**IMPORTANT NOTE:** The following is a thorough annotation of a subset of clusters we identified within our transcriptional heterogeneity analysis project. Currently we are considering publishing this series of documents independently of the main paper, as an online resource for users to reference when diving into our pan-cancer atlas. These will be then updated with time as new and better information emerges. Alternatively, a stripped-down version could be published as supplemental information.

**T005 Leukemia**

We observe a separation between lymphoblastic leukemias, **T119 ALL** (n = 334), and myeloid leukemias, **T120 AML** (n = 472) at the second hierarchical level. A significant difference in age is expected due to the different etiologies (median age 7.16 vs 16.76 y.o., MWU p-value = 2.98e-23) and the presence of both adult and pediatric populations in both groups to different degrees. No significant difference in survival is observed.

**Acute lymphoblastic leukemia**

Along the lymphoblastic branch we immediately observe the separation of a small group of infant leukemias with *MLL* rearrangements, found in **T121 ALL INF MLLr** (n =14), from all other diagnoses, in **T122 ALL A** (n = 320) (Fig. S25a, b). **T121** contains most samples marked as infant (6 vs 1 χ2 p-val < 2.20e-16) and mixed-lineage leukemia (4 vs 0, χ2 p-val < 2.20e-16) and has a significantly younger median age (0.73 vs 7.20 y.o., MWU p-value = 4.37e-02). We confirmed this annotation with gene sets, as **T121** is highly enriched for *MLL* downstream targets (medNES = 1.50 , MWU adj. p-val = 2.47e-09)1 (Fig. S25e).

**T122** further splits into two subclasses, **T123 ALL B** (n = 127) and **T124 ALL TRG** (n = 193) (Fig. S25a, b), containing most of samples from TARGET. Gene sets analysis between all TARGET leukemia samples and the remaining cohort shows enrichment (MWU adj. p-val < 1.00e-10) of poly-A RNA binding, ribonucleoprotein complex, RNA processing, ribosomal and mitochondrial pathways, and oxidative phosphorylation (Gene Ontology2,3) in **T124**. Furthermore, **T124 ALL TRG** has a lower median age (6.41 vs 13.17 y.o., MWU adj. p-val = 5.04e-08). We couldn’t identify with statistical certainty any biological driver behind the split between **T123** and **T124**; stringent low variance genes removal or more advanced batch effect removal methods (e.g. COMBaT4) weren’t enough to assure complete compatibility between the TARGET cohort and the rest of the dataset without the loss of information and damage to the subtyping process. We decided to keep the clusters separate as by choice of the algorithm and further investigate their subtypes independently, to maintain tumour subtypes that were exclusive of one or the other cohorts and increase the classifier range.

At the next level within **T123**, we observe the separation of **T126 ALL ETV6-RUNX1** (n = 20) a small class of samples marked with *ETV6-RUNX1* fusion (χ2 p-value < 2.20e-16) from the remaining ALLs in **T125 ALL C** (n=107) (Fig. S25a, b). The t(12;21)(p13;q22) translocation which results from this fusion is often accompanied by copy number gains in *RUNX1*, which is overexpressed in this group (logFC = 4.17e-01, FDR = 3.33e-03). Compared to samples in **T125**, **T126** is significantly younger (14.5 vs. 4.46 y.o., MWU p-val = 3.29e-08)5.

Following a similar pattern, samples in **T128** **ALL ERGdel** (n = 36), a child class of **T126**, separate from the remaining samples **T127 ALL Ph-like** (n = 71) (Fig. S25a, b). **T128** is characterized by tumours carrying *ERG* deletions (15 vs. 55, χ2 p-val < 2.20e-16), and exhibits characteristic overexpression of *CHST2* (logFC = -4.48, FDR = 5.742e-33), *PTPRM* (logFC = -7.64, FDR = 2.987e-32), and GPR49/*AGAP1* (logFC = -6.23, FDR = 3.201e-31)6. The majority of samples in **T127** are composed of Ph-like tumours of various classes (χ2 p-val < 2.2e-16)7.

**T128** is further subdivided in two child nodes, **T129 ALL Ph-like A** (n = 41) and **T130 ALL Ph-like IKZF1/JAK2** (n = 29) (Fig. S25b, d). Both contain small populations *of BCR-ABL1* fusion samples (Ph+) (11 and 5, ns) and Philadelphia-like (Ph-like) samples (13 and 14, ns). While **T129** contains the majority of Ph-like non-*CRFL2* tumours (11/28 vs. 14/19, χ2 p-val = 4.32e-02), there is no corresponding enrichment of this signature via gene sets analysis (Fig. S25e). However, the two differ by some specific lesions known to be present in the Ph-like group: **T130** contains 6 *JAK2* fusion samples (0/13 vs 6/14, χ2 p-val = 2.69e-02), while **T131** contains all *EPO* fusion samples (4/13 vs. 0/14, FET p-val 4.07e-02) (Fig. S25d). Both contain other *JAK/STAT* alterations (4/13 vs. 3/14, ns), and two of other *ABL1/2* fusion samples each (Fig. S25d). **T130** is also enriched for tumours with concurrent *IKZF1* alterations 11/28 vs. 14/19, χ2 p-val = 4.32e-02).

**T129** then divides into two further subtypes, **T131 ALL Ph-like JAK/STAT** (n=23) **and T132 ALL Ph+/Ph-like** **EPOR** (n =12) (Fig. S25b, d). **T132** contains the majority of *BCR-ABL1* fusion samples (3/23 vs. 8/12, p-val = 4.23e-03) (Fig. S25d). Of the Ph-like samples for which we have annotation, **T131** contains 4 unspecified *JAK/STAT* mutants along with an additional *CRLF2-JAK* mutant, a *CRFL2* rearranged sample with no *JAK* rearrangements, and a *RAS* mutant (Fig. S25e). **T132** contains 3 *EPOR-IGH* fusion samples, while **T131** contains an *EPOR-IGK* fusion (ns) (Fig. S25e). Both groups contain one *ABL* fusion without *CRFL2* rearrangement, while Ph-like non-*CRLF2* samples are evenly divided between the clusters (7/20 vs. 4/8, n.s.) (Fig. S25e). Another interesting distinction is that **T131** is enriched for tumours with cell-cycle related lesions, either in *TP53*, *CDK2NA/B*, or *RB1* (14/20 vs. 1/8, χ2 p-val = 1.95e-2). **T132**, however, is enriched for samples with concurrent *IKZF1* alterations (5/20 vs. 6/8, χ2 p-val = 4.35e-02), though these are heterogeneous and have some overlap between the two clusters8.

Gene set enrichment analysis shows that **T131** is enriched for non-Ph-like *CRFL2* rearranged samples (medNES = 1.57, MWU adj. p-val = 4.70e-05), while **T132** is enriched for Ph-like samples with *CRFL2* rearrangments (medNES = 2.68, MWU adj. p-val = 1.61e-07)9 (Fig. S25e), suggesting that **T132** may contain *CRFL2*-rearranged samples which have not been annotated as such.

**TARGET cohort**

On the other hand, we see the TARGET branch, **T124**, splitting in four different subtypes (Fig. S25b, c).

**T133 ALL TRG A** (n = 109) is the largest cluster and contains a mixture of genomic alterations: ALL with hyperdiploidy without trisomy of chr4 and ch10 (χ2 p-val = 3.31e-4), ALL with hyperdiploidy with trisomy chr4 and ch10, samples with iAMP21, plus a number of unspecified samples (Fig. S25c). The cluster is characterized by significant overexpression of *CRLF2* (edgeR glmQLFTest, logFC= 4.96, FDR ≤ 7.749e-04). Indeed, gene set enrichment analysis confirmed this cluster contains a sizeable population of Ph+ and Ph-like samples (medNES = 79.08, adj. p-val = 7.03e-14, Dunn adj. p-val < 1.00e-03).

**T134 ALL TRGZNF384** (n = 13) is the smallest cluster and contains the oldest group of patients (median age 13.23 y.o., KW p-val = 1.13e-03). Patients with ALL in this cluster display the best overall survival (lrt p-val < 1e-04). Gene set enrichment analysis of genes upregulated and downregulated in *ZNF384*-rearanged ALLs reveals this cluster to be highly enriched for *ZNF384*-fusion downstream upregulated targets (medNES ≥ 1.51, KW adj. p-val = 5.56e-16 , Dunn adj. p-val < 1.00e-04) and impoverished for downregulated targets (medNES ≤ 4.81e-01, KW adj. p-val = 2.59e-16, Dunn adj. p-val < 1.00e-04) (Fig. S25e).

**T135 ALL TRG TCF3** (n = 30) is comprised of samples harbouring *TCF3-PBX1* (n = 19, χ2 p-val < 2.2e-16) and *TCF3-HLF* (n = 3, χ2 p-val = 1.60e-02) fusions (Fig. S25c). Out of all TARGET ALL subgroups, **T135** contains the patient group with theworst overall survival, reaching median OS at 483 days (lrt p-val = 6.30e-22 at 4383 days, post-hoc pairwise lrt p-val ≤ 1.5e-06); Interally to this class, patients with *TCF3-HLF* fusions exhibited significantly worse survival (lrt p-val = 4.89e-02), consistent with literature10.

**Finally, T136 ALL TRG ETV6-RUNX1** (n = 27) is the class comprising the youngest patients (median 3.1 y.o., KW p-val = 1.13e-03) and contains samples with *ETV6-RUNX1* fusions (n = 20, χ2 p < 2.2e-16) (Fig. S25c).

**T133** separates in further components (Fig. S25b, c).

**T137 ALL TRG Ph-like *CRLF2* (**n=29) contains all samples labelled as harbouring *BCR-ABL1* fusions (n = 3), *MLL* rearranged ALLs (n=3), and the highest proportion of otherwise unspecified ALLs (n = 23, χ2 p-val = 2.95e-05) (Fig. S25c). It shows overexpression of *CRLF2* (logFC= 2.99, FDR = 1.48e-02) and enrichment of *CRLF2*-rearrangment signatures in Ph-like ALL (Ph+ *CRFL2*pos, medNES = 2.21, KW adj. p-val = 3.05e-03)9 (Fig. S25e). It also exhibits overexpression of *IDH1* (logFC= 1.28, FDR = 3.66e-05), *JAK1* (logFC = 0.641, FDR = 4.15e-02) and is enriched for Ph-like gene signatures (medNES = 2.88, KW adj. p-val = 9.79e-06, Dunn adj. p-val < 1.00e-03)8,11 when compared to its siblings (Fig. S25e).

**T138 ALL HYPRD** (n=21) is enriched for tumours with hyperdiploidy without trisomy of both chromosomes 4 and 10 (1/29 vs. 11/20 vs. 7/22, χ2 p-val =2.66e-4) (Fig. S25c). Patients in **T138** are also significantly younger than its siblings (3.59 y.o., KW p-val = 1.14e-02). Furthermore, **T138** exhibits the highest DNA index of its siblings, an indicator of hyperdiploidy (median = 1.17, KW p-val = 3.97e-07, Dunn adj. p-val ≤ 4.18e-03)12.

**T139 ALL TRG Ph-like EPOR** (n = 22) is characterized by overexpression of *EPOR* (median logFC = 2.06, FDR ≤ 1.20e-04), as well as enrichment of erythrocyte developmental gene sets (medNES = 1.22, KW adj. p-val = 2.06-06, Dunn adj. p-val < 5.00e-02)2,3(Fig. S25e)It also exhibits overexpression of *IDH2* (median logFC= 1.65, FDR ≤ 3.40e-11).

**Acute myeloid leukemia**

Myeloid malignancies in **T120** separate instead in 9 different classes at the very next level (Fig. S26a, b). Similar to ALLs, we observe among these two classes made up exclusively of TARGET samples: **T144 AML TRG** and **T146 AML TRG IDH2low**, which are discussed at the end of this section.

**T140 AML MLLr** (n = 52) has a median age of 60.00 y.o (KW p-val =1.54e-48). It contains a number of samples marked for *MLL* fusions (most of them high risk, χ2 p-val = 4.45e-08), and is highly enriched (medNES > 1.08, KW adj. p-val < 1.00e-40, Dunn adj. p-val < 1.00e-04) for their matching pathways1,13 (Fig. S26d). It is also enriched for *NPM1* mutated pathways (medNES = 1.07, KW adj. p-val = 1.97e-38, Dunn adj. p-val < 1.00e-04)13 suggesting a large cohort within this class may be *NPM1* mutated. Indeed, all samples in this cluster for which we have *NPM1* and *FLT3* mutation data are mutated for either *NPM1* (n=23) or *FLT3* (n=16). This class displays poor overall survival (lrt p-val = 6.31e-11at 4022 days), reaching median OS at 327days.

**T140** splits into two subclasses (Fig. S26b). **T149 AML MLL 11q23** (n = 8) is a very small cluster and is considerably younger (45.00 vs 62.00 MWU p-val = 7.24e-03) than **T150 AML MLL NPM1/FLT3** (n = 44); this is also reflected in the percentage of samples marked as paediatric (50.00% vs. 4.55%, χ2 p-val = 7.25e-03). While 5 samples are marked as AMLs, it also contains 3 samples marked as mixed lineage leukemias (χ2 p-val = 7.79e-04). **T149** contains 4 samples from TCGA, all of which are annotated with *MLL* fusions (two *MLL10-MLL* and one *MLL-MLLT3*, and one *MLL-MLLT4*), while **T150** contains 40 samples from TGCA, 10 of which have reported gene fusions, with seven involving *MLL* genes. When compared to **T150**, **T149** is significantly enriched for genes ets involving chr11q23 rearrangement (medNES = 8.46, adj. p-val = 1.06e-08)14 and VALK AML cluster 16 (medNES = 4.03, adj. p-val = 2.66e-09), which is composed of samples with 11q23 rearrangements15. Cluster **T150 AML MLL NPM/FLT3** inherits all of **T140’s** *NPM1* and *FLT3* mutants16, and is enriched for their corresponding gene sets (medNES =2.34, KW adj. p-val =7.97e-08, medNES = 1.85, KW adj. p-val = 1.25e-04, respectively)15,17.

**T141 AML BM** (n = 30) is a mixed-lineage cluster. It comprises myeloid, megakaryoblastic, non-specific, and lymphoblastic leukemias along with a few lymphomas and osteosarcomas. It is not enriched for any leukemia associated gene sets, suggesting this class may contain samples contaminated by normal blood or bone marrow tissue.

**T142 AML MATlow** (n = 105) is largely composed of FAB subtypes M1 (n = 33, χ2 p-val = 7.44e-04), AML with minimal maturation, and M2 (n = 34, χ2 p-val = 1.60e-06), AML with maturation, and a smaller subpopulation of undifferentiated M0 (n = 15, χ2 p-val = 1.15e-04). It also contains older patients, with a median age of 57 y.o. It is enriched for samples classified as intermediate (n = 54, χ2 p-val = 1.43e-07) and high-risk (n = 37, χ2 p-val 1.61e-09). It also contains two *BCR-ABL1* fusion samples, 24 *FLT3* mutants - all of which are from the TCGA, though the mutations themselves are heterogenous – 24 *NMP1* mutants, 21 of which are W288F (χ2 p-val < 2.2e-16), along with 9 *WT1* mutants (χ2 p-val = 1.56 e-4). All samples in this cluster for which we have *NPM1* and *FLT3* mutation data have mutations in either gene. This cluster displays intermediate low overall survival, reaching median OS at 417 days.

This group’s two child classes, **T151AML MATlow NPM1mut and T152 AML MATlow noNPM1** (Fig. S26b)**,** are separated by the presence or absence of *NPM1* mutations, as well as karyotypic complexity. **T151AML MATlow NPM1mut** (n = 34) has a higher ratio of FAB M1 samples, AMLs with minimal maturation, (16/32 vs 17/62, FET p-val = 4.04e-02) and inherits all *NPM1* mutate samples except for one, a p.K263R (χ2 p-val = 6.67e-13); all samples for which we have *NPM1* data within this cluster (n=25) are *NPM1* mutated. As expected, we confirmed this annotation through significance (medNES = 1.25, 1.MWU adj. p-val = 7.83e-16) in *NPM1* mutation pathways13. Its sibling, **T152 AML MATlow noNPM1** (n = 71), has a higher proportion of FAB M0 samples, undifferentiated AML (1 vs. 14, FET p-val = 3.21e-02), and possibly is carrying equivalent samples without *NPM1* mutation. M2 samples are evenly split between the clusters (χ2 p-val = 6.51e-01), suggesting maturation is not a critical determinant of this split. Samples with *FLT3* and *WT1* mutations are more common in **T151** than in **T152**, confirmed by gene sets for *FLT3* mutation (medNES = 1.90, MWU adj. p-val = 2.29e-13)15. We observe no significant separation in survival between the two clusters.

Finally, **T152** further splits into **T153 AML FLT3-ITD** (n = 58) and **T154 AML CEBPA** (n = 13) (Fig. S26b), which differ significantly in age (63 vs 32 y.o. MWU p-val = 7.80e-05). Cluster **T153** contains all M0 samples (n = 14 vs 0) while **T154** is enriched for FAB M2 samples (n = 12 vs 9, χ2 p-val = 3.81e-03). **T153** also contains five acute megakaryoblastic leukemias and two mixed lineage leukemias. **T153** carries more samples with complex cytogenetics (χ2 p-val < 0.001) and has significantly reduced survival (lrt p-val = 2.00e-02). In line with findings described in literature, **T153** exhibits a higher mutation burden (median = 17.00 vs. 8.50, MWU p-val = 2.06e-03), which is largely related to age in AML18. **T153** contains 6 *FLT3* mutant samples (three of which have in frame insertions), while **T154** contains only one. It also shows overexpression of a myriad of genes (21/39, FDR < 0.05), which are known to be upregulated in samples harbouring *FLT3* internal tandem duplications (*FLT3*-ITD*)*, as well as enrichment of *FLT3*-ITD gene sets (medNES = 3.11, KW adj. p-val = 5.99e-08)15. On the other hand, **T153** contains only 3 *CEBPA* mutated samples, while **T154** contains 8 (χ2 p-val = 3.28e-06)

Cluster **T143 AMKL** (n = 49) is exclusively composed of megakaryoblastic samples (n = 41, χ2 p-val < 2.20e-16) while 8 samples are unlabelled, and as expected is enriched for AMKL pathways (medNES ≥ 1.70 , KW adj. p-val = 1.01e-38)1. Note that all cases are non-down syndrome. This cluster displays the worst overall survival of all its siblings, reaching median OS at 313 days (lrt p-val = 6.31e-11). **T143** then splits into **T155 AMKL CBFA2T3-GLIS2** (n = 12) and **T156 AMKL HOX** (n = 37). Though both are entirely paediatric, the former cluster is significantly younger than the latter, median age of 0.97 vs 2.17 y.o. (MWU p-val = 2.08e-02). All samples in **T155** for which genomic data are available are characterized by a *CBFA2T3-GLIS2* fusion (9/9 vs. 0/25, χ2 p-val = 7.03e-08)19. This cluster has the worst survival of its siblings, reaching median OS at just 313 days post diagnosis. **T156** is composed of other driver events: two *GATA1* mutants, four HOXr (*HOX* fusion) samples, eight *KMT2A-MLLT3/10* fusions, four *NUP98-KDM5A* fusions, two *RBM15-MKL1* fusions, and four samples with other driver mutations. With a greater sample size its possible these mutations would form their own clusters as well. When comparing these two classes, **T156** exhibits overexpression of *HOXA* (11/11 genes upregulated, median logFC ≤ -5.67, FDR ≤ 8.47e-03 ) and *HOXB* genes (8/10 upregulated, median logFC = -5.65, FDR ≤ 7.31e-03)19.

The remaining classes within this branch are defined by clear fusion events. All samples within **T145 AML CBFB-MYH11** (n = 14) are marked as core binding factor positive, *CBFB-MYH11*. As we expected, it is enriched (medNES ≥ 1.35 , KW adj. p-val = 3.65e-27) for *CBFB-MYH11* associated gene sets1,15. All samples in **T147 APML** (n = 15), except for one, are positive for *PML-RARA* fusions (χ2 p-val < 2.20e-16) and marked as FAB M3 (χ2 p-val < 2.20e-16), acute promyelocytic leukemia. This class also contains 5 samples with *FLT3* mutations, four of which are p600 in frame insertions (from TCGA); these seem to be exclusive to this cluster. This class has the best overall survival of the cohort, with >60% of patients surviving at 4022 days post diagnosis. Finally, **T148** **AML RUNX1-RUNX1T1** (n = 13) exclusively contains *RUNX1-RUNX1T1* fusion AMLs (χ2 p-val < 2.20e-16). It has moderate-good OS, reaching median OS 2910 days.

**TARGET cohort**

We observe two classes within the AML branch with an exclusive TARGET composition (Fig. S26b).

**T146 AML TRG IDH2low** (n = 23) is composed by samples with various diagnostic categories: three *MLL* fusions (n =3), eight normal karyotypes, and 10 other lesions, including two t(X;10)(p11.2;p11.2), add(17)(p11.2) and two inv(17)(p13.1q11.2), both exclusive to this group. However, it contains the highest proportion of *WT1* mutations (7/23, χ2 p-val = 1.39e-3) and *FLT3*-ITDs (8/23, χ2 p-val = 2.427e-05) amongst the TARGET cohort. It also exhibits the lowest expression of *IDH2* (logFC = -0.836, p-val = 2.58e-2 against **T155**-**T159** and **T162**). This group displays intermediate survival, reaching median OS at 1394 days post diagnosis.

**T144 AML TRG** (n = 163) is the largest cluster of all the AMLs and is composed largely of unspecified AMLs (n=154), and surprisingly 5 ALLs. It is an entirely paediatric cluster (median age 9.36 y.o.) and has excellent survival, with >50% of patients surviving at 4022 days post diagnosis.

Diving deeper into this class (Fig. S26b, c), we observe first the singling out of AML with *MLL* translocations (23/33 v 12/120, χ2 p-val = 2.623e-12) in **T158 AML TRG MLLr** (n = 33), from everything else in **T157 AML TRG A** (n = 130). As expected, **T158** shows enrichment (MWU adj. p-val ≤ 1.00e-03) of *MLL*-associated gene sets1,13. There is no difference in survival.

Down by another level (Fig. S26b), we observe **T157** splitting into three small subclasses, characterized by different molecular aberrations: **T159 AML TRG KMT2Ar/MPAL** (n = 65), **T160 AML TRG CFB-MYH11** (n = 36) and **T161 AML TRG RUNX-RUNX1T1** (n = 29).

Aside from myeloid malignancies, **T159** contains 4 ALL samples, one unspecified leukemia and one lymphoma. It has the highest proportion of intermediate risk samples (n = 36, χ2 p-val = 1.581e-06) and exhibits a significantly worse survival rate than either of its siblings (lrt p-val = 2.20e-04). This cluster also inherits all *NPM1* mutant samples, while *FLT3*-ITD and *WT1* mutants are spread across all three clusters.

This class also contains samples labelled as *MLL*-rearranged (n = 11/56, χ2 p-val = 4.103e-03). It shows overexpression of a wide variety of *HOX* genes (24/39 *HOX* genes with median logFC > 0 & FDR < 0.05, 22/39 FDR < 1e-04, median logFC = 4.62), a phenotype previously described in AMLs with *MLL* partial (internal) tandem duplication (*MLL-*PTD)20. The characteristic expression patterns of *MLL-*PTD could explain the inclusion of a handful of ALL samples, which may also harbour non-canonical *MLL* aberrations.

Indeed, manual inspection of a subsample of 8 mRNA sequences (5 labelled as AML, 3 as ALL) from TARGET revealed the majority of samples (4/8) harbouring complex lesions in *KMT2A* or (2/8) with rearrangments to exon 7 and 8 associated with *MLL-*PTDs.

The transcriptional profile of *KMT2A* lesions in this class departs from that most commonly described by literature, as most gene sets involving *MLL* mutated leukemias agree an impoverishment in this class when compared to the bona-fide *MLL*-rearranged AML class **T158** (medNes ≥ 1.27 for positive signatures in **T158**, ≥1.61 for negative signatures in **T159**, MWU p-val ≤ 3.32e-15)1,13.

A single sample harbours a BSG-CDC34 fusion. While no MLL mutation was reported, CDC34 is known to mediate stability and degradation of MLL21,22, supporting the idea of a group of tumours whose MLL pathways are affected by alternative mechanisms.

*KMT2A* rearrangements are also common in Mixed Phenotype Acute Leukemias (MPAL)23,24; to assess whether some of these samples are mixed phenotype ALs, we interrogated a number of gene sets (Fig. S26b). Indeed, MPAL expression sets were significantly upregulated in AMLs within **T159** when compared to all other AMLs in **T120** (medNES = 1.20, MWU p-val = 2.94-12), which in turn have higher markers of AML vs MPAL (medNES = 1.04, MWU p-val = 5.55e-05)25. Furthermore, these samples carry higher lymphocyte differentiation against other AMLs from their family class (**T159** vs **T120**, medNES = 2.63, MWU p-val = 1.36e-03)2,3. In turn, the four ALL samples within this same class have significant enrichment for myeloid differentiation when compared to all other ALLs in **T119** (medNES = 1.25, MWU p-val = 9.01e-04)2,3.

Furthermore, we report enrichment of T-cell development and differentiation gene sets when comparing samples of matching reported lineage to either **T120 AML** (medNES ≥ 1.10, MWU p-val ≤ 9.82e-08) and **T119** **ALL** (medNES ≥ 1.27, MWU p-val ≤ 2.77e-02)2,3, composed exclusively of b-cell lymphoid leukemias (Fig. S26b). These results support the hypothetical presence of T-cell MPALs within this group. While limited information is given by the labelling of these samples, we can confidently speculate this class includes *KMT2A*-rearranged B-cell and/or T-cell MPALs, or at least samples of either linage expressing both myeloid and lymphoid markers.

Finally, a more straightforward annotation allows us to recognize **T160 AML TRG CBF-MYH11** as harbouring core binding factor-mutated samples. It contains a majority of samples originally labelled as carrying such fusion (n = 26/35, χ2 p-val = 9.70e-15), and shows enrichment (medNES ≥ 1.84, KW adj. p-val = 9.376e-16, Dunn adj. p-val < 1.00e-04) of gene sets associated with this fusion1. Similarly, **T161 AML TRG RUNX1-RUNX1T1** contains most samples labelled as harbouring *RUNX1-RUNX1T1* fusions (n = 18/29, χ2 p-val = 1.77e-11) and is enriched for respective gene sets (medNES ≥ 1.01, KW adj. p-val = 5.83e-04)26. It also contains 6 *CEBPA* mutants (χ2 p-val = 8.21e-3).

**Bibliography**

1. Ross, M. E. *et al.* Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* **104,** 3679–3687 (2004).

2. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

3. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47,** D330–D338 (2019).

4. Lazar, C. *et al.* Batch effect removal methods for microarray gene expression data integration: a survey. *Brief. Bioinformatics* **14,** 469–490 (2013).

5. Sun, C., Chang, L. & Zhu, X. Pathogenesis of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia and mechanisms underlying its relapse. *Oncotarget* **8,** 35445–35459 (2017).

6. Yeoh, E.-J. *et al.* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1,** 133–143 (2002).

7. Jain, N. *et al.* Ph-like acute lymphoblastic leukemia: a high-risk subtype in adults. *Blood* **129,** 572–581 (2017).

8. Harvey, R. C. *et al.* Development and Validation Of a Highly Sensitive and Specific Gene Expression Classifier To Prospectively Screen and Identify B-Precursor Acute Lymphoblastic Leukemia (ALL) Patients With a Philadelphia Chromosome-Like (“Ph-like” or “BCR-ABL1-Like”) Signature For Therapeutic Targeting and Clinical Intervention. *Blood* **122,** 826–826 (2013).

9. Sadras, T. *et al.* Differential expression of MUC4, GPR110 and IL2RA defines two groups of CRLF2-rearranged acute lymphoblastic leukemia patients with distinct secondary lesions. *Cancer Lett.* **408,** 92–101 (2017).

10. Inukai, T. *et al.* Hypercalcemia in childhood acute lymphoblastic leukemia: frequent implication of parathyroid hormone-related peptide and E2A-HLF from translocation 17;19. *Leukemia* **21,** 288–296 (2007).

11. Harvey, R. C. *et al.* Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood* **116,** 4874–4884 (2010).

12. Rachieru-Sourisseau, P. *et al.* DNA Index in childhood acute lymphoblastic leukaemia: a karyotypic method to validate the flow cytometric measurement. *Int. J. Lab. Hematol.* **32,** 288–298 (2010).

13. Mullighan, C. G. *et al.* Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias. *Leukemia* **21,** 2000–2009 (2007).

14. Yagi, T. *et al.* Identification of a gene expression signature associated with pediatric AML prognosis. *Blood* **102,** 1849–1856 (2003).

15. Valk, P. J. M. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N. Engl. J. Med.* **350,** 1617–1628 (2004).

16. Braoudaki, M. *et al.* The frequency of NPM1 mutations in childhood acute myeloid leukemia. *J Hematol Oncol* **3,** 41 (2010).

17. Verhaak, R. G. W. *et al.* Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* **94,** 131–134 (2009).

18. Shaver, A. C. *et al.* Mutational burden in acute myeloid leukemia is largely age dependent. *Blood* **126,** 2605–2605 (2015).

19. de Rooij, J. D. E. *et al.* Pediatric non-Down syndrome acute megakaryoblastic leukemia is characterized by distinct genomic subsets with varying outcomes. *Nat. Genet.* **49,** 451–456 (2017).

20. Dorrance, A. M. *et al.* Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations. *J. Clin. Invest.* **116,** 2707–2716 (2006).

21. Meyer, C. *et al.* The MLL recombinome of acute leukemias in 2017. *Leukemia* **32,** 273–284 (2018).

22. Sugeedha, J., Gautam, J. & Tyagi, S. SET1/MLL family of proteins: functions beyond histone methylation. *Epigenetics* **16,** 469–487 (2021).

23. Winters, A. C. & Bernt, K. M. MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front. Pediatr.* **5,** 4 (2017).

24. Yang, W., Tran, P., Khan, Z., Rezk, S. & O’Brien, S. MLL-rearranged mixed phenotype acute leukemia masquerading as B-cell ALL. *Leuk. Lymphoma* **58,** 1498–1501 (2017).

25. Bian, S. *et al.* Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362,** 1060–1063 (2018).

26. Tonks, A. *et al.* Transcriptional dysregulation mediated by RUNX1-RUNX1T1 in normal human progenitor cells and in acute myeloid leukaemia. *Leukemia* **21,** 2495–2505 (2007).