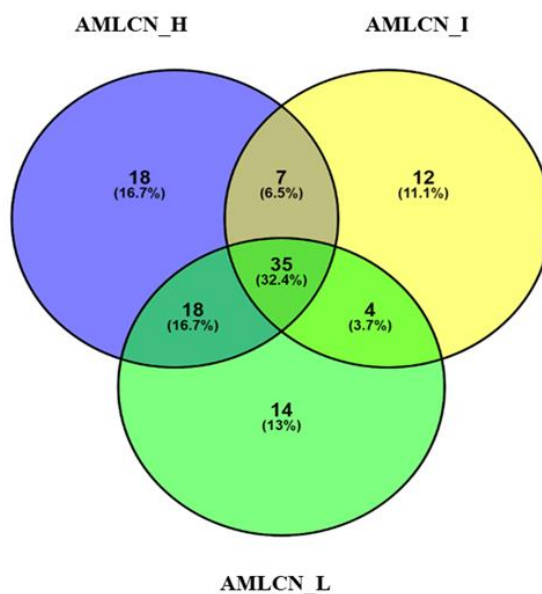
Figure 3



735
736

737 **Figure 3:** Volcano plots representing all dysregulated transcripts among AMLCN_H (A),
738 AMLCN_I (B) and AMLCN_L (C) groups compared to controls. Green dots represent
739 significantly differentially expressed transcripts in the three subgroups of AML compared to
740 controls. High AMLCN_H (more than 75th percentile); intermediate AMLCN_I (50th to 75th
741 percentile) and low AMLCN_L (lower than 50th percentile) relative mitochondrial DNA copy number
742

Figure 4



743

744

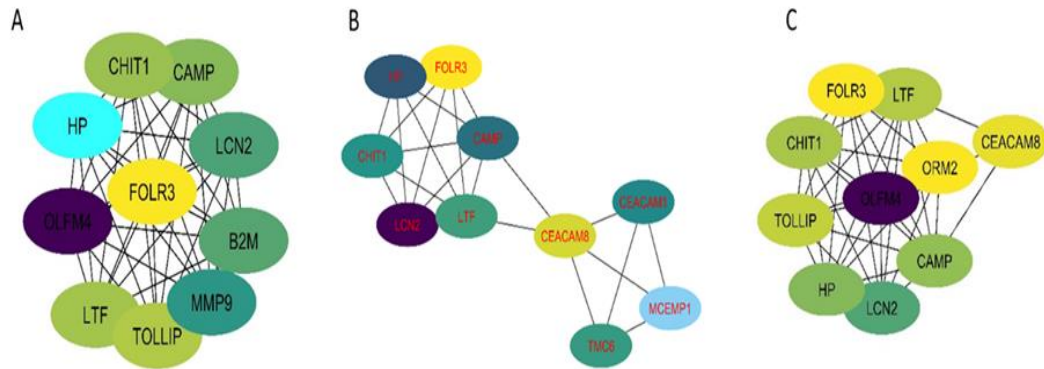
745 **Figure 4:** Venn diagram representing mitochondrial associated differentially expressed genes (DEGs)
746 among three groups of AML (AMLCN_H; AMLCN_I and AMLCN_L). High AMLCN_H (more
747 than 75th percentile); intermediate AMLCN_I (50th to 75th percentile) and low AMLCN_L (lower than
748 50th percentile) relative mitochondrial DNA copy number

749

750

751

Figure 5



752

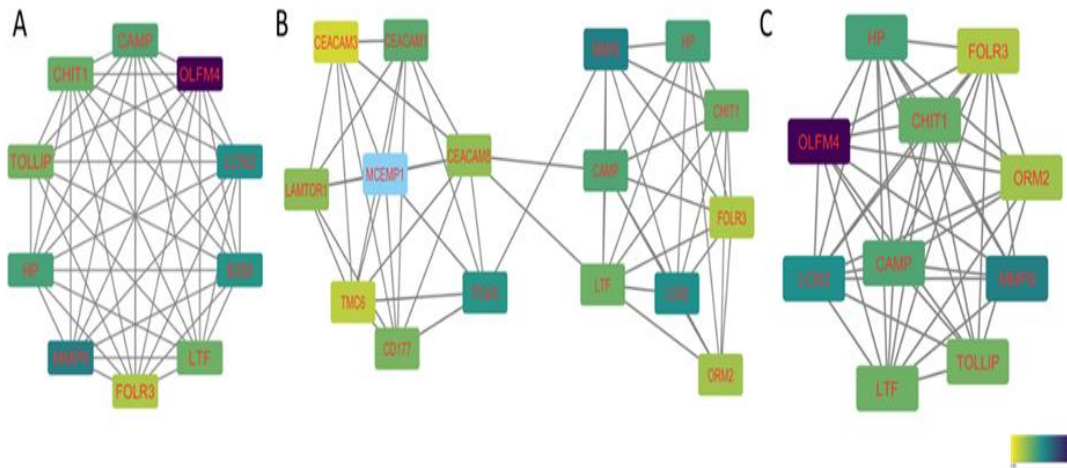
753

754 **Figure 5:** Hub gene based on Maximal Clique Centrality (MCC) (A): AMLCN_H; (B) AMLCN_I;
755 (C) AMLCN_L. High AMLCN_H (more than 75th percentile); intermediate AMLCN_I (50th to 75th
756 percentile) and low AMLCN_L (lower than 50th percentile) relative mitochondrial DNA copy number

757

758

Figure 6

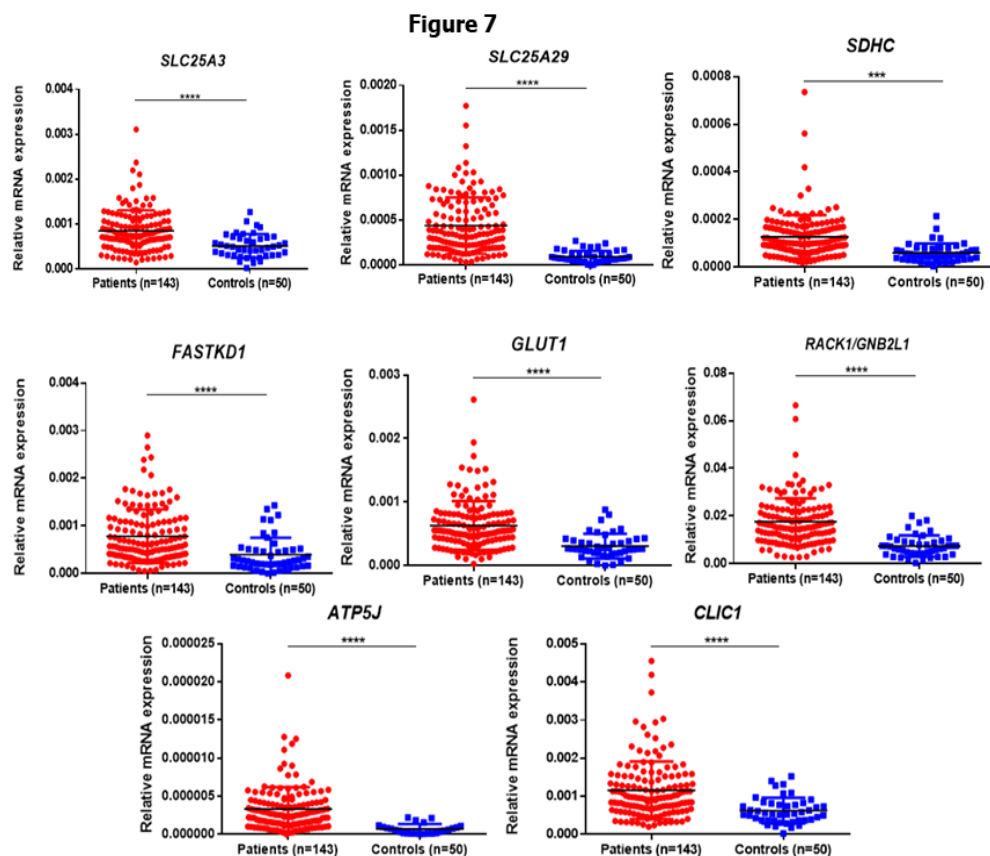


759

760 **Figure 6:** MCODE (molecular complex detection) analysis showing most interactive clusters
761 among A: (A): AMLCN_H; (B) AMLCN_I; (C) AMLCN_L. High AMLCN_H (more than 75th
762 percentile); intermediate AMLCN_I (50th to 75th percentile) and low AMLCN_L (lower than 50th
763 percentile) relative mitochondrial DNA copy number

764

765



766

767

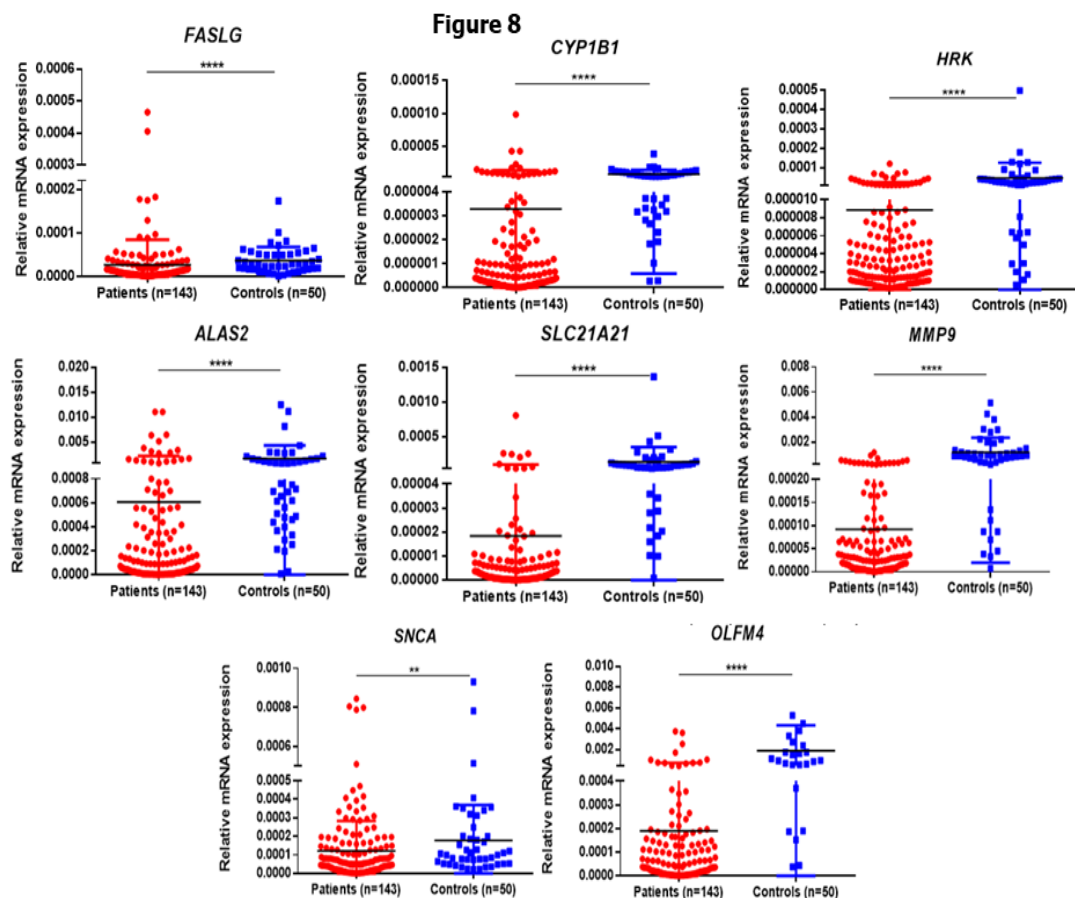
768 **Figure 7:** Validation of selected upregulated differentially expressed genes (DEGs) in patients as
769 compared to controls. *SLC25A3*, *SLC25A29*, *SDHC*, *FASTKD1*, *GLUT1*, *RACK1*, *ATP5J* and *CLIC1*
770 were significantly upregulated in pediatric AML patients (n=143) compared to controls (n=50). *:
771 P<0.05; **: P< 0.01; ***: P<0.001; ****: P<0.0001.

772

773

774

775



776

777

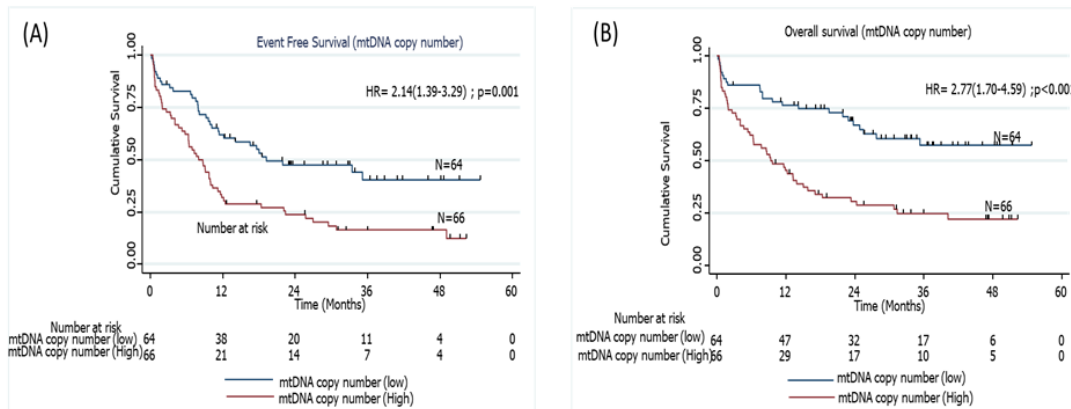
778 **Figure 8:** Validation of selected downregulated differentially expressed genes (DEGs) in patients as
779 compared to controls. *FASLG*, *CYP1B1*, *HRK*, *ALAS2*, *SLC25A21*, *MMP9*, *SNCA* and *OLFM4* were
780 significantly downregulated in pediatric AML patients (n=143) compared to controls (n=50) *:
781 P<0.05; **: P< 0.01; ***: P<0.001; ****: P<0.0001

782

783

784

Figure 9

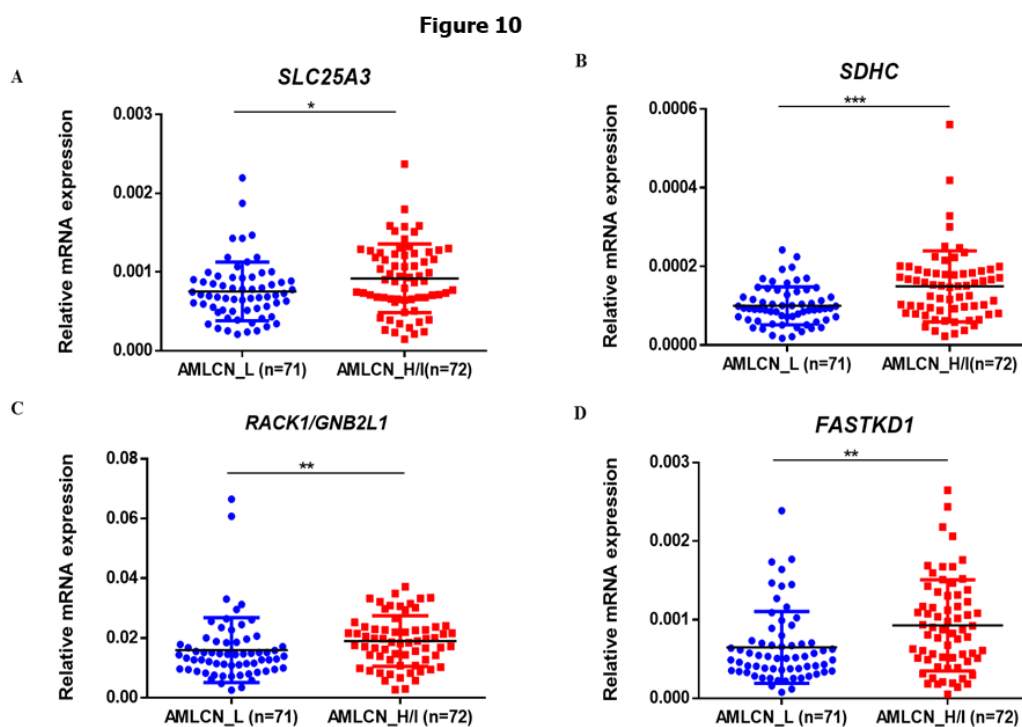


785

786 **Figure 9:** Kaplan Meier curves representing association of mtDNA copy number with (A) event free
 787 survival and (B) overall survival of pediatric AML patients

788

789

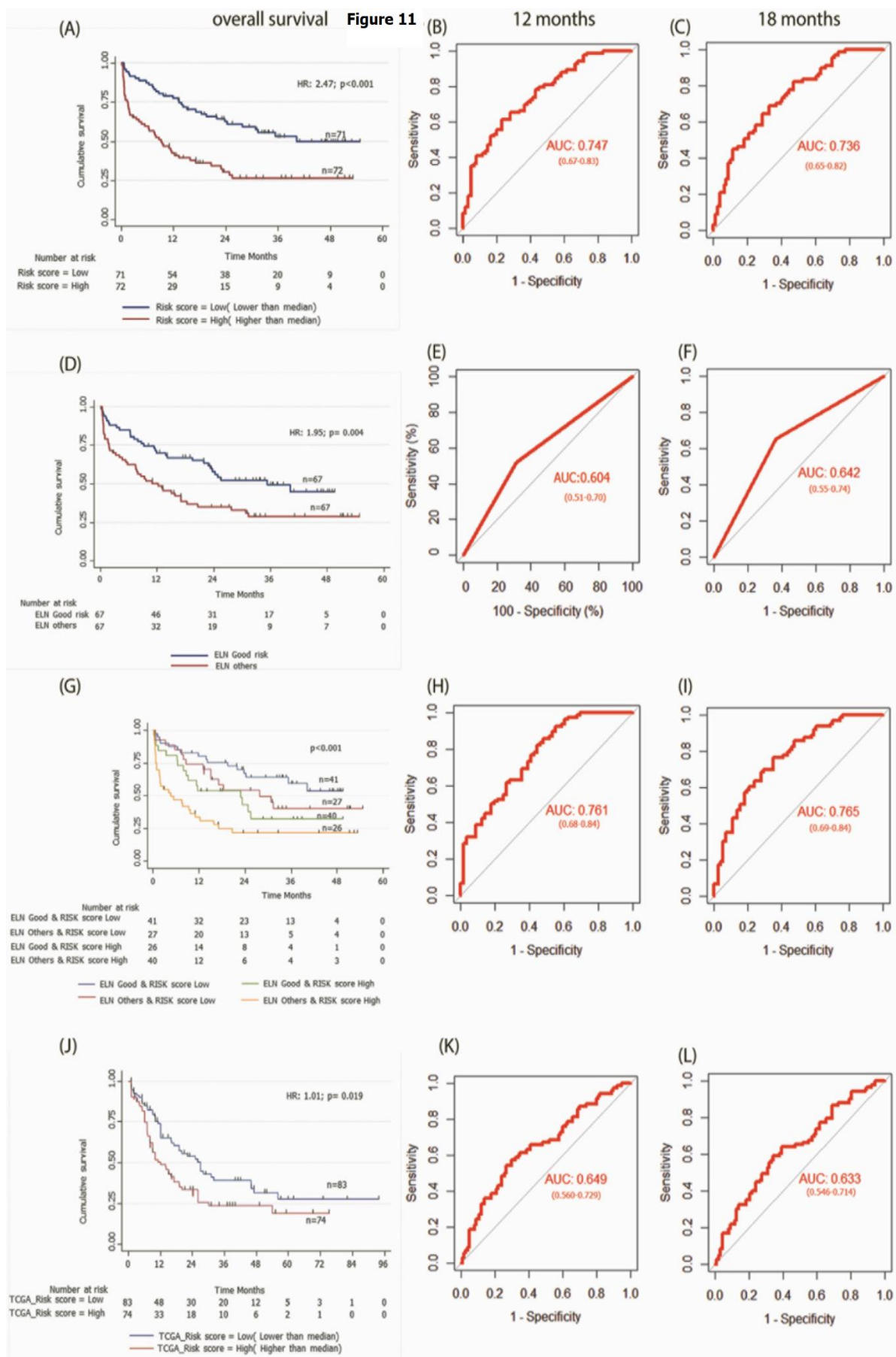


790

791 **Figure 10:** Subgroup analysis of acute myeloid leukemia showing exclusive deregulation of validated
792 genes among AMLCN_L and AMLCN_H group. The expression of *SLC25A3*(A), *SDHC*(B),
793 *RACK1*(C) and *FASTKD1*(D) were significantly higher in AMLCN_H group as compared to
794 AMLCN_L

795

796



798 **Figure 11:** Kaplan Meier curves representing association of mtDNA copy number with (A) event free
799 survival and (B) overall survival of pediatric AML patients **Figure 11.A 3-gene based gene**
800 **signature stratifies survival in pediatric and adult AML patients along with clinically**
801 **established European LeukemiaNet (ELN) risk categories.** (A) Kaplan Meier estimates of overall
802 survival in pediatric AML patient's subgroup into high Risk-score and low Risk-score. (B) and (C)
803 AUC curves quantify the ability of our 3-gene based risk score to predict outcome in individual
804 patients (specificity and sensitivity) within the first 12 months(B) and 18 months(C) of treatment
805 initiation respectively. (D) Kaplan Meier estimates of overall survival in pediatric AML patient's
806 subgroup into ELN good risk and ELN intermediate or poor risk categories. (E) and (F) AUC curves
807 quantify the ability of ELN risk categories to predict outcome in individual patients (specificity and
808 sensitivity) within the first 12 months (E) and 18 months(F) of treatment initiation respectively. (G)
809 Kaplan Meier estimates of overall survival in pediatric AML patient's subgroup by combining ELN
810 risk categories with our 3 gene-based risk score. (H) and (I) AUC curves quantify the ability of
811 combined model of ELN risk categories and our 3 gene-based risk score to predict outcome in
812 individual patients (specificity and sensitivity) within the first 12 months(H) and 18 months(I) of
813 treatment initiation respectively. (J) Kaplan Meier estimates of overall survival in external adult The
814 Cancer Genome Atlas (TCGA) AML patient's subgroup into high Risk-score and low Risk-score
815 using our 3 gene-based gene signature model. (K) and (L) AUC curves quantify the ability of our 3-
816 gene based risk score to predict outcome in individual patients of TCGA adult AML datasets
817 (specificity and sensitivity) within the first 12 months(K) and 18 months(L) of treatment initiation
818 respectively. AUC = 1.0 would denote perfect prediction, and AUC = 0.5 would denote no predictive
819 ability.

820

821

822

823

824

825

826

827

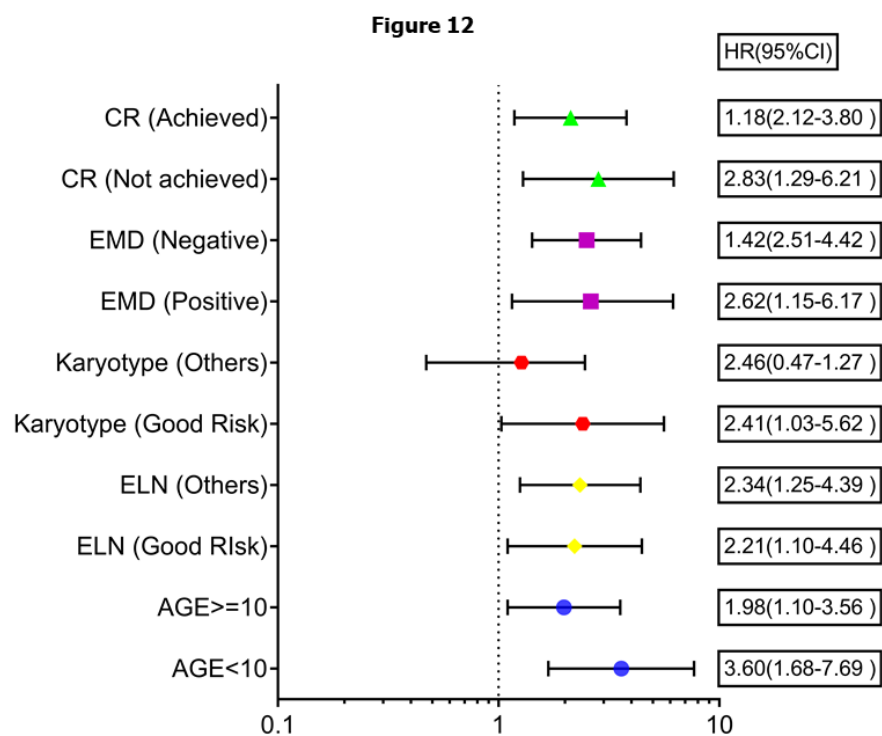
828

829

830

831

832



833

834 **Figure 12:** Forest plot showing the impact of mitochondrial prognostic gene signature risk score
835 category on survival outcome in various clinically relevant subgroups of pediatric Acute myeloid
836 leukemia

837

838

839

840

841 **Table 1: Baseline characteristic features of pediatric AML patients (n=143)**

Characteristics	Number (%)
Median age (years)	10 (0.8-18)
Sex	
Male	87 (60.8)
Female	56 (39.2)
Haematological parameters	
Median haemoglobin, (g/dL)	7.4 (2.1-14.6)
Median total leucocyte count, ($\times 10^3/\mu\text{L}$)	24.38 (0.76-314.3)
Hyperleukocytosis, ($>50 \times 10^3/\mu\text{L}$)	45 (31.46)
Median platelet count, ($\times 10^3/\mu\text{L}$)	35.00(4-276)
Clinical features at presentation	
Fever	112(78.3)
Chloroma	27 (18.9)
Cytogenetics (n=130)	
Normal	36 (27.5)
t(8;21) & inv (16)*	52 (39.7)
Complex karyotype	8 (6.1)
Other	19 (14.5)
Failed cytogenetics	16(12.2)
Molecular analysis (n= 122)	
<i>FLT3 ITD</i>	17 (13.8)
<i>RUNX1-RUNX1T1</i>	53 (43.1)
<i>CBFB-MYH11</i>	6 (4.9)
<i>NPM1</i>	5 (4.1)
Negative	41 (33.6)
ELN risk stratification** (n=134)	
Good	67 (50)
Intermediate	37 (27.6)
Poor	30 (22.4)
Complete remission (n=143)	
Achieved	104 (72.7%)
Not achieved	39 (27.3%)

842 * t(8;21)= t(8;21)(q22;q22) *RUNX1-RUNX1T1*; inv(16)= inv(16)(p13.1;q22) *CBFB-MYH11*

843 **ELN risk stratification was done using both cytogenetics and molecular markers in 134 patients. However,
844 for 12 patients risk stratification was done with only cytogenetics and in 20 patients, it was done by only
845 molecular analysis.

846 Median values were reported with range

847 n: number of patients; AML: Acute Myeloid Leukemia; Normal cytogenetics: 46,XX/46,XY; Failed
848 cytogenetics: metaphases could not be isolated; *FLT3 ITD*: FMS-like tyrosine kinase internal tandem
849 duplication; *RUNX1-RUNX1T1*: runt-related transcription factor 1-RUNX1 partner transcriptional co-repressor
850 1 fusion transcript; *CBFB-MYH11*: core binding factor beta-myosin heavy chain 11 fusion transcript; *NPM1*:
851 Nucleophosmin 1; ELN: European LeukemiaNet.

852

853

854 **Table 2: List of genes selected for validation among all the groups based on mitochondrial**
 855 **compartment score, *Cytohubba* and *MCODE***

S no.	Gene name	Compartment mitochondria score (0-5)	Groupwise expression	Expression in RNA seq data
1	<i>SLC25A3</i>	5	AMLCN_H	Upregulation
2	<i>LONP1</i>	5	AMLCN_H	Downregulation
3	<i>SDHC</i>	5	AMLCN_H	Upregulation
4	<i>GNB2L1/RACK1</i>	4.2	AMLCN_H	Upregulation
5	<i>FASTKD1</i>	5	AMLCN_I	Upregulation
6	<i>MRPL51</i>	5	AMLCN_I	Downregulation
7	<i>ATP5J</i>	5	AMLCN_I	Upregulation
8	<i>FASLG</i>	2.64	AMLCN_L	Downregulation
9	<i>CLIC1</i>	3.12	AMLCN_L	Upregulation
10	<i>HRK</i>	5	Common in all AML samples	Downregulation
11	<i>ALAS2</i>	5	Common in all AML samples	Downregulation
12	<i>SLC25A21</i>	5	Common in all AML samples	Downregulation
13	<i>CYP11B1</i>	4.35	Common in all AML samples	Downregulation
14	<i>GLUD1</i>	5	Common in AMLCN_H & AMLCN_I	Upregulation
15	<i>SLC25A29</i>	5	Common in AMLCN_H & AMLCN_I	Upregulation
16	<i>LIG1</i>	3.6	Common in AMLCN_H & AMLCN_I	Downregulation
17	<i>SNCA</i>	4.6	Common in AMLCN_L & AMLCN_H	Downregulation
18	<i>DHFR</i>	4.7	Common in AMLCN_L & AMLCN_H	Downregulation
19	<i>MMP9</i>	2.8	Hub gene and Seed gene in MCODE cluster of AMLCN_H & AMLCN_I	Downregulation
20	<i>OLFM4</i>	5	MCODE cluster gene	Downregulation

856 All the genes selected for validation are listed in the table. Selection of genes were based on their mitochondrial
 857 compartment score as determined by *Cytoscape*. *SLC25A3*, *LONP1*, *SDHC*, *GNB2L1/RACK1* were selected
 858 exclusively from AMLCN_H subgroup. Three genes (*FASTKD1*, *MRPL51* AND *ATP6J*) belongs to AMLCN_I
 859 subgroup and two genes (*FASLG* & *CLIC1*) were selected from AMLCN_L subgroup. Furthermore, three genes
 860 (*SLC25A29*, *GLUD1* and *LIG1*) which were common among AMLCN_H and AMLCN_I and two genes, *DHFR*
 861 and *SNCA* were selected from the common genes among AMLCN_H and AMLCN_L cohort. *MMP9* and
 862 *OLFM4* were selected based on the Hub gene and MCODE analysis.

863

864

865

866 **Table 3: Median expression of validated genes in patients (n=143) compared to controls(n=50)**
 867 **and their comparison with TCGA LAML dataset(n=179)**

S.n o.	Gene list	Expression	Median expression (IQR) [#]	P value ^{##}	Expression in our cohort	Expression in TCGA dataset (LAML)*
1.	<i>SLC25A3</i>	Patients	7.45E-04 (5.46E-04-1.08E-03)	< 0.0001	Overexpression	Non-significant
		Controls	4.83E-04(3.08E-04-6.95E-04)			
2.	<i>SDHC</i>	Patients	1.03E-04(7.21E-05-1.66E-04)	< 0.0001	Overexpression	Non-significant
		Controls	5.13E-05(3.39E-05-7.81E-05)			
3.	<i>RACK1/ GNB2L1</i>	Patients	1.56E-02(1.07E-02-2.26E-02)	< 0.0001	Overexpression	Non-significant
		Controls	6.43E-03(4.12E-03-9.42E-03)			
4.	<i>FASTKD1</i>	Patients	6.10E-04(3.75E-04-1.09E-03)	< 0.0001	Overexpression	Non-significant
		Controls	2.71E-04(1.60E-04-5.06E-04)			
5.	<i>ATP5J</i>	Patients	2.67E-06(1.48E-06-4.23E-06)	< 0.0001	Overexpression	Downregulation
		Controls	5.63E-07(1.44E-07 -1.04E-06)			
6.	<i>FASLG</i>	Patients	1.24E-05(5.82E-06-2.23E-05)	< 0.0001	Downregulation	Non-significant
		Controls	2.83E-05(1.72E-05-5.11E-05)			
7.	<i>CLIC1</i>	Patients	9.63E-04(6.70E-04-1.48E-03)	< 0.0001	Overexpression	Downregulation
		Controls	5.46E-04(3.68E-04-7.91E-04)			
8.	<i>HRK</i>	Patients	2.99E-06(1.30E-06-7.32E-06)	< 0.0001	Downregulation	Non-significant
		Controls	2.64E-05(7.26E-06-4.44E-05)			
9.	<i>ALAS2</i>	Patients	9.00E-05(1.68E-05-4.16E-04)	< 0.0001	Downregulation	Downregulation
		Controls	7.61E-04(4.68E-04-1.71E-03)			
10.	<i>SLC25A21</i>	Patients	2.27E-06(6.51E-07-7.02E-06)	< 0.0001	Downregulation	Downregulation
		Controls	8.28E-05(3.86E-05-1.32E-04)			
11.	<i>CYP1B1</i>	Patients	4.87E-07(1.94E-07-1.87E-06)	< 0.0001	Downregulation	Non-significant
		Controls	4.57E-06(3.22E-06-1.05E-05)			
12.	<i>GLUT1</i>	Patients	5.46E-04(3.88E-04-8.10E-04)	< 0.0001	Overexpression	Non-significant
		Controls	2.65E-04(1.91E-04-4.08E-04)			
13.	<i>SLC25A29</i>	Patients	3.48E-04(1.94E-04-6.53E-04)	< 0.0001	Overexpression	Upregulation
		Controls	7.30E-05(5.35E-05-1.26E-04)			
14.	<i>SNCA</i>	Patients	5.73E-05(2.30E-05-1.59E-04)	0.0023	Downregulation	Non-significant
		Controls	1.06E-04(5.29E-05-2.25E-04)			
15.	<i>DHFR</i>	Patients	5.01E-04(2.58E-04-7.72E-04)	0.0115	Non- significant	Non-significant
		Controls	7.15E-04(4.32E-04-1.25E-03)			
16.	<i>MMP9</i>	Patients	2.70E-05(7.74E-06-7.15E-05)	< 0.0001	Downregulation	Non-significant
		Controls	9.05E-04(3.79E-04-1.37E-03)			
17.	<i>OLFM4</i>	Patients	3.78E-05(1.16E-05-1.47E-04)	< 0.0001	Downregulation	Non-significant
		Controls	9.37E-04(4.46E-04-2.54E-03)			
18.	<i>LIG1</i>	Patients	1.57E-04(9.12E-05- 2.46E-04)	0.8696	Non- significant	Non-significant
		Controls	1.38E-04(1.06E-04- 2.46E-04)			
19.	<i>MRPL51</i>	Patients	6.95E-04(5.46E-04-1.07E-03)	0.0116	Non- significant	Downregulation
		Controls	5.25E-04 (3.56E-04-1.00E-03)			
20.	<i>LONPI</i>	Patients	2.93E-04(2.10E-04-4.11E-04)	< 0.0001**	Upregulation	Downregulation
		Controls	1.79E-04(1.03E-04-2.59E-04)			

868 #IQR= Interquartile Range
869 ##Level of significance was set by adjusting alpha error for multiple comparisons by Bonferroni correction ($p < 0.05/20$) i.e. $p < 0.0025$ were considered as significant)
870 (0.05/20) i.e. $p < 0.0025$ were considered as significant)
871 *LAML = adult AML data available on TCGA (The Cancer Genome Atlas) database accessed from *Gepia*
872 ** The expression of *LONPI* showed reverse expression trend in validation cohort when compared to RNA
873 sequencing data of test cohort hence considered not validated.
874
875
876

877 **Table 4: List of genes showing significant correlation with mitochondrial DNA copy number**
878 **using Pearson's correlation**

Gene name	Mitochondrial DNA Copy number (n=143)	
	Correlation coefficient (r)	p value
<i>ATP5J</i>	0.133	0.024
<i>FASTKD1</i>	0.158	0.008
<i>CLIC1</i>	0.165	0.005
<i>RACK1/GNB2L1</i>	0.174	0.003
<i>SDHC</i>	0.172	0.004
<i>SLC25A3</i>	0.155	0.009

879 n= number of patients

880

881

882

883

884 **Table 5: Impact of expression of individual genes and overall risk score on overall survival and**
885 **event free survival of the test cohort (pediatric cohort) and overall survival in validation cohort**
886 **(TCGA adult LAML cohort)**

Gene Name	Overall survival (Pediatric cohort) n=143		Overall survival (TCGA adult LAML cohort) n=179		Event Free Survival (Pediatric cohort) n=143	
	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value
<i>SDHC</i>	1.29(1.14-1.41)	<0.001	0.994(0.941-1.050)	0.826	1.225(1.100-1.363)	<0.001
<i>SLC25A29</i>	0.88(0.83-0.93)	<0.001	0.988(0.981-0.996)	0.003	0.905(0.860-0.952)	<0.001
<i>CLIC1</i>	1.20(1.04-1.38)	0.013	1.002(1.00-1.004)	0.069	1.136(0.984-1.312)	0.082
Risk Score	1.010(1.007-1.014)	<0.001	1.011(1.002-1.021)	0.019	1.008(1.001-1.012)	<0.001

887 CI: Confidence interval; n= number of patients; TCGA The cancer genome atlas; LAML Adult AML dataset

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903 **Table 6: Univariable and multivariable analysis of clinical features with overall survival**

Variables (n=143)	Categories (n)	Univariable analysis			Multivariable analysis	
		Median	Hazard (95% CI)	p value	Hazard (95% CI)	p value
Age(years)	≤10 (64)	20.63	0.98(0.63-1.52)	0.93	-	-
	>10(79)	21.93				
Gender	Male (87)	15.07	0.72(0.45-1.143)	0.16	-	-
	Female (56)	40.23				
Total leukocyte count (/mm³)	<50000(98)	23.77	0.8(0.517-1.238)	0.32	-	-
	>50000(45)	13.88				
Platelets (/μL)	≤ 50000(92)	19.53	0.94(0.59-1.48)	0.789	-	-
	>50000(51)	23.33				
Haemoglobin(g/dl)	≤8(95)	13.87	0.73(0.47-1.12)	0.16	-	-
	>8(48)	24.87				
Fever	Negative (26)	Not reached	1.52(0.82-2.83)	0.18	-	-
	Positive (112)	19.53				
Chloroma (n=132)	Negative (105)	20.63	0.56(0.29-1.07)	0.083	1.45(0.74-2.84)	0.27
	Positive (27)	Not reached				
ELN Risk group (n=134)	Good (67)	40.23	1.95(1.23-3.09)	0.004	0.59(0.36-0.98)	0.041
	Others (67)	12.27				

904 CI: Confidence interval; ELN: European LeukemiaNet.

905

906

907

908

909 **Table 7: Association of 3-gene risk score with clinical and demographic parameters**

Characteristics (n=143)	Risk score Low (%) (n=71)	Risk score High (%) (n=72)	χ^2	P value
Age (years)				
<10 Years (64)	26(36.6)	38(52.8)	3.775	0.052
≥10 years (79)	45(63.4)	34(47.2)		
Sex				
Male (87)	42(59.2)	45(62.5)	0.168	0.682
Female (56)	29(40.8)	27(37.5)		
Hyperleukocytosis, (>50×10³/μL)				
TLC <50×10 ³ /μL (98)	53(74.6)	45(62.5)	2.446	0.118
TLC ≥50×10 ³ /μL (45)	18(25.4)	27(37.5)		
Fever(n=138)				
Negative (26)	17(24.6)	9(13.0)	3.033	0.082
Positive (112)	52(75.4)	60(87.0)		
Chloroma				
Negative (116)	57(80.3)	59(81.9)	1.781	0.182
Positive (27)	14(19.7)	13(18.1)		
Cytogenetics (n=130) *				
Good Risk (50)	32(54.2)	18(32.7)	5.349	0.021
Others (64)	27(45.8)	37(67.3)		
Molecular analysis (n= 122) **				
<i>FLT3ITD</i>				
Negative (105)	54(87.1)	51(85.0)	0.112	0.738
Positive (17)	8(12.9)	9(15.0)		
<i>RUNX1-RUNX1T1</i>				
Negative (69)	29(46.8)	40(66.7)	4.911	0.027
Positive (53)	33(53.2)	20(33.3)		
<i>CBFB-MYH11 #</i>				
Negative (113)	58(93.5)	55(91.7)	-	1.00 ^{##}
Positive (6)	3(4.83)	3(5.0)		
<i>NPM1</i>				
Negative (117)	61(98.4)	56(93.3)	-	0.203 ^{##}
Positive (5)	1(1.61)	4(6.7)		

ELN Risk stratification (n=134) ***				
Good Risk (67)	41(60.3)	26(39.4)	5.852	0.016
Intermediate and poor risk (67)	27(39.7)	40(60.6)		
Complete remission (n=143)				
Achieved (104)	58(81.7)	46(63.9)	5.711	0.017
Not achieved (39)	13(18.3)	26(36.1)		

910 *Cytogenetics failed (n=16) and not done in n=13 cases.
 911 ** Molecular analysis was not done in 19 cases; molecular mutation was absent in n=43 cases
 912 #*CBFB-MYH11* mutation was assessed in n=119 cases
 913 ## Fisher's Exact Test
 914 ***ELN (European LeukemiaNet) risk stratification was done using both cytogenetics and molecular markers in
 915 134 patients. However, 12 patients risk stratification was done with only cytogenetics and in 20 patients, it was
 916 done by only molecular analysis
 917 χ^2 : Chi square value; TLC: Total leukocyte count; *FLT3 ITD*: FMS-like tyrosine kinase internal
 918 tandem duplication; *RUNX1-RUNX1T1*: runt-related transcription factor 1-RUNX1 partner transcriptional co-
 919 repressor 1 fusion transcript; *CBFB-MYH11*: core binding factor beta-myosin heavy chain 11 fusion transcript;
 920 *NPM1*: Nucleophosmin 1;

921

922

923 **Table 8: Predictive ability of combined gene risk score group and ELN risk category on**
 924 **survival outcome in pediatric AML cohort**

925

Variables(n)	HR (95%CI)	P Value	P value (overall)	Median OS (months)	Predicted 12 months survival	Predicted 18 months survival
ELN Good risk & gene signature risk group Low (41)	1	--	<0.001	Not reached	80%±6%	75%±7%
ELN Others & gene signature risk group Low (27)	1.58 (0.78-3.21)	0.20	-	27.77(±7.82)	74% ±8%	58% ±10%
ELN Good risk & gene signature risk group High (26)	2.12 (1.06-4.26)	0.034	-	22.90(±8.31)	54% ±10%	44% ±10%
ELN Others & gene signature risk group High (40)	3.83 (2.07-7.07)	<0.001	-	4.67(±3.71)	33% ± 8%	25% ± 7%

926 HR: Hazard Ratio; CI: Confidence interval; OS: overall survival; ELN: European LeukemiaNet; AML: Acute
 927 myeloid leukemia

928

929

930

931

932

933

934

935

936

937

938

939