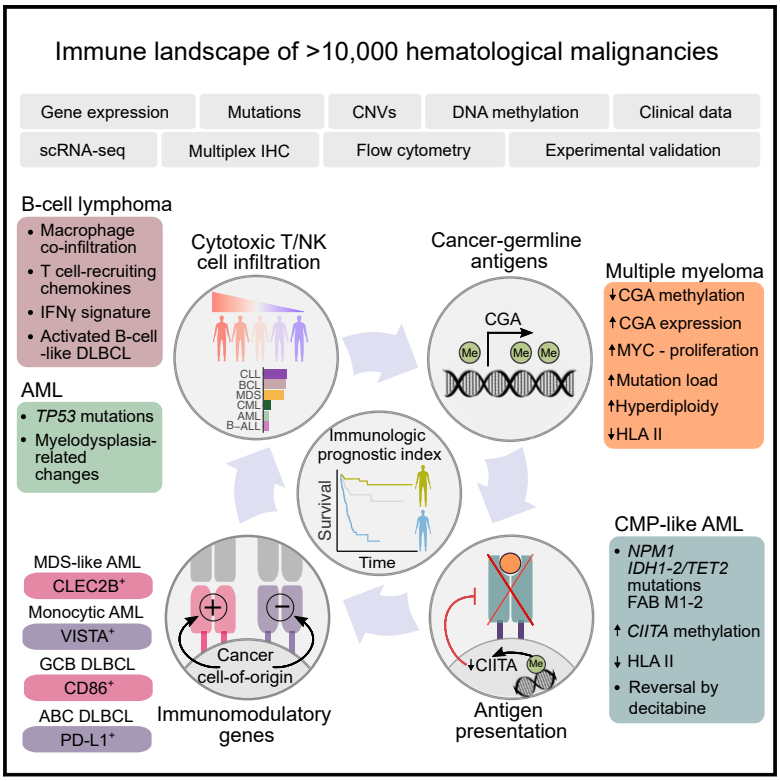
Cancer Cell

Article

Immunogenomic Landscape of Hematological Malignancies

Graphical Abstract



Highlights

- Molecular subtypes and genetics shape immune landscape in hematological malignancies
- Cytotoxic T/NK cell infiltration in MDS-like AML with *TP53* mutations and ABC DLBCL
- Methylation changes suppress HLA genes in AML and induce cancer antigens in myeloma
- Cancer type-specific targets such as VISTA in myeloid and CD70 in lymphoid cancers

Authors

Olli Dufva, Petri Pölönen, Oscar Brück, ..., Olli Lohi, Merja Heinäniemi, Satu Mustjoki

Correspondence

merja.heinaniemi@uef.fi (M.H.), satu.mustjoki@helsinki.fi (S.M.)

In Brief

Dufva et al. define the immune landscape of hematological malignancies by integrating genomic, epigenomic, and transcriptomic data and describe how this influences immune cell infiltration, antigen presentation, anti-tumor immune responses, and patient outcomes.

Dufva et al., 2020, Cancer Cell *38*, 380–399
September 14, 2020
https://doi.org/10.1016/j.ccell.2020.06.002

CellPress

CellPress

Cancer Cell

Article

Immunogenomic Landscape of Hematological Malignancies

Olli Dufva,[1,2,3,11] Petri Pölönen,[4,11] Oscar Brück,[1,2,3,12] Mikko A.I. Keränen,[1,2,12] Jay Klievink,[1,2] Juha Mehtonen,[4] Jani Huuhtanen,[1,2] Ashwini Kumar,[5] Disha Malani,[5] Sanna Siitonen,[6] Matti Kankainen,[1,2,3] Bishwa Ghimire,[5] Jenni Lahtela,[5] Pirkko Mattila,[5] Markus Vähä-Koskela,[5] Krister Wennerberg,[5] Kirsi Granberg,[7] Suvi-Katri Leivonen,[8,9] Leo Meriranta,[8,9] Caroline Heckman,[3,5] Sirpa Leppä,[3,8,9] Matti Nykter,[7] Olli Lohi,[10] Merja Heinäniemi,[4,13,*] and Satu Mustjoki[1,2,3,13,14,*]

[1]Hematology Research Unit Helsinki, Helsinki University Hospital Comprehensive Cancer Center (HUH CCC), 00029 Helsinki, Finland
[2]Translational Immunology Research Program and Department of Clinical Chemistry and Hematology, University of Helsinki (UH), 00029 Helsinki, Finland
[3]iCAN Digital Precision Cancer Medicine Flagship, Helsinki, Finland
[4]Institute of Biomedicine, School of Medicine, University of Eastern Finland, 70211 Kuopio, Finland
[5]Institute for Molecular Medicine Finland, UH, 00014 Helsinki, Finland
[6]Department of Clinical Chemistry, UH and HUSLAB, HUH, 00029 Helsinki, Finland
[7]Laboratory of Computational Biology, Faculty of Medicine and Health Technology, Tampere University (TU), 33014 Tampere, Finland
[8]Department of Oncology, HUH CCC, 00029 Helsinki, Finland
[9]Applied Tumor Genomics Research Program, Faculty of Medicine, UH, 00014 Helsinki, Finland
[10]Tampere Center for Child Health Research, TU and Tays Cancer Center, Tampere University Hospital, 33521 Tampere, Finland
[11]These authors contributed equally
[12]These authors contributed equally
[13]Senior authors
[14]Lead Contact
*Correspondence: merja.heinaniemi@uef.fi (M.H.), satu.mustjoki@helsinki.fi (S.M.)
https://doi.org/10.1016/j.ccell.2020.06.002

SUMMARY

Understanding factors that shape the immune landscape across hematological malignancies is essential for immunotherapy development. We integrated over 8,000 transcriptomes and 2,000 samples with multilevel genomics of hematological cancers to investigate how immunological features are linked to cancer subtypes, genetic and epigenetic alterations, and patient survival, and validated key findings experimentally. Infiltration of cytotoxic lymphocytes was associated with *TP53* and myelodysplasia-related changes in acute myeloid leukemia, and activated B cell-like phenotype and interferon-γ response in lymphoma. *CIITA* methylation regulating antigen presentation, cancer type-specific immune checkpoints, such as VISTA in myeloid malignancies, and variation in cancer antigen expression further contributed to immune heterogeneity and predicted survival. Our study provides a resource linking immunology with cancer subtypes and genomics in hematological malignancies.

INTRODUCTION

Immune checkpoint blockade therapies are revolutionizing cancer therapy in several tumor types, demonstrating that the immune system can be successfully harnessed for effective anti-cancer treatment (Ribas and Wolchok, 2018). In hematological malignancies, immune checkpoint inhibition has demonstrated efficacy in classical Hodgkin lymphoma (CHL) with high PD-L1 expression

Significance

Immunotherapeutic strategies are actively investigated in hematological malignancies, yet the most relevant therapy targets and patient subsets remain unclear. In this study, we explored large-scale genomic datasets to uncover factors explaining immunological heterogeneity in hematological malignancies. We identified microenvironmental and genetic differences between immune-infiltrated and immune-excluded cancers and demonstrate how epigenetic makeup is linked to antigen presentation. Prognostic models highlighted the significance of immunological properties in predicting survival. Our findings thus underline the importance of integrating data of genetic and epigenetic aberrations, as well as the tumor microenvironment (TME), for a complete understanding of factors that may impact immunotherapy responsiveness. This understanding has implications for the development of precision immune intervention strategies in hematological malignancies.

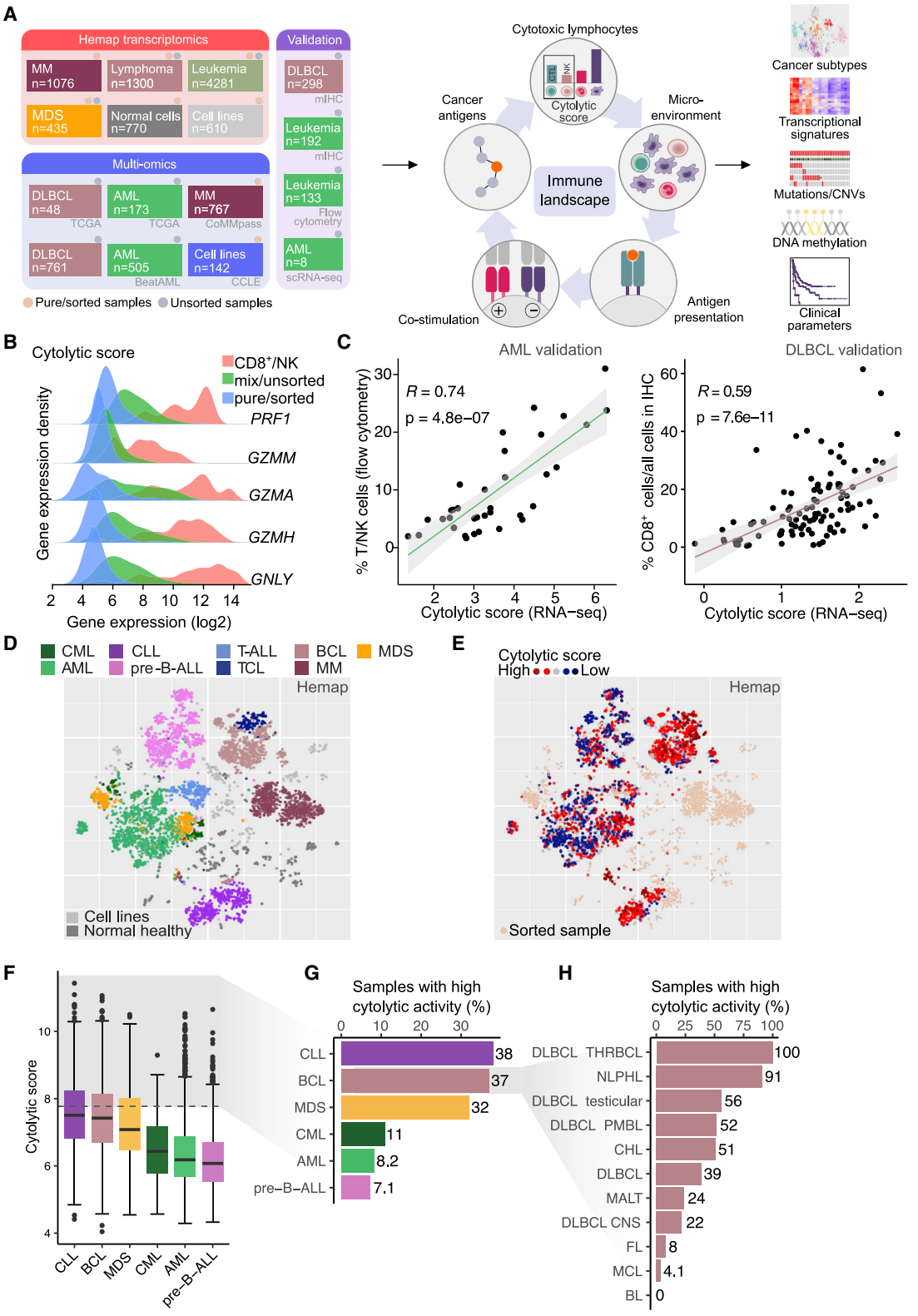


Figure 1. Identification of Cytotoxic Lymphocyte Infiltration in Hematological Malignancies

(A) Schematic overview and case numbers of used datasets.

(B) Distributions of log2 gene expression levels of genes included in cytolytic score for sorted and unsorted cancer cells, and sorted CD8+ and NK cells.

(legend continued on next page)

(Ansell et al., 2015). Moreover, chimeric antigen receptor T cell therapy has been successful in several B cell malignancies (Maude et al., 2014; Schuster et al., 2017). Allogeneic hematopoietic stem cell transplantation (allo-HSCT) is also considered to rely on the immune system by inducing the graft-versus-leukemia effect (Casucci et al., 2013). In contrast, responses to immune checkpoint inhibition in diffuse large B cell lymphoma (DLBCL) (Ansell et al., 2019), follicular lymphoma (FL) (Kline et al., 2020), chronic lymphocytic leukemia (CLL) without Richter's transformation (Younes et al., 2019), and multiple myeloma (MM) (Mateos et al., 2019), have been limited. As only some cancer types or a subset of patients within a cancer type typically achieve responses to immunotherapies, rational patient selection based on the immune milieu of each tumor type is crucial to achieve optimal benefit. However, in hematological cancers the immunological diversity and underlying mechanisms resulting in distinct immune landscapes are unclear.

The cancer immune landscape comprises various elements influencing the anti-cancer immune response (Chen and Mellman, 2013, 2017). The composition of the immune infiltrate, importantly cytotoxic lymphocytes that mediate elimination of cancer cells, has been associated with favorable outcomes (Fridman et al., 2012; Galon et al., 2006) and with immunotherapy responses (Tumeh et al., 2014; van Allen et al., 2015). Furthermore, antigen presentation is essential for adaptive immune responses, demonstrated by somatic mutations in the human leukocyte antigen (HLA) genes in solid tumors (Garrido et al., 2010; Shukla et al., 2015). Hematopoietic cancer cells may present antigen also in the context of HLA class II (Bachireddy et al., 2015). HLA II expression has been linked to prognosis (Rimsza et al., 2004) and response to PD-1 blockade in lymphoma (Roemer et al., 2018), demonstrating the importance of HLA II in hematological malignancies. In addition to antigen presentation, activating and inhibitory signals from tumor cells and antigen-presenting cells (APCs) regulate T and natural killer (NK) cell responses (Chen and Flies, 2013). Finally, cancer antigens, including neoantigens derived from somatic mutations and cancer-germline antigens (CGAs) (Simpson et al., 2005), are essential to initiate and maintain the adaptive immune response.

Emerging evidence from solid tumors suggests that cancer cell-intrinsic genetic and epigenetic aberrations influence tumor immune landscapes (Wellenstein and de Visser, 2018). Recently, several studies have integrated the genetics of solid tumors with immunological properties by leveraging extensive genomic datasets (Charoentong et al., 2017; Gentles et al., 2015; Li et al., 2016; Rooney et al., 2015; Thorsson et al., 2018). In contrast, large-scale studies of genotypic-immunophenotypic connections in hematological malignancies have not been conducted. Here, we perform a comprehensive immunogenomic analysis in hematological cancers, investigating cytotoxic immune infiltration, antigen presentation, immune cell co-stimulation, and cancer antigen expression patterns in relation to cancer subtypes and genomics.

RESULTS

Assessment of Cytotoxic Lymphocyte Infiltration across Hematological Malignancies

We used 7,092 samples from 36 hematological malignancies, with 770 healthy donor hematological cell populations and 610 cell lines as controls (Pölönen et al., 2019), to comprehensively analyze immunological properties in hematological cancer transcriptomes (Figures 1A and S1A; Table S1).

$CD8^+$ cytotoxic T lymphocytes (CTLs) and NK cells are considered essential for effective anti-tumor immunity and response to immunotherapy (Joyce and Fearon, 2015; Morvan and Lanier, 2016; Tumeh et al., 2014; van Allen et al., 2015). We first aimed to quantify the cytolytic immune infiltrate in the tumor microenvironment (TME) from bulk transcriptomes across hematological malignancies using genes specifically expressed in CTLs and NK cells (Figure S1B). Based on the high specificity of the genes *GZMA*, *GZMH*, *GZMM*, *PRF1*, and *GNLY* to CTLs/NK cells compared with hematopoietic cancer cells and their essential role in cytolytic effector functions, we defined the geometric mean of these five genes as the cytolytic score reflecting CTL/NK abundance (Figures 1B, S1B, and S1C).

To validate our strategy for inferring cytotoxic lymphocyte abundance, we analyzed T and NK cell fractions using flow cytometry from AML bone marrow (BM) and performed paired RNA sequencing (RNA-seq). Cytolytic score correlated highly with the combined fraction of T and NK cells out of all BM cells, indicating good performance in leukemia samples (Figures 1C, S1D, and S1E). Cytolytic score also correlated with the immunohistochemistry-based T cell content in DLBCL (Figure 1C) and mucosa-associated lymphoid tissue lymphoma (Figure S1F). Furthermore, cytolytic score agreed with previous methods of estimating immune cell subset abundance (Bindea et al., 2013; Becht et al., 2016) (Figures S1G and S1H). Correlation to the deconvolution method CIBERSORT (Newman et al., 2015), designed to infer relative fractions rather than immune cell type abundance, was lower (Figure S1I). This can be attributed to the similarity of hematologic cancer cells to reference immune cell types. In conclusion, cytolytic score robustly estimated the abundance of CTLs and NK cells in bulk transcriptomes of hematological malignancies, enabling its use for immunogenomic analyses.

Across hematological malignancies, we observed the highest cytolytic score in CLL, B cell lymphomas (BCL), and myelodysplastic syndrome (MDS) (Figures 1D–1F; Table S1). In contrast, acute leukemias and chronic myeloid leukemia (CML) were characterized

(C) Correlation of cytolytic score from RNA-seq and combined T and NK cell fraction by flow cytometry from 35 BM samples (AML) or $CD8^+$ T cell fraction on a tissue microarray spot from 100 samples (DLBCL). Gray shading represents the 95% confidence interval.
(D) t-SNE plot of Hemap cancer types, cell lines, and normal cell populations.
(E) Cytolytic score colored on the Hemap t-SNE map.
(F) Cytolytic score across main cancer types in Hemap. Gray shading indicates samples with Z score > 1. Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 × IQR from the quartiles.
(G) Percentages of samples with high cytolytic score (Z score > 1) across main cancer types.
(H) Percentages of samples with high cytolytic score (Z score > 1) across Hemap BCL subtypes.
See also Figure S1 and Table S1.

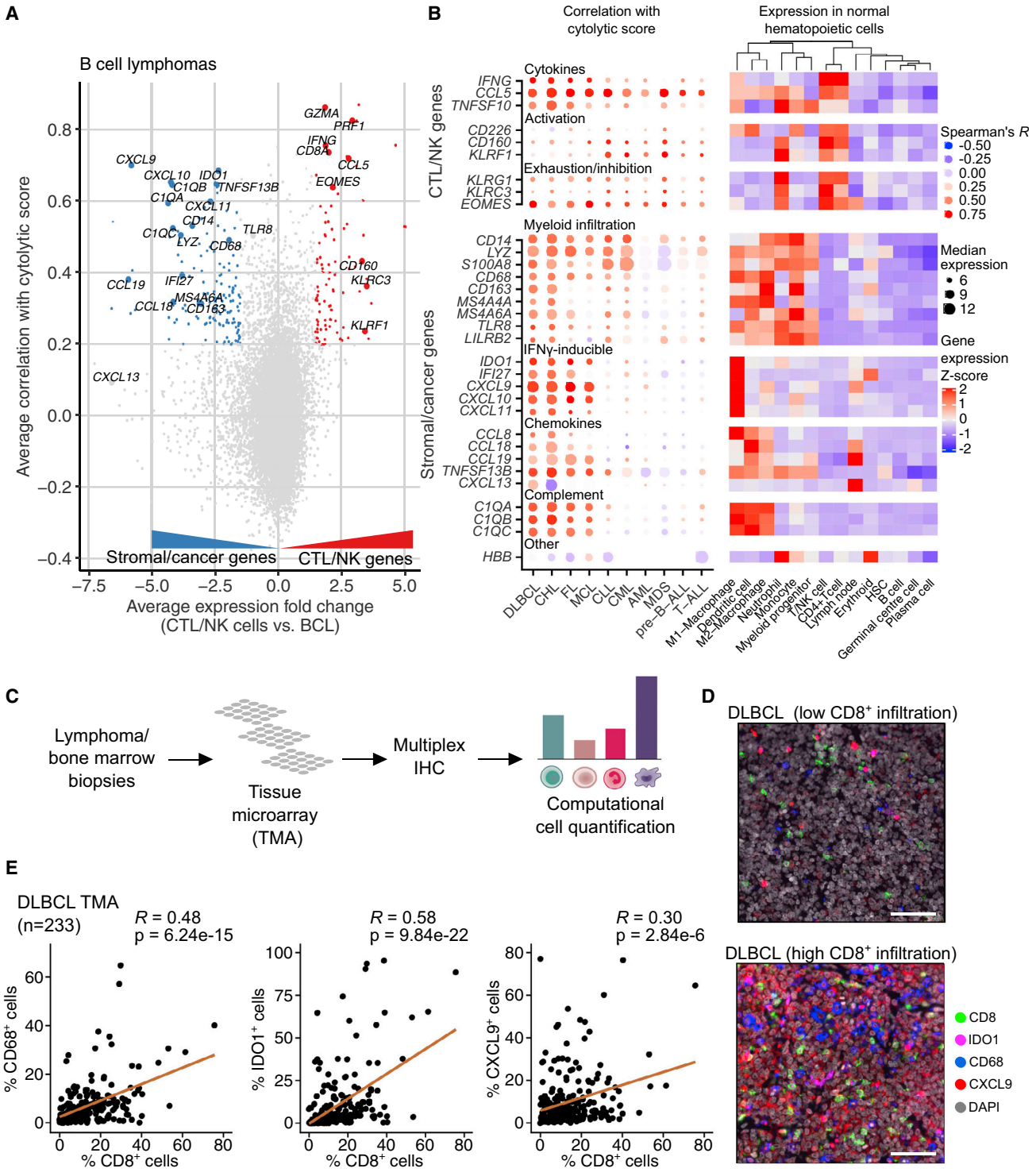


Figure 2. Distinct Immune TME Associates with Cytolytic Infiltration in Lymphoma and Leukemia

(A) Correlation of gene expression with cytolytic score in BCL. Correlation between expression of each gene and cytolytic score in BCL samples (y axis) is compared with the expression fold change between purified CTLs/NK cells and bulk BCL transcriptomes (x axis). Genes more specific to CTLs/NK cells are colored in red and genes more specific to other cell types in the tumor sample (stromal/cancer cells) are colored in blue.

(B) Dot plot of correlation of cytolytic score with selected genes across Hemap lymphoma and leukemia subtypes (left) and heatmap of mean expression in normal cell types (right).

(legend continued on next page)

by lower scores. Importantly, we observed substantial variation in cytolytic infiltrate within cancer types, with most cancer types, including acute leukemias, harboring a subset of samples with high cytolytic score (Figure 1G). Within BCL, we observed the highest levels in T cell/histiocyte-rich BCL (THRBCL) and CHL, and the lowest in Burkitt lymphoma (Figure 1H), consistent with known characteristics of the immune infiltrate across BCL subtypes (Scott and Gascoyne, 2014). Activated B cell (ABC)-like DLBCL showed higher cytolytic score compared with the germinal center B cell (GCB)-like subtype. Testicular DLBCL demonstrated high cytolytic score, in contrast to central nervous system DLBCL, indicating differences related to tumor site. These data show that cytolytic score captures variation in cytolytic infiltrates across hematological malignancies and indicate that, even in disease entities with generally low cytolytic activity, a subset of cases with abundant cytotoxic lymphocyte infiltration can be identified.

Interferon-γ Signature Linked to Cytolytic Infiltration Distinguishes Lymphoma TME from Leukemias

To characterize cancers with abundant cytolytic infiltrate in more detail, we explored genes whose expression correlated with cytolytic score. As expected, these genes were enriched in signatures reflecting T cell activation and inflammatory response, also confirmed at the protein level (Figures S2A and S2B; Table S2).

To dissect genes expressed in different cell types, we contrasted the transcripts correlated with cytolytic score with the difference in expression between purified CTLs/NK cells and the unsorted tumor samples (Figures 2A and S2C; Table S2). We also investigated which normal cell types expressed the identified genes to define cell types co-infiltrating with cytolytic cells (Figure 2B; Table S2). Genes expressed in monocytes and macrophages correlated highly with cytolytic score both in BCL and chronic leukemias (e.g., *CD14*, $R > 0.6$, false discovery rate [FDR] = 0.0), suggesting frequent co-infiltration. Single-cell analysis confirmed the myeloid cell origin of these genes, such as *LYZ* and *S100* genes expressed in monocytes in CLL, whereas genes negatively correlated with cytolytic score were expressed in cancer cells (Figures S2D and S2E). In contrast to BCL, CLL, and CML, correlations of TME genes with cytolytic score were much more modest in AML, pre-B cell acute lymphoblastic leukemia (pre-B-ALL), T cell ALL (T-ALL), and MDS (e.g., $CD14R \leq 0.3$) (Figure S2C). Genes expressed in T/NK cells, including activating receptors (*CD226/DNAM-1*, *CD160*, and *KLRF1/NKp80*) and inhibitory receptors or transcription factors related to exhaustion (*KLRG1*, *KLRC3/NKG2E*, and *EOMES*) correlated with cytolytic score across cancer types, consistent with both activation and exhaustion linked to infiltration by cytotoxic lymphocytes.

Genes associated with CTL infiltration in lymphomas included those encoding the immunosuppressive tryptophan-catabolizing enzyme *IDO1* (Munn and Mellor, 2016), T cell-recruiting chemokines highly expressed in proinflammatory M1-type macrophages (*CXCL9*, *CXCL10*, and *CXCL11*), and complement components expressed in macrophages and dendritic cells (*C1QA*, *C1QB*, and *C1QC*) (Figure 2B). The expression of *IDO1* and the CXCR3-ligand chemokines is known to be strongly induced by interferon-γ (IFN-γ) (Groom and Luster, 2011; Spranger et al., 2013), suggesting a TME response to IFN-γ associated with cytolytic infiltration. Expression of these genes in normal lymph nodes and both acute and chronic leukemias was low, indicating cancer-associated modulation of the lymphoma immune microenvironment. Consistently, *IFNG* was increased and showed a stronger correlation with cytolytic score in lymphomas compared with leukemias. Genes such as *CD163* and *CCL18* linked to M2 polarization of macrophages (Gordon, 2003), *TLR8*, and the immunoinhibitory *LILRB2* (Chen et al., 2018) distinguished the more heavily immune-infiltrated CHL and DLBCL from MCL and FL harboring lower cytolytic score. *CXCL13* was associated with low cytolytic infiltration specifically in CHL.

To validate the distinct lymphoma TME characterized by an IFN-γ-induced expression signature and myeloid infiltration, we analyzed tissue microarrays from DLBCL and AML BM biopsies (Figure 2C). Consistent with the gene expression data, CTLs ($CD8^+$) correlated with macrophages ($CD68^+$) and $IDO1^+$ and $CXCL9^+$ cells in DLBCL (Figures 2D and 2E). In AML, however, $IDO1^+$ or $CXCL9^+$ cells were generally sparse and did not correlate with CTLs (Figure S2F). These data indicate that macrophage/monocyte infiltration is associated with cytolytic cells and a distinct immunological TME in lymphomas characterized by IFN-γ-responsive genes.

Cytolytic Infiltration Is Associated with Genetic Alterations and Molecular Subtypes

We next asked whether specific genetic alterations or molecular subtypes could be associated with increased abundance of cytotoxic lymphocytes. We examined cytolytic score in relation to mutations and copy number variations (CNVs) in DLBCL (Figure 3A; Table S3) (Chapuy et al., 2018), which showed high cytolytic activity in comparison with other cancers. *BCL2* translocations found in GCB DLBCL, correlated negatively with cytolytic infiltration, consistent with the lower cytolytic score observed in this molecular subtype and fewer $CD8^+$ cells assessed by multiplexed immunohistochemistry (mIHC) (Figure S3A). Several CNVs and GCB-associated mutations (*BCL2*, *KMT2D*, and *CREBBP*) were negatively associated with cytolytic infiltration, whereas mutations in *ETV6*, *ETS1*, and *DTX1* showed positive correlations. *CREBBP*-mutated cases showed lower cytolytic infiltration also in another DLBCL dataset (Table S3) (Reddy et al., 2017). Given the strong impact of the molecular subtype on cytolytic infiltrate, we analyzed both ABC and GCB subtypes separately (Figures S3B and S3C). 7q amplifications were preferentially found in GCB, and in ABC with low cytolytic infiltration (Figure 3B). The correlation of *DTX1* mutations with cytolytic infiltration was even stronger in the GCB subtype alone compared with all DLBCL (Figure 3C).

We next examined cytolytic score in relation to mutations and CNVs in The Cancer Genome Atlas (TCGA) AML dataset (Table

(C) Workflow of mIHC validation of protein-level expression of genes correlating with cytolytic score.
(D) mIHC images of representative DLBCL samples with low or high percentage of $CD8^+$ cells. Scale bars, 50 μm.
(E) Scatterplots of the percentages of $CD8^+$, $CD68^+$, $IDO1^+$, and $CXCL9^+$ cells out of total cells in the DLBCL IHC cohort (n = 233).
See also Figure S2 and Table S2.

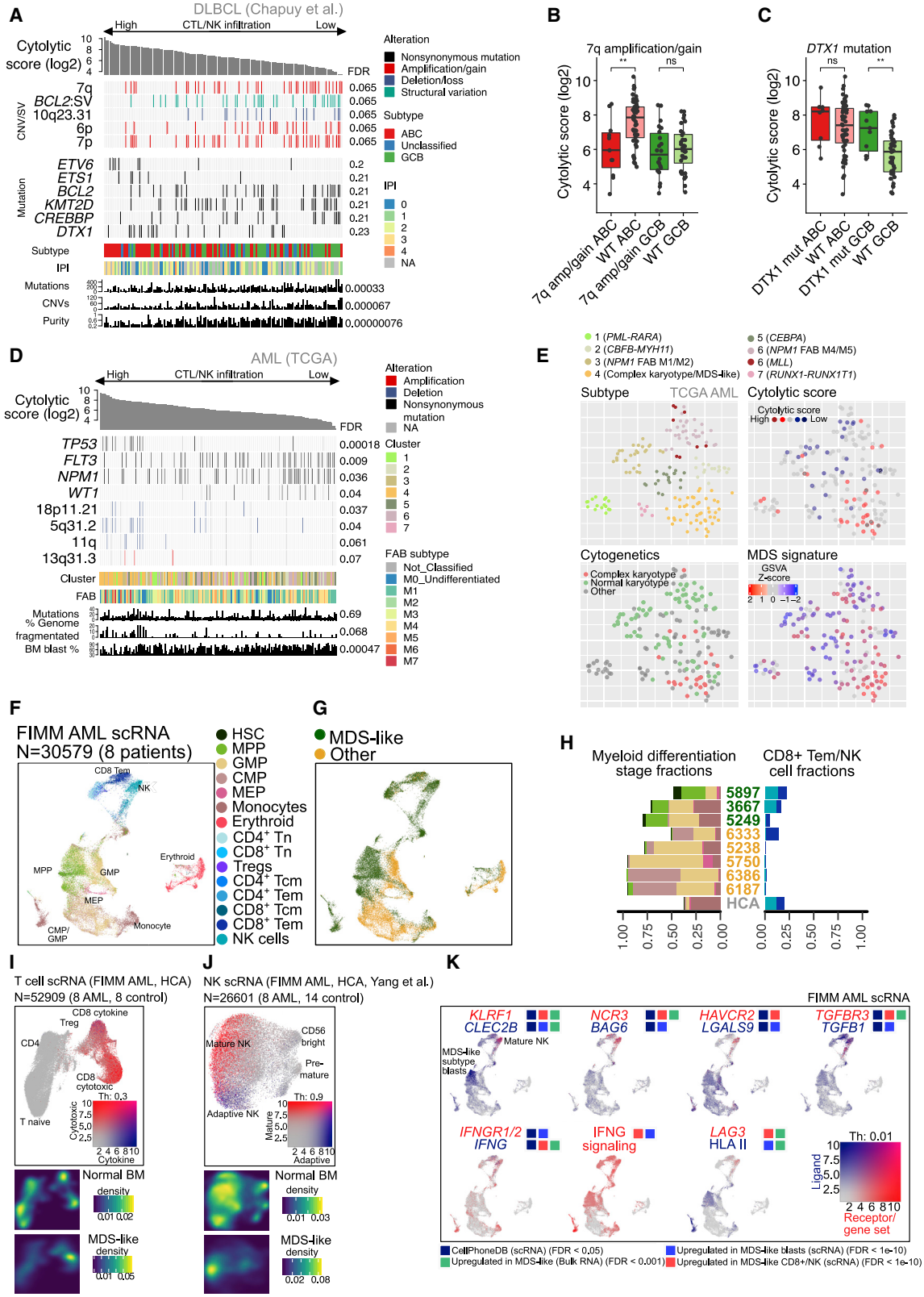


Figure 3. Genetic Alterations Associated with Cytolytic Infiltration

(A) Top genetic alterations associated with cytolytic score in DLBCL.

(B) Cytolytic score in DLBCL stratified by 7q amplification and molecular subtype.

(legend continued on next page)

S3). Cytolytic score positively correlated with mutations in the *TP53* tumor suppressor gene, as well as deletions located in the long arm of chromosome 5 (Figure 3D). These alterations often co-occurred with complex cytogenetics and increased genome fragmentation, whereas no correlation to mutation load was detected. In contrast, the common AML driver mutations *FLT3* and *NPM1* preferentially occurred in samples with low cytolytic activity. The samples with high cytolytic score were enriched in a cluster characterized by an MDS-like transcriptomic phenotype (Figure 3E, FDR = 10^{-6}, Fisher's exact test) that we have previously identified (Mehtonen et al., 2019).

To validate the association of cytolytic infiltration to an MDS-like transcriptomic subtype, we identified matching transcriptional clusters in the Hemap AML and BeatAML (Tyner et al., 2018) datasets (Mehtonen et al., 2019) (Figures S3D–S3G). The cases with high cytolytic score were enriched in a cluster corresponding to TCGA MDS-like cluster (FDR = 0.0044, Fisher's exact test in BeatAML) with frequent complex cytogenetics and prior MDS cases. Cytolytic score correlated with diagnosis of AML with myelodysplasia-related changes (FDR = 0.05), further suggesting a link between an MDS-like/secondary AML subtype and increased cytolytic infiltration (Table S3). The MDS-like subtype was associated with MDS-related mutations, such as *RUNX1*, *TP53*, *U2AF1*, and *SRSF2* in the BeatAML dataset (Figure S3H). We further confirmed the associations between low versus high T/NK cell infiltrate in AML with *FLT3*, *NPM1* mutations or 5q deletion, respectively, using flow cytometry (Figure S3I). These data suggest that specific cancer cell-intrinsic genetic alterations and molecular subtypes are linked to cytotoxic infiltrate both in DLBCL and AML.

To better understand the underlying mechanisms of increased cytolytic infiltrate in the MDS-like subtype, we used single-cell RNA-seq (scRNA-seq) to compare three samples harboring an MDS-like signature and MDS-related *RUNX1*, *ASXL1*, and *BCOR* mutations with five other AML samples (Figure S3J). AML blasts from MDS-like samples were classified primarily as HSC or progenitor-like cells, such as multipotent progenitors (MPPs), megakaryocyte-erythroid progenitors, or granulocyte-monocyte progenitors (GMPs), in contrast to common myeloid progenitors (CMPs) and GMPs in other samples (Figures 3F–3H). Accordingly, the genes of the MDS-like signature were mainly associated with stem cell-like cells in both AML and Human Cell Atlas normal BM (Figures S3K–S3L; Table S3). As expected, we observed a high frequency of NK and $CD8^+$ T cells in the MDS-like scRNA samples compared with other AMLs (Figure 3H). MDS-like AML T cells were biased toward a cytotoxic and T effector memory (Tem)-like phenotype (Szabo et al., 2019) (Figures 3I and S3M). NK cells in MDS-like patients were mostly mature NK cells clustering close to adaptive NK cells (Yang et al., 2019) (Figures 3J and S3N).

We next sought to identify potential receptor-ligand pairs between AML blasts and cytotoxic cells. Using CellPhoneDB, we identified 41 and 47 candidate interactions between mature NK or cytotoxic/cytokine $CD8^+$ Tem cells and blasts in MDS-like AML samples (Table S3). The ligands *CLEC2B* and *BAG6* were expressed in MDS-like subtype AML blasts with concomitant expression of their respective activating receptors *KLRF1* and *NCR3* in NK cells, which could be linked to the higher NK cell infiltration and NK cell regulation (Figure 3K). *CLEC2B* expression was significantly associated with the cluster composed of MPP cells and with the MDS-like subtype in Hemap and BeatAML (Table S3), suggesting that the blast differentiation stage in different AML subtypes may influence the immune interactions. As a sign of exhaustion potentially balancing the activation, both $CD8^+$ Tem and NK cells expressed the inhibitory receptor *LAG3*, and NK cells additionally expressed *HAVCR2 (TIM-3)* and *TGFBR3*. *LAG3* and *TGFBR3* were also significantly upregulated in BeatAML, Hemap, and TCGA for MDS-like cases (Table S3). Their respective ligands HLA II, *LGALS9*, and *TGFB1* were expressed on MDS-like AML blasts (Figure 3K). Interestingly, the *IFNG* signaling gene set was upregulated in MDS-like AML blasts, while *IFNG* was mainly expressed by $CD8^+$ Tem cells, suggesting AML blast response to IFN-γ secreted by cytolytic cells (Figures 3K and S3O). Taken together, scRNA-seq revealed receptor-ligand interactions between cytotoxic lymphocytes and blasts of distinct differentiation stages in MDS-like AML patients that could underlie the increased cytolytic infiltrate.

Epigenetic Modification of the HLA Class II Transactivator *CIITA* Regulates Antigen Presentation

We next analyzed the expression of HLA genes to detect potential transcriptional downregulation driving immune evasion. We constructed an HLA I score, comprised of *B2M*, *HLA-A*, *HLA-B*, and *HLA-C*, and an HLA II score, containing HLA II genes significantly upregulated in APCs highly correlating with each other (Figures S4A and S4B; Table S4).

(C) As in (B) for *DTX1* mutations.
(D) Top genetic alterations associated with cytolytic score in TCGA AML.
(E) t-SNE plot of TCGA AML clusters, cytolytic score, cytogenetics, and MDS signature.
(F) UMAP plot of cell types identified by SingleR in eight AML scRNA-seq samples.
(G) Location of cells from MDS-like AML cases (samples 3,667, 5,897, and 5,249) and other AML cases (samples 6,333, 5,238, 5,750, 6,386, and 6,187) on the UMAP plot.
(H) Myeloid lineage differentiation stages and $CD8^+$ Tem and NK cell proportions for each AML and Human Cell Atlas normal BM sample.
(I) UMAP plot of the average expression for $CD8^+$ cytotoxic (red) and cytokine gene sets (blue) (Szabo et al., 2019) (top panel). Density of normal BM T cells and MDS-like AML T cells on the UMAP is shown below.
(J) UMAP plot of the average expression for NK mature and adaptive gene sets (Yang et al., 2019). The normal BM NK cell and MDS-like AML NK cell densities are shown as in (I).
(K) Gene expression for selected significant ligand-receptor pairs and related gene sets on the FIMM AML UMAP plot. Significant associations identified by different analyses are annotated for each feature as colored squares.
Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 × IQR from the quartiles. See also Figure S3 and Table S3.

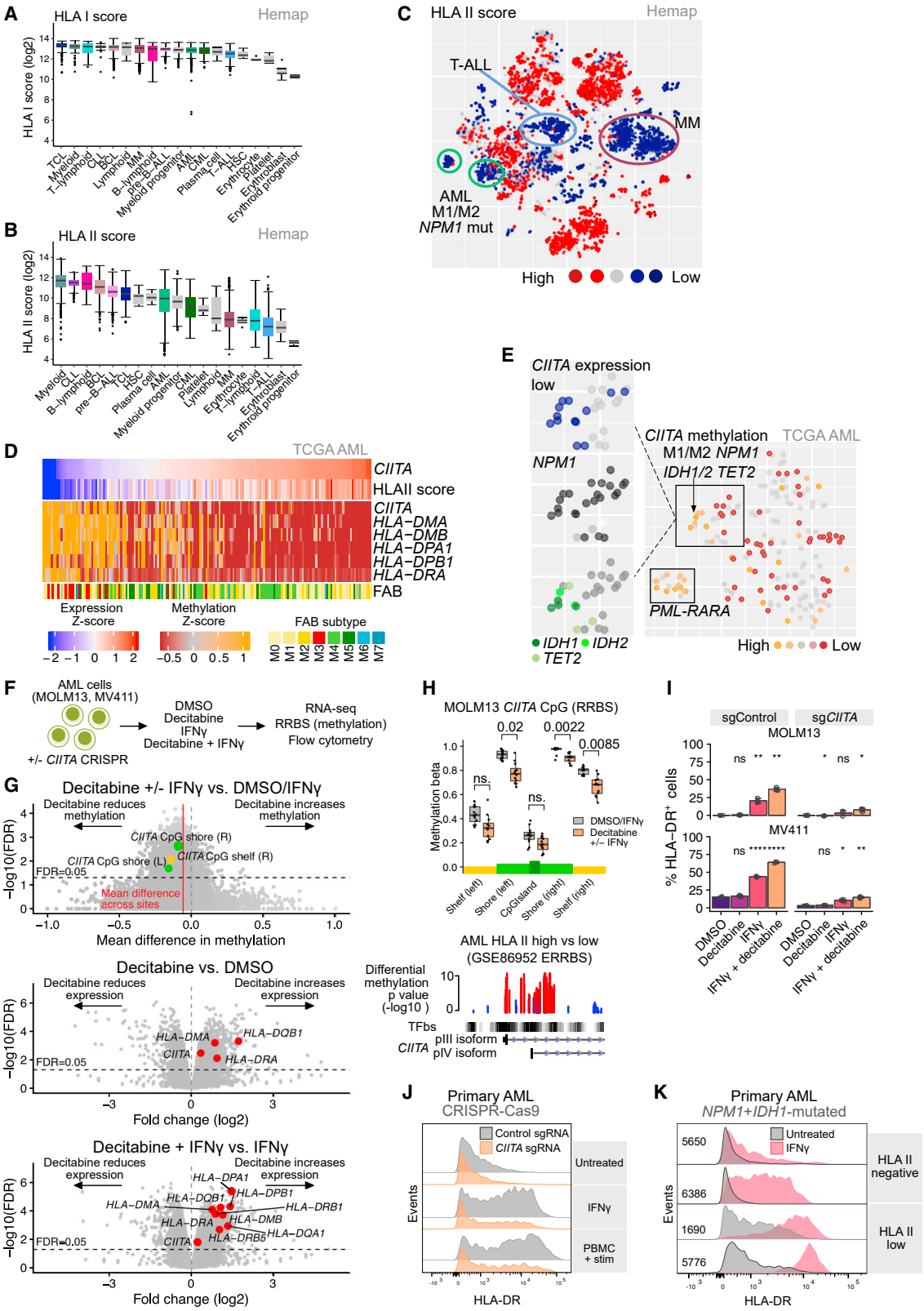


Figure 4. Expression of Antigen-Presenting HLA Genes Is Linked to Molecular Subtypes and Epigenetic Regulation

(A) HLA I score in main cancer types and normal cell populations in Hemap.

(B) HLA II score as in (A).

(legend continued on next page)

We observed a lower HLA I score in cells of the erythroid lineage, hematopoietic progenitors, and T-ALL compared with other cell populations (Figure 4A; Table S1). While the differences in HLA I expression were rather modest, HLA II score varied more substantially (Figure 4B), accurately reflecting variation in surface HLA II protein levels (Figure S4C). B cell malignancies, including pre-B-ALL, CLL, and BCL, had high HLA II score as expected by their APC origin, whereas in MM HLA II was downregulated consistent with HLA II loss upon plasmacytic differentiation (Silacci et al., 1994) (Figures 4B and 4C; Table S1). In AML, specific transcriptomic clusters showed downregulated HLA II expression, including the acute promyelocytic leukemia cluster harboring PML-RARA fusion known to be characterized by low surface HLA-DR (Wetzler et al., 2003), and a cluster characterized by *NPM1* mutations and M1 or M2 FAB subtype (Figure 4C).

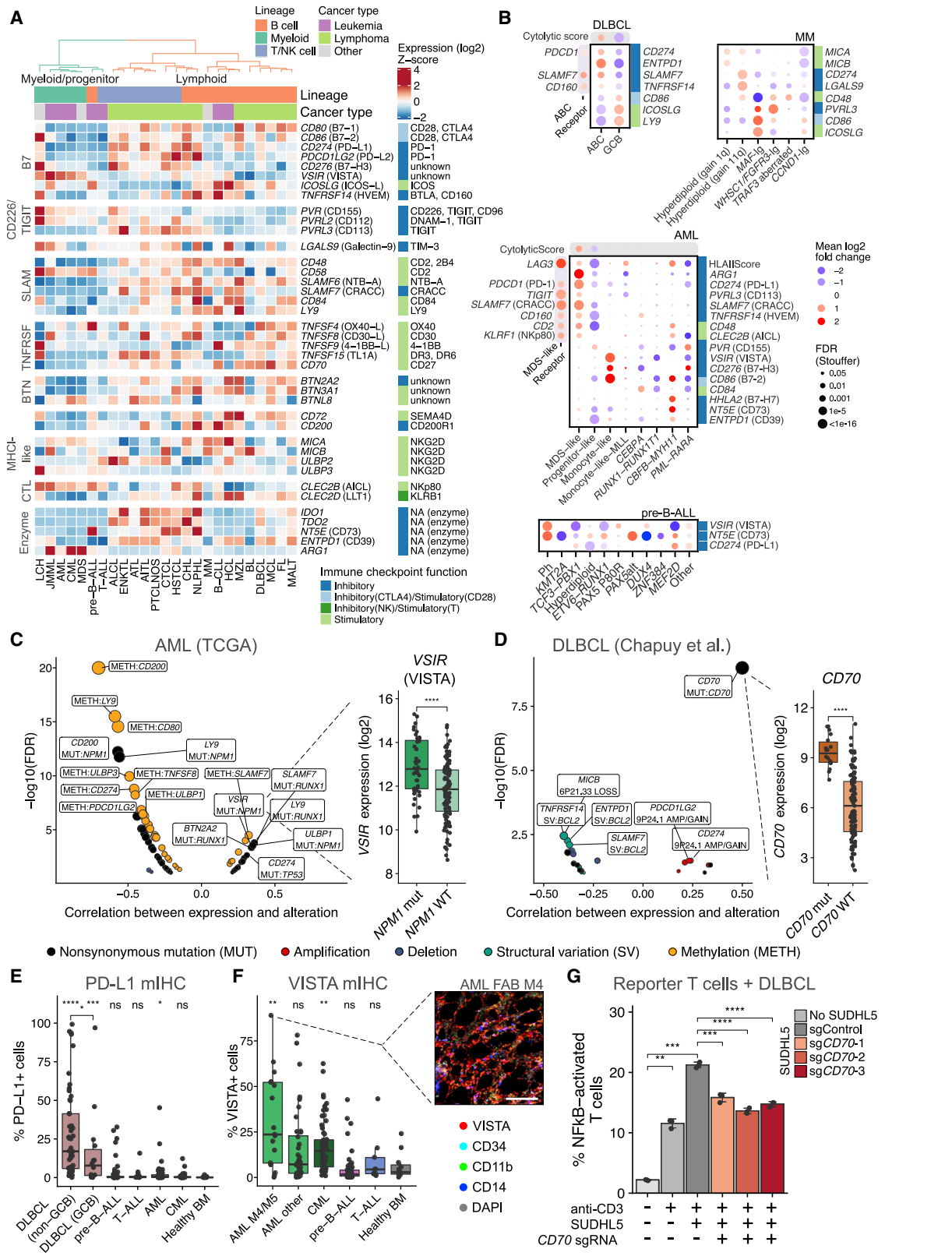
We next correlated molecular features, including mutations, CNVs, and DNA methylation in TCGA AML cohort to the HLA II score to shed light on the molecular mechanisms leading to downregulation of the HLA II genes (Table S4). Expression of the HLA class II transactivator *CIITA* strongly correlated with the HLA II score ($R = 0.84$, FDR = 6.1×10^{-43}, Figure 4D). Furthermore, methylation of promoter regions of *CIITA* ($R = -0.54$, FDR = 4.6×10^{-10}, probe cg01351032) and several *HLA II* genes correlated negatively with the HLA II score. *CIITA* methylation was enriched in transcriptomic clusters with low HLA II score corresponding to those identified in Hemap, harboring PML-RARA fusion or M1/M2 FAB subtype co-occurring with mutations in *NPM1* (Figures 4E and S4D). Upon closer examination, highest *CIITA* hypermethylation occurred in a subcluster with mutations in the DNA methylation regulators *IDH1*, *IDH2*, and *TET2* (Figure 4E). In contrast, AML harboring CBFB-MYH11 or RUNX1-RUNX1T1 translocations were characterized by high HLA II and *CIITA* hypomethylation, and *RUNX1* mutations correlated with high HLA II score (Table S4).

Validation in an independent dataset (Glass et al., 2017) demonstrated a differentially methylated region between AML patients with high and low HLA II score encompassing a CpG island and *CIITA* promoter III, active in lymphocytes, and the IFN-γ-inducible promoter IV (Muhlethaler-Mottet et al., 1997) (Figures S4E and S4F). We observed a similar clustering of epigenetic modifier mutations in cases with low *CIITA* and HLA II score in the BeatAML dataset (Figure S4G). scRNA-seq verified the low *CIITA* and HLA II expression in AML blasts with *NPM1* and *IDH1*/*TET2* mutations, which were mainly classified as CMP and GMP cells (Figures S4H, S4I, and 3B). In addition to AML, *CIITA* was methylated in T-ALL with low HLA II expression (Holling et al., 2004), suggesting potential epigenetic regulation of antigen presentation in various hematological cancer types (Figures S4J–S4K).

Given the potential epigenetic nature of HLA II silencing, we tested whether the hypomethylating agent decitabine could reduce *CIITA* methylation and induce HLA II expression in AML cell lines (Figure 4F). Decitabine induced a global demethylation in MOLM13 cells, including significant changes at the CpG island at the *CIITA* promoter (Figures 4G, 4H, and S4L). Consistently, RNA-seq demonstrated increased *CIITA* and HLA II gene expression with decitabine treatment in MOLM13 and MV411 cells, and further increase from initial upregulation by treatment with IFN-γ, a known inducer of HLA II (Steimle et al., 1994) (Figures 4G and S4M). *CIITA* disruption by CRISPR/Cas9 in MOLM13 and MV411 resulted in reduced HLA II induction by decitabine or IFN-γ, indicating that *CIITA* mediates the HLA II upregulation in response to these stimuli (Figures 4I and S4N). Interestingly, HLA II was effectively induced by IFN-γ and/or decitabine in AML cell lines with moderate *CIITA* methylation, such as MOLM13, whereas cells with high *CIITA* methylation, such as NB4 were resistant to HLA II upregulation, suggesting that CIITA methylation may limit antigen presentation by AML cells despite IFN-γ stimulation (Figures S4O and S4P).

CRISPR-mediated silencing of *CIITA* in primary AML cells inhibited HLA II induction in the presence of IFN-γ or activated T cells, further suggesting that *CIITA* silencing by means such as methylation can prevent activation of HLA II-mediated antigen presentation upon T cell encounter (Figure 4J). Consistently, primary AML cells from patients harboring *NPM1* and *IDH1* mutations with initially absent HLA II expression showed only modest HLA II induction upon IFN-γ treatment, indicating impaired antigen presentation in the *CIITA*-methylated AML subgroup even in the presence of activating cytokines (Figure 4K). These data show that AML cells may evade antigen presentation through transcriptional downregulation of *HLA II* genes that mechanistically could result from *CIITA* methylation in distinct genetic and

(C) HLA II score colored on Hemap t-SNE map. Clusters with low HLA II score are circled.
(D) *CIITA* expression, HLA II score, and methylation of *CIITA* and *HLA II* genes in TCGA AML.
(E) t-SNE plot of *CIITA* methylation (TCGA AML).
(F) Schematic of CRISPR-mediated *CIITA* silencing and IFN-γ/hypomethylating treatment experiments in AML cell lines.
(G) Volcano plots comparing MOLM13 AML cells with and without decitabine treatment. Methylation at CpG island regions by RRBS (top), gene expression without IFN-γ treatment (middle), and gene expression with IFN-γ treatment (bottom). Significantly differentially methylated CpG regions close to *CIITA* promoter are colored with green and yellow (top), and significantly differentially expressed HLA II genes and *CIITA* are colored with red (middle and bottom).
(H) Methylation at CpG regions close to *CIITA* promoter with and without decitabine treatment in MOLM13 AML cells. The t test FDR values across all tested sites genome-wide are shown. Below, significantly differentially methylated cytosines comparing AML patient samples with low and high HLA II score in the same region as −log10 p values with red and blue colors indicating hypermethylated and hypomethylated cytosines. Transcription factor binding sites from ENCODE (TFbs) and *CIITA* transcripts pIII and pIV are shown.
(I) Percentages of HLA-DR$^+$ MOLM13 or MV411 AML cells expressing control or *CIITA* single-guide RNAs (sgRNAs) after indicated treatments. Data are mean ± SD, dots represent technical replicates, and p values are from Welch's t test.
(J) HLA-DR expression in primary AML blasts (patient 4,374) expressing control or *CIITA* sgRNAs compared after treatment with IFN-γ, anti-CD3+CD28-stimulated peripheral blood mononuclear cells (PBMCs), or untreated.
(K) HLA-DR expression of primary AML blasts with *NPM1* and *IDH1* mutations after treatment with DMSO or IFN-γ. Samples are marked as HLA II negative or HLA II low based on HLA-DR baseline expression.
Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 × IQR from the quartiles. See also Figure S4 and Table S4.



(legend on next page)

transcriptional subtypes of AML and across hematological cancers.

Immune Checkpoint Ligands Are Linked to Cancer Subtypes and Genetic Alterations

We surveyed the expression of ligands for T and NK cell co-stimulatory and co-inhibitory receptors and other immunomodulators to identify potentially targetable immune checkpoints. An initial analysis of immune checkpoints highlighted distinct patterns of immunomodulatory genes in myeloid malignancies and acute leukemias in comparison with mature lymphoid malignancies (Figure 5A; Table S5). The observed clustering by lineage suggests that the cell-of-origin influences the repertoire of immunomodulatory genes expressed by cancer cells.

The inhibitory ligand *CD274* (PD-L1) was preferentially expressed in lymphomas compared with myeloid malignancies. Within cancer types, increased *CD274* expression characterized the more heavily cytolytic-infiltrated ABC subtype compared with GCB in DLBCL (Figure 5B), whereas in MM, *CD274* and other ligands for inhibitory checkpoints, such as *LGALS9* (galectin-9, ligand for TIM-3) were expressed in the hyperdiploid subtype with 11q gain (Figure 5B). Other inhibitory molecules found in mature lymphoid malignancies included the NK cell inhibitory receptor KLRB1 ligand *CLEC2D* (LLT1), the T cell inhibitory butyrophilin *BTN2A2*, and *PVRL3* encoding a ligand for the inhibitory receptor TIGIT in myeloma and T cell malignancies (Figures S5A and S5B).

In contrast, myeloid malignancies AML, CML, juvenile myelomonocytic leukemia (JMML), and MDS, highly expressed *VSIR (C10orf54*, VISTA*)*, encoding an inhibitory T cell checkpoint of the B7 family (Figure 5A). *ARG1*, encoding the immunosuppressive enzyme arginase and expressed particularly in MDS, JMML, and CML, represented another potential myeloid-specific immune evasion mechanism. Subtype-level analysis of AML revealed expression of *VISTA* and *CD86* in monocytic AML and several inhibitory genes (*CD274* and *ARG1*) and their corresponding receptors in the MDS-like subgroup (Figure 5B and S5B), highlighting the influence of the blast differentiation stage. *CLEC2B* (AICL), encoding an activating ligand for the NKp80 receptor, showed a myeloid-specific expression pattern, consistent with its expression in the HSC, MPPs, and GMPs found in MDS-like AML (Figure S5C). Langerhans cell histiocytosis highly expressed several immune checkpoint ligands of the B7 and TIGIT families, likely related to its origin as a professional APC. *NT5E*, encoding the immunosuppressive adenosine-producing enzyme CD73 (Beavis et al., 2012), was enriched in pre-B-ALL, especially in Ph/Ph-like and PAX5-altered subtypes (Figure 5B). Together, these results suggest cancer type and lineage as major determinants of immune checkpoint ligand expression in hematological malignancies.

To investigate potential mechanisms of immunomodulatory gene regulation, we examined correlations with DNA methylation in AML and DLBCL in TCGA dataset (Table S5). Comparison of AML and DLBCL revealed differential methylation at gene promoters linked to the cancer type-specific expression of several immunomodulators, such as *PDCD1LG2* and *CD80* (Figures S5D and S5E). Promoter methylation also correlated with expression of immunomodulators within cancer types. In AML, promoter methylation correlated negatively with expression of *CD200*, *CD274*, the NKG2D ligands *ULBP1* and *ULBP3*, and *PDCD1LG2* and *CD80* (Figure 5C). Similarly in DLBCL, expression of *ULBP1* and *CD200* correlated negatively with promoter methylation, suggesting that DNA methylation contributes to variation in immune checkpoints across cancer types (Figure S5F).

In addition to epigenetic modification, several recurrent genetic alterations were linked to distinct immune checkpoints (Table S5). In DLBCL, the co-stimulatory *CD70* was often mutated when highly expressed (Figure 5D), suggesting immune evasion through somatic mutations. Other alterations potentially enabling immune evasion included downregulation of *MICB*, encoding an activating ligand for the NKG2D receptor, through 6p21.33 losses containing *MICB*, and 9p24.1 amplifications or gains associated with increased expression of PD-1 ligands (Figure S5G). In AML, *NPM1* mutations were linked to increased expression of *VISTA* and the NKG2D ligand *ULBP1*, independent of monocytic morphology (Figures 5C and S5H). *RUNX1*-mutated AML highly expressed the B cell-associated *BTN2A2*, *SLAMF7*, and *LY9* in addition to HLA II, suggesting that the lineage infidelity and B lineage transcriptional program induced by *RUNX1* mutations influences also co-inhibitory signaling by

Figure 5. Immune Checkpoint Ligands Are Linked to Cancer Subtypes and Genetic Alterations

(A) Expression of immunomodulatory genes (Hemap).

(B) Expression of selected immunomodulatory genes in molecular subtypes of DLBCL, MM, AML, and pre-B-ALL. Genes shown are significant in at least two datasets. Both receptor and ligand expression are shown for subtypes with high cytolytic activity (ABC DLBCL, MDS-like AML).

(C) Volcano plot of correlations of immunomodulatory gene expression with genetic alterations and DNA methylation in TCGA AML. Dot size is proportional to the FDR value. *VSIR* expression stratified by *NPM1* mutation status.

(D) Volcano plot of correlations of immunomodulatory gene expression with genetic alterations in DLBCL as in (C). *CD70* expression stratified by *CD70* mutation status.

(E) Percentages of PD-L1$^+$ cells out of all cells in non-GCB DLBCL (testicular, n = 50), GCB DLBCL (testicular, n = 14), pre-B-ALL (n = 54), T-ALL (n = 14), AML (n = 70), CML (n = 59), and healthy BM (n = 14) tissue microarrays by mIHC. p value significance codes indicate comparisons to healthy BM and between non-GCB and GCB DLBCL.

(F) Percentages of VISTA$^+$ cells out of all BM cells in AML (n = 57), CML (n = 62), pre-B-ALL (n = 51), T-ALL (n = 9), and healthy BM (n = 11) tissue microarrays by mIHC. p value significance codes indicate comparisons to healthy BM. Inset shows BM from a patient with M4 FAB subtype AML with high VISTA expression. Scale bar, 100 μm.

(G) Percentages of NF-κB-activated reporter T cells upon co-culture with SUDHL5 DLBCL cells with or without CRISPR-mediated *CD70* disruption. Anti-CD3 indicates T cell stimulator cells expressing membrane-bound anti-CD3. Data are mean ± SD, dots represent technical replicates, and p values are from Welch's t test.

Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 × IQR from the quartiles. See also Figure S5 and Table S5.

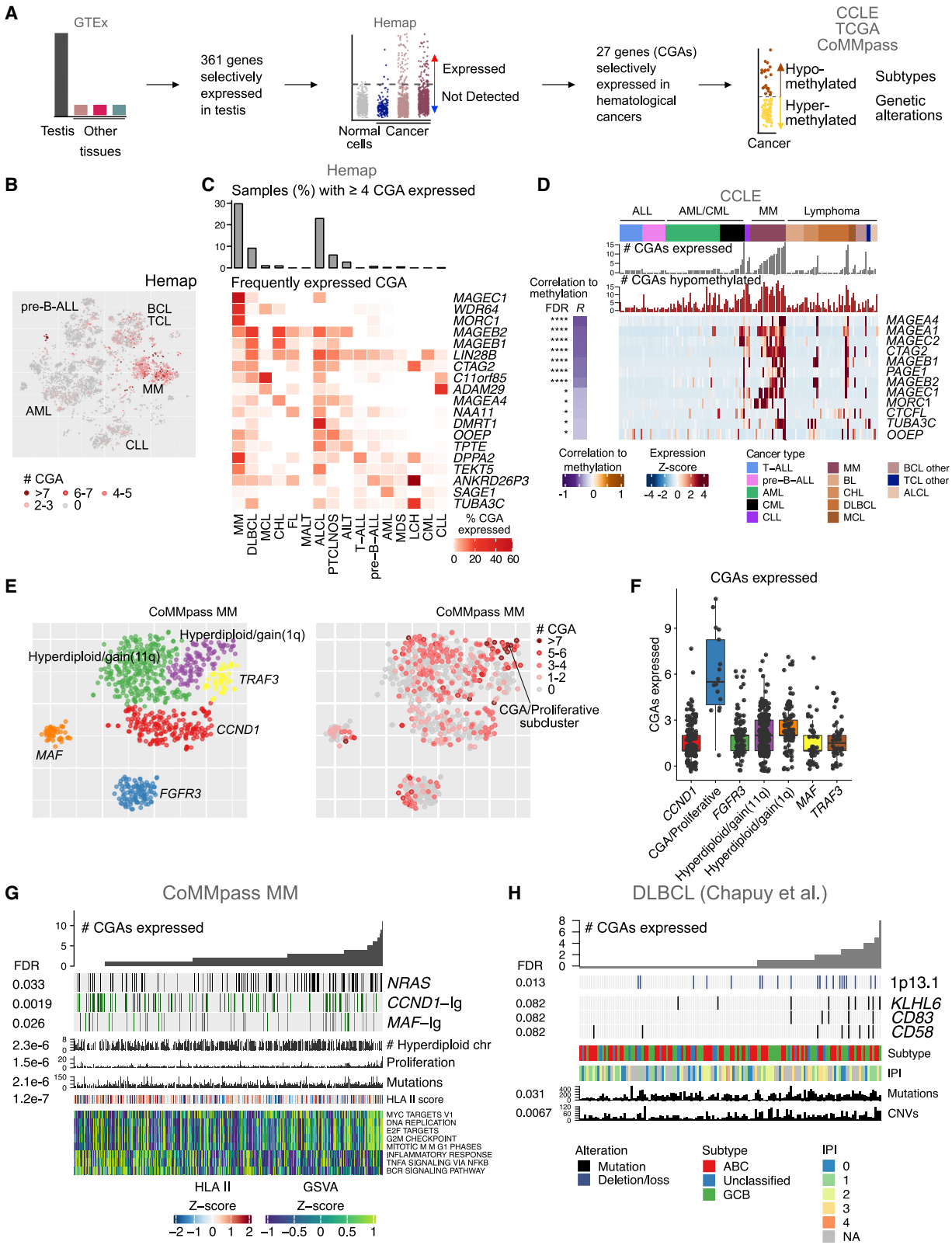


Figure 6. CGA Expression Is Frequent in MM and Linked to DNA Methylation

(A) Schematic of identification of genes with cancer-germline expression pattern.

(B) t-SNE plot of CGA expression (Hemap).

(legend continued on next page)

AML cells (Silva et al., 2009). *TP53* mutations were linked to higher *CD274* expression, potentially related to increased cytolytic activity. Thus, genetic alterations contribute to variation in immunomodulatory gene expression in AML and DLBCL.

Finally, we sought to validate the cancer type-specific immune checkpoints at the protein level and functionally. Quantitative mIHC confirmed increased PD-L1$^+$ cells in DLBCL, particularly in non-GCB tumors, compared with leukemias (Figure 5E). Moreover, VISTA expression was enriched in monocyte-like AML and CML BM (Figures 5F and S5I). In contrast to *VISTA* in myeloid malignancies, *CD70* expression was restricted to lymphoid cancers where it was associated with somatic mutations in the same gene. To investigate whether *CD70* in DLBCL cells influences T cell co-stimulation, we introduced inactivating *CD70* mutations using CRISPR/Cas9 in SUDHL5 DLBCL cells highly expressing CD70. Co-culture of SUDHL5 with CD27$^+$ Jurkat T cell receptor signaling reporter cells promoted nuclear factor κB (NF-κB) transcriptional activity together with stimulator cells expressing membrane-bound anti-CD3 (Figure 5G). However, with CRISPR-mediated *CD70* disruption, this response was significantly attenuated, indicating that *CD70* mutations can promote immune evasion by preventing CD27-mediated T cell co-stimulation. These data suggest that several immune checkpoints, such as PD-L1, VISTA, and CD70 are expressed in a cancer type-specific fashion and may be influenced by DNA methylation or genetic alterations.

Frequent Expression of CGAs in MM

To evaluate potential targets of the adaptive cytotoxic immune response, we investigated CGAs that have not been systematically studied in hematological malignancies. We integrated the Genotype-Tissue Expression (GTEx) (Melé et al., 2015) and Hemap data to define genes with a cancer-germline expression pattern by first selecting genes expressed exclusively in testis using GTEx and then requiring the genes to be expressed in at least 5% hematological cancers but not in normal hematopoietic cells (Figure 6A), recovering 27 CGA genes. Most are included in the CTdatabase (Almeida et al., 2009), which however contained several non-testis-restricted genes (Figure S6A).

CGAs were strikingly most frequently expressed in MM (Figures 6B and S6B). One-third of MM patients expressed more than four CGAs. Both B and T cell lymphomas also showed frequent CGA expression, whereas CGAs were largely transcriptionally silent in leukemias. Several CGAs were expressed in a cancer type-specific manner, including *MAGEC1*, *MORC1*, *DPPA1*, *COX7B2*, *PAGE1*, and *GAGE1* in MM, *ADAM29* in CLL (Vasconcelos et al., 2005), *SAGE1* in AML, *DMRT1* in anaplastic large-cell lymphoma, and *MAGEB2* and *MAGEB1* in DLBCL (Figures 6C and S6C).

To understand mechanisms leading to aberrant CGA expression, we studied whether alterations in DNA methylation were associated with activated CGA transcription. CGAs were frequently hypomethylated and expressed in MM cell lines in CCLE, consistent with primary MM samples (Figure 6D). Similarly, expression of the most frequent CGAs in DLBCL patients, *MAGEB1* and *MAGEB2*, was linked to hypomethylation (Figure S6D).

We next explored genetic alterations and transcriptional signatures correlated with the number of expressed CGAs (Table S6). Analysis of CoMMpass MM data (Manojlovic et al., 2017) revealed increased CGA expression in specific transcriptomic clusters of MM, including proliferative and hyperdiploid MM with 1q gain (Figures 6E, 6F, and S6E). Moreover, *NRAS* mutations, the number of hyperdiploid chromosomes, and mutation load correlated with CGA expression (Figure 6G). Consistent with enrichment in the proliferative cluster, CGA expression was associated with signatures reflecting cell-cycle activity and MYC targets, whereas inflammatory response and NF-κB signaling gene sets were downregulated in CGA-high MM. Interestingly, HLA II score correlated negatively with the number of expressed CGAs, implying potential immune evasion through HLA II downregulation in MM with increased CGAs. Associations to transcriptomic subtypes, HLA II score, and proliferation-related signatures were confirmed in Hemap MM (Figures S6F and S6G).

In DLBCL, number of CGAs correlated with mutation and CNV load and specific alterations, such as 1p13.1 deletions containing *CD58* and *CD58* mutations (Figure 6H), providing a potential immune evasion mechanism for CGA-expressing cancers through disruption of the CD2-CD58 interaction with T cells. In the ABC subtype, a mutational signature, including 6q deletions and *MYD88*, *HLA-A*, and *ETV6* mutations, resembling cluster 5 (Chapuy et al., 2018), was enriched in cases expressing multiple CGAs (Figure S6H), suggesting a link between a distinct genetic subtype and activation of germline-restricted genes. Cytolytic score and gene sets reflecting inflammatory response were downregulated in ABC DLBCL expressing multiple CGAs (Figures S6H and S6I). In the GCB subtype, 1p13.1 deletions and *KLHL6* mutations correlated with CGA expression, similarly as in all DLBCLs (Table S6). These data suggest that CGA expression is activated in myelomas and lymphomas harboring genomic aberrations or distinct genetic alterations associated with immune evasion, often involving promoter hypomethylation.

Cancer Subtype-Specific Immunological Features Are Associated with Survival

Finally, we aimed to delineate how immunological features are associated with overall survival. We first identified survival associations of individual immunological features using univariate Cox proportional hazards models in multiple datasets of AML, MM, and DLBCL (Figure 7A; Table S7). Tested features included cytolytic score, HLA scores, number of expressed CGAs and

(C) Heatmap of CGA expression (Hemap). Color indicates the percentage of samples expressing a given antigen.
(D) Expression and methylation of CGAs across cell lines of hematological malignancies (CCLE).
(E) t-SNE plot of molecular subtypes and CGA expression in MM.
(F) Numbers of expressed CGAs in molecular subtypes of MM. Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 × IQR from the quartiles.
(G) Genetic alterations and gene sets correlated with the number of expressed CGAs in MM.
(H) Genetic alterations correlated with the number of expressed CGAs in DLBCL.
See also Figure S6 and Table S6.

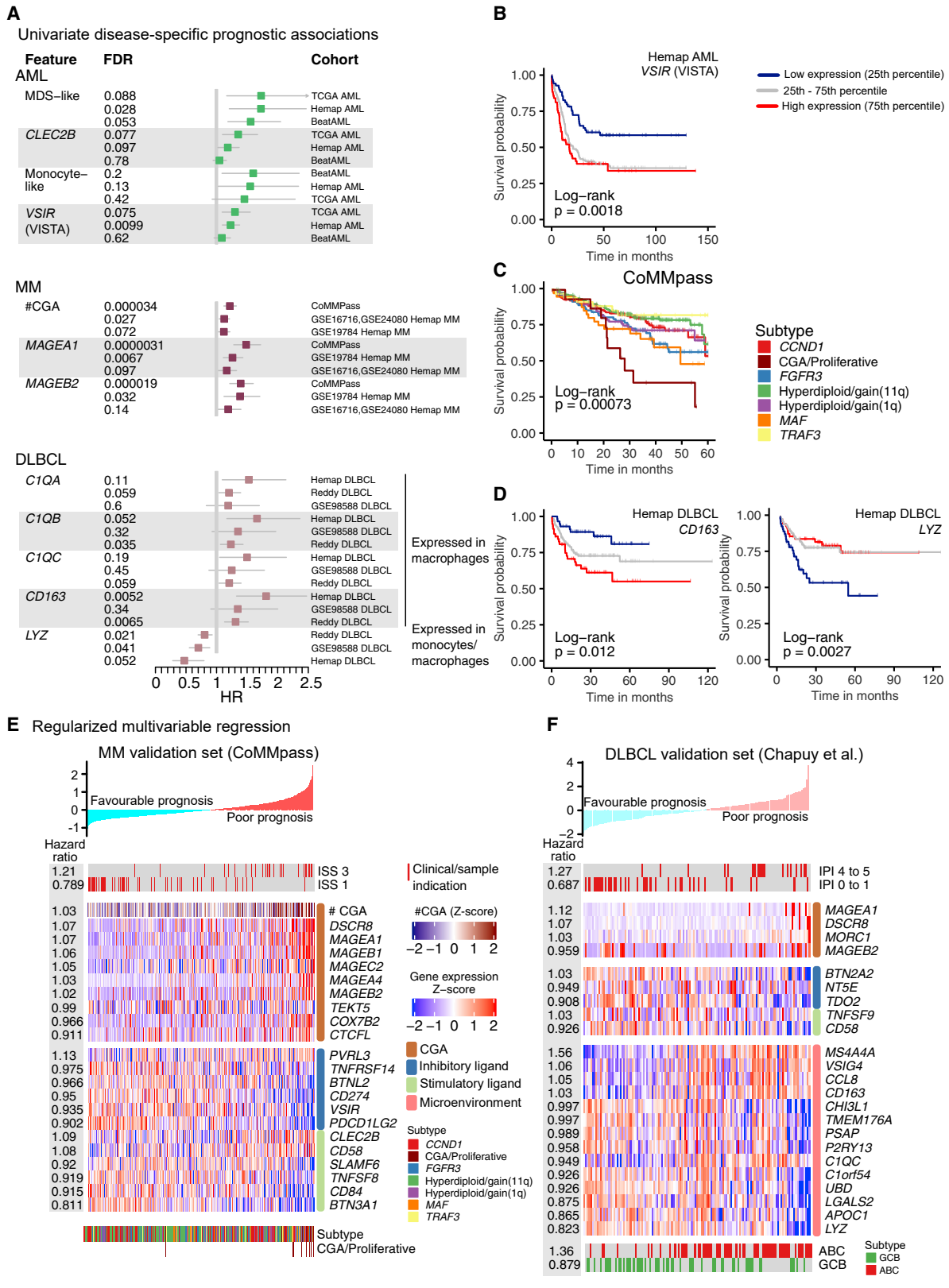


Figure 7. Immunological Properties Are Associated with Survival

(A) Forest plots of univariate Cox proportional hazards models. Selected features with FDR < 0.2 in at least two cohorts are shown.

(B) Kaplan-Meier curves of patients stratified by *VISTA* expression in Hemap AML.

(legend continued on next page)

individual CGAs, immunomodulatory genes, TME genes linked to cytolytic infiltrate and molecular cancer subtypes.

Several immunological prognostic associations were cancer type specific: in AML, the immunomodulatory genes *CLEC2B* and *VISTA* and the MDS-like and monocytic subtypes were linked to poor prognosis (Figures 7A and 7B). In MM, the CGA-high/proliferative subtype, the number of expressed and individual CGAs were associated with a particularly poor prognosis (Figures 7A and 7C). In DLBCL, several macrophage-expressed genes correlating with cytolytic score, including *C1QA*, *C1QB*, *C1QC*, and the M2 polarization-associated *CD163*, were linked to worse survival (Figures 7A and 7D). However, genes expressed in both monocytes and macrophages, such as *LYZ*, were associated with better overall survival, indicating that distinct types or states of infiltrating myeloid cells characterized by these genes have opposing impacts on outcomes. Our analysis also highlighted genes that showed significant associations across cancers, such as *CD274 (PD-L1)* linked to superior survival in MM and DLBCL (Figure S7B).

To profile the prognostic associations of immune properties in relation to each other and established clinical risk scores, we obtained survival models using multiple regression elastic net Cox proportional hazards modeling in Hemap MM, DLBCL, and AML. We validated the results in independent validation cohorts (Figures S7C–S7E). Immunological features significantly improved outcome predictions compared with clinical risk scores alone in all studied cancer types and further stratified patients within existing risk groups, including the cell-of-origin subtype and International Prognostic Index in DLBCL, International Staging System in MM, and European Leukemia Net in AML (Figure S7F).

In MM, several immune checkpoint ligands associated with superior survival, including *CD274 (PD-L1)*, *PDCD1LG2 (PD-L2)*, and *VISTA* (Figure 7E), whereas the TIGIT ligand *PVRL3 (CD113)*, highly expressed in MM compared with other cancers, was linked to poor survival. The expression of several CGAs was linked to worse outcomes in both MM and DLBCL, suggesting increased CGA expression in more aggressive cancers (Figures 7E and 7F). In DLBCL, also the multiple regression model highlighted the predictive value of the distinct monocyte- and macrophage-associated genes (Figure 7F). Importantly, genes such as *CD163* expressed in macrophages were strongly linked to the ABC subtype, suggesting that myeloid cell populations in the TME may contribute to the well-established inferior survival of this subtype. The AML multivariate model highlighted the poor prognosis associated with MDS-like signature genes and *VISTA* (Figure S7G). These findings demonstrate how immunological properties characterizing distinct molecular subtypes of hematological malignancies are linked to patient outcomes.

DISCUSSION

Understanding the determinants that shape the immunological landscape in cancer subtypes could enable more precise development of immune intervention approaches. Our exploration of large-scale genomic datasets uncovered TME properties and cancer cell-intrinsic genetic and epigenetic features explaining immunological heterogeneity in hematological malignancies.

In hematological malignancies, cancer cells transcriptionally resemble normal immune cells, hindering the use of standard deconvolution methods. Furthermore, tools aiming at immune cell fraction inference were found to mix signal from different $CD4^+$, $CD8^+$, and NK cell fractions (Sturm et al., 2019). The cytolytic score introduced here represents a simple but robust metric to capture the cytolytic infiltrate, corresponding to mature cytolytic NK and $CD8^+$ cells.

The higher cytolytic score in lymphomas compared with other hematological malignancies was associated with an IFN-γ signature and M2 macrophage-associated genes in the most heavily infiltrated CHL and DLBCL. A similar IFN-γ-related profile has been shown to predict clinical response to PD-1 blockade (Ayers et al., 2017). A distinct TME response might thus influence efficacy of immunotherapies in lymphomas as opposed to acute leukemias. In addition, cancer cell-intrinsic genetic alterations and molecular subtypes were linked to cytotoxic infiltration. In DLBCL, the ABC subtype showed higher infiltration compared with the GCB subtype and *CREBBP*-mutated cases. *CREBBP* mutations result in reduced HLA II-mediated antigen presentation in DLBCL (Hashwah et al., 2017; Jiang et al., 2017) and associated with reduced T cell infiltration in FL (Green et al., 2015), providing a potential mechanistic explanation to our observations. The high cytolytic infiltration in the MDS-like transcriptomic subtype of AML suggests that the cancer cell molecular phenotype may influence the immune infiltrate. The T and NK cells infiltrating MDS-like AML BM were enriched in effector and mature phenotypes, consistent with a previous report of T cells in AML (Knaus et al., 2018; Brück et al., 2020). The receptor-ligand interactions involving the inhibitory receptors LAG-3 and TIM-3 provide potential therapeutic targets in this subtype. Increased BM T cell infiltrate has been linked with favorable responses to PD-1 blockade combined with hypomethylating therapy in AML (Daver et al., 2019).

In line with our findings, *CIITA* methylation has been suggested to impair HLA II expression in hematopoietic cells (Morimoto et al., 2004). Here, we identified a distinct transcriptomic AML subgroup harboring *CIITA* methylation, characterized by a CMP/GMP phenotype and concurrent *NPM1* and *IDH1/2* and *TET2* mutations. As HLA loss has been linked to AML relapse after allo-HSCT (Vago et al., 2009), low expression level already at diagnosis could restrict $CD4^+$ T helper cell-mediated recognition both during an autologous immune response and in the allogeneic setting. Moreover, *CIITA* hypermethylation could be responsible for the transcriptional downregulation of HLA II upon relapse after allo-HSCT (Christopher et al., 2018; Toffalori et al., 2019). Reversal of promoter methylation by hypomethylating treatment could augment HLA II-dependent immunity. Of interest, combining PD-1 blockade with hypomethylating agents has demonstrated efficacy in AML (Daver et al., 2019).

(C) Kaplan-Meier curves of MM molecular subtypes.
(D) Kaplan-Meier curves of patients stratified by *CD163* and *LYZ* expression in Hemap DLBCL.
(E) Heatmap of features included in the MM immunological risk model. Patients are ordered by the risk index.
(F) Immunological risk model heatmap as in (E) for DLBCL.
See also Figure S7 and Table S7.

Several co-inhibitory ligands were expressed in a cancer type- and lineage-specific fashion, further impacted by genetic and epigenetic factors, suggesting that targeting different inhibitory interactions might be required for maximizing immunotherapy benefit in each disease. *VISTA* emerged as a checkpoint enriched in myeloid malignancies. In AML, VISTA has been implicated in immune suppression through its expression in myeloid-derived suppressor cells (Wang et al., 2018), whereas our analysis indicates expression particularly in leukemic monocytes. Highest *VISTA* expression was found in JMML, again suggesting association with monocytic differentiation. In healthy hematopoiesis, *VISTA* is expressed in monocytes and neutrophils (Flies et al., 2014), and could be utilized by cancer cells of these lineages for immune evasion. VISTA has been implicated as a potential immunotherapy target in several solid tumors (Blando et al., 2019; Gao et al., 2017; Hmeljak et al., 2018). In contrast, CD70 showed a B-lymphoid expression pattern, with inactivating *CD70* mutations promoting escape from T cell-mediated immunity. Co-stimulation through the CD70 receptor CD27 may therefore be critical for anti-lymphoma immunity, possibly by activating naive T cells highly expressing CD27 (Buchan et al., 2018). CD27 agonists may have potential to overcome the immune evasion conferred by *CD70* mutations.

CGA expression and anti-CGA immune responses have been demonstrated in MM (Atanackovic et al., 2007; van Duin et al., 2011). Our comparison highlighted this property in comparison with other cancers. In both MM and DLBCL, CGA expression correlated with promoter hypomethylation, high mutation, and CNV load/hyperdiploidy, as well as HLA II downregulation and *CD58/HLA* mutations, suggesting immune evasion. MM and DLBCL are more highly mutated than other hematological malignancies (Alexandrov et al., 2013), suggesting higher genomic complexity and immunogenicity in CGA-high cancers. MM treatment could potentially benefit from immunotherapies leveraging the high number of expressed antigens, such as vaccines or T cell receptor therapies tailored for most common CGAs. In contrast, hypomethylating agents could be used to increase antigenicity in leukemias.

The immunological risk models highlighted the unfavorable VISTA-associated MDS-like and monocytic subtypes in AML, further underlining VISTA as a potential target for immune checkpoint blockade. Macrophage infiltration in ABC DLBCL predicted poor survival, suggesting a distinct immune TME together with the increased cytolytic infiltrate and PD-L1 expression found in this subtype. The CGA-high cluster of MM displayed strikingly inferior survival, indicating a need for immunotherapies leveraging the high antigen load.

Our approach of estimating immune cell composition from bulk gene expression data is limited in the analysis of rare cell types. We anticipate that single-cell studies characterizing simultaneously transcriptional, genomic, and epigenetic states (Cao et al., 2018; Petti et al., 2019; Zachariadis et al., 2020) will further illuminate the immune cell-cancer crosstalk. We envision that the genotypic-immunophenotypic associations presented here can guide further investigation into the underlying characteristics that modulate inter-tumor heterogeneity in immune landscapes.

In summary, our integrative analysis provides evidence of genomic and microenvironmental factors associated with variation in the immune contexture between different tumors. The findings of this study highlight the need to integrate genetic, epigenetic, and transcriptomic data of different aspects of the immune landscape to understand potential determinants of responsiveness to cancer immunotherapies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patients
  - Cell Lines
- METHOD DETAILS
  - Processing of Genome-Wide Multilevel Data
  - Sample Stratification Based on Gene Expression Profiles
  - Statistical Analysis Using Discrete Gene Expression Features
  - Development of Immunological Scores
  - Validation of Immunological Scores
  - Gene Set Enrichment Analysis
  - Analysis of Microenvironment Genes Correlated to Cytolytic Score
  - Multiplexed Immunohistochemistry (mIHC)
  - Genomic Correlations with Immunological Features
  - Single-cell RNA Sequencing
  - Differential Methylation Analysis
  - Decitabine and IFNγ Treatment Experiments
  - CRISPR-Cas9-Mediated Gene Silencing
  - Immune Checkpoint Gene List Curation
  - TCR Signaling Reporter Experiments
  - Antigen Expression Analysis
  - Survival Analysis
  - Data Visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.ccell.2020.06.002.

ACKNOWLEDGMENTS

We thank the Helsinki Biobank and the Digital and Molecular Pathology Unit (University of Helsinki [UH] and Biocenter Finland [BF]). The computational infrastructure was provided by the UEF Bioinformatics Center and CSC IT Center for Science, Finland. We acknowledge the FIMM Sequencing Laboratory, Technology Center, and the High-Throughput Biomedicine unit (HiLIFE, UH). The Genome Biology Unit (HiLIFE) generated constructs. We acknowledge Prof. Peter Steinberger and Dr. Judith Leitner for providing TCR reporter cell lines. The results are in part based upon data generated by the Multiple Myeloma Research Foundation Personalized Medicine Initiative and The Cancer Genome Atlas project established by the NCI and NHGRI. This study was supported by the Cancer Foundation Finland, Sigrid Jusélius Foundation,

Relander Foundation, State funding for university-level health research in Finland, Academy of Finland (321553), Biomedicum Helsinki Foundation, Hi-LIFE (UH) fellow funds, University of Eastern Finland, and Orion Research Foundation.

AUTHOR CONTRIBUTIONS

O.D., P.P., M.A.I.K., M.H., and S.M. conceived the study and wrote the article with contributions from all authors. O.D. and P.P. designed and performed the analyses. O.D. performed experimental validations and visualized data. P.P. performed computational analyses (multi-omics, RRBS, scRNA-seq, elastic nets). O.B. performed mIHC. J.K. generated CRISPR/Cas9-edited cell lines. J.M. developed workflows for scRNA-seq analysis. A.K., J.H., and M.K. performed AML RNA-seq data analysis and C.H. supervised RNA-seq collection. D.M., M.V.-K., and K.W. provided samples and reagents. S.S. generated flow cytometry data. B.G. processed RRBS data. J.L. and P.M. performed scRNA-seq. L.M., S.-K.L., and S.L. constructed the DLBCL tissue microarray. M.N. and O.L. supervised Hemap data harmonization. S.M. and M.H. supervised the project.

DECLARATION OF INTERESTS

S.M. has received honoraria and research funding from Novartis, Pfizer, and Bristol-Myers Squibb, and C.H. received funding from Celgene, Novartis, Oncopeptides, Orion Pharma, and the Innovative Medicines Initiative project HARMONY. The remaining authors declare no competing interests.

Received: April 1, 2019
Revised: March 27, 2020
Accepted: May 29, 2020
Published: July 9, 2020; corrected online: September 1, 2020

REFERENCES

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

van Allen, E.M.V., Miao, D., Schilling, B., Shukla, S.A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M.H.G., Goldinger, S.M., et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science *350*, 207–211.

Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C.C., Mundstein, A.S., Cohen, T., Chen, Y.-T., Chua, R., Gurung, S., Gnjatic, S., et al. (2009). CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res *37*, D816–D819.

Ansell, S.M., Lesokhin, A.M., Borrello, I., Halwani, A., Scott, E.C., Gutierrez, M., Schuster, S.J., Millenson, M.M., Cattry, D., Freeman, G.J., et al. (2015). PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. N. Engl. J. Med. *372*, 311–319.

Ansell, S.M., Minnema, M.C., Johnson, P., Timmerman, J.M., Armand, P., Shipp, M.A., Rodig, S.J., Ligon, A.H., Roemer, M.G.M., Reddy, N., et al. (2019). Nivolumab for relapsed/refractory diffuse large B-cell lymphoma in patients ineligible for or having failed autologous transplantation: a single-arm, phase II study. J. Clin. Oncol. *37*, 481–489.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172.

Atanackovic, D., Arfsten, J., Cao, Y., Gnjatic, S., Schnieders, F., Bartels, K., Schilling, G., Faltz, C., Wolschke, C., Dierlamm, J., et al. (2007). Cancer-testis antigens are commonly expressed in multiple myeloma and induce systemic immunity following allogeneic stem cell transplantation. Blood *109*, 1103–1112.

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al. (2017). IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. J. Clin. Invest. *127*, 2930–2940.

Bachireddy, P., Burkhardt, U.E., Rajasagi, M., and Wu, C.J. (2015). Haematological malignancies: at the forefront of immunotherapeutic innovation. Nat. Rev. Cancer *15*, 201–215.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Beavis, Paul.A., Stagg, J., Darcy, P.K., and Smyth, M.J. (2012). CD73: a potent suppressor of antitumor immune responses. Trends Immunol. *33*, 231–237.

Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. *17*, 218.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity *39*, 782–795.

Blando, J., Sharma, A., Higa, M.G., Zhao, H., Vence, L., Yadav, S.S., Kim, J., Sepulveda, A.M., Sharp, M., Maitra, A., et al. (2019). Comparison of immune infiltrates in melanoma and pancreatic cancer highlights VISTA as a potential target in pancreatic cancer. Proc. Natl. Acad. Sci. U S A *116*, 1692–1697.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Brück, O., Blom, S., Dufva, O., Turkki, R., Chheda, H., Ribeiro, A., Kovanen, P., Aittokallio, T., Koskenvesa, P., Kallioniemi, O., et al. (2018). Immune cell contexture in the bone marrow tumor microenvironment impacts therapy response in CML. Leukemia *32*, 1643–1656.

Brück, O., Dufva, O., Hohtari, H., Blom, S., Turkki, R., Ilander, M., Kovanen, P., Pallaud, C., Ramos, P.M., Lähteenmäki, H., et al. (2020). Immune profiles in acute myeloid leukemia bone marrow associate with patient age, T-cell receptor clonality, and survival. Blood Adv. *4*, 274–286.

Buchan, S.L., Rogel, A., and Al-Shamkhani, A. (2018). The immunobiology of CD27 and OX40 and their potential as targets for cancer immunotherapy. Blood *131*, 39–48.

Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science *361*, 1380–1385.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.

Casucci, M., Perna, S.K., Falcone, L., Camisa, B., Magnani, Z., Bernardi, M., Crotta, A., Tresoldi, C., Fleischhauer, K., Ponzoni, M., et al. (2013). Graft-versus-leukemia effect of HLA-haploidentical central-memory T-cells expanded with leukemic APCs and modified with a suicide gene. Mol. Ther. *21*, 466–475.

Chapuy, B., Stewart, C., Dunford, A.J., Kim, J., Kamburov, A., Redd, R.A., Lawrence, M.S., Roemer, M.G.M., Li, A.J., Ziepert, M., et al. (2018). Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat. Med. *24*, 679–690.

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., and Trajanoski, Z. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. *18*, 248–262.

Chen, L., and Flies, D.B. (2013). Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat. Rev. Immunol. *13*, 227–242.

Chen, D.S., and Mellman, I. (2013). Oncology meets immunology: the cancer-immunity cycle. Immunity *39*, 1–10.

Chen, D.S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. Nature *541*, 321–330.

Chen, H.-M., van der Touw, W., Wang, Y.S., Kang, K., Mai, S., Zhang, J., Alsina-Beauchamp, D., Duty, J.A., Mungamuri, S.K., Zhang, B., et al. (2018). Blocking immunoinhibitory receptor LILRB2 reprograms tumor-associated myeloid cells and promotes antitumor immunity. J. Clin. Invest. *128*, 5647–5662.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. IEEE Trans. Pattern Anal. Mach. Intell. *17*, 790–799.

Chng, W.J., Remstein, E.D., Fonseca, R., Bergsagel, P.L., Vrana, J.A., Kurtin, P.J., and Dogan, A. (2009). Gene expression profiling of pulmonary mucosa-associated lymphoid tissue lymphoma identifies new biologic insights with potential diagnostic and therapeutic applications. Blood *113*, 635–645.

Christopher, M.J., Petti, A.A., Rettig, M.P., Miller, C.A., Chendamarai, E., Duncavage, E.J., Klco, J.M., Helton, N.M., O'Laughlin, M., Fronick, C.C., et al. (2018). Immune escape of relapsed AML cells after allogeneic transplantation. N. Engl. J. Med. *379*, 2330–2341.

Daver, N., Garcia-Manero, G., Basu, S., Boddu, P.C., Alfayez, M., Cortes, J.E., Konopleva, M., Ravandi-Kashani, F., Jabbour, E., Kadia, T., et al. (2019). Efficacy, safety, and biomarkers of response to azacitidine and nivolumab in relapsed/refractory acute myeloid leukemia: a nonrandomized, open-label, phase II study. Cancer Discov. *9*, 370–383.

Deuse, T., Hu, X., Gravina, A., Wang, D., Tediashvili, G., De, C., Thayer, W.O., Wahl, A., Garcia, J.V., Reichenspurner, H., et al. (2019). Hypoimmunogenic derivatives of induced pluripotent stem cells evade immune rejection in fully immunocompetent allogeneic recipients. Nat. Biotechnol. *37*, 252–258.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

van Duin, M., Broyl, A., de Knegt, Y., Goldschmidt, H., Richardson, P.G., Hop, W.C.J., van der Holt, B., Joseph-Pietras, D., Mulligan, G., Neuwirth, R., et al. (2011). Cancer testis antigens in newly diagnosed and relapse multiple myeloma: prognostic markers and potential targets for immunotherapy. Haematologica *96*, 1662–1669.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat. Protoc. *15*, 1–23.

Eklund, A.C., and Szallasi, Z. (2008). Correction of technical bias in clinical microarray data improves concordance with known biological information. Genome Biol. *9*, R26.

Flies, D.B., Han, X., Higuchi, T., Zheng, L., Sun, J., Ye, J.J., and Chen, L. (2014). Coinhibitory receptor PD-1H preferentially suppresses $CD4^+$ T cell-mediated immunity. J. Clin. Invest. *124*, 1966–1975.

Fridman, W.H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. Nat. Rev. Cancer *12*, 298–306.

van Galen, P., Hovestadt, V., Wadsworth, M.H., II, Hughes, T.K., Griffin, G.K., Battaglia, S., Verga, J.A., Stephansky, J., Pastika, T.J., Lombardi Story, J., et al. (2019). Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. Cell *176*, 1265–1281.e24.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J Stat Softw *33*, 1–22.

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science *313*, 1960–1964.

Gao, J., Ward, J.F., Pettaway, C.A., Shi, L.Z., Subudhi, S.K., Vence, L.M., Zhao, H., Chen, J., Chen, H., Efstathiou, E., et al. (2017). VISTA is an inhibitory immune checkpoint that is increased after ipilimumab therapy in patients with prostate cancer. Nat. Med. *23*, 551–555.

Garrido, F., Cabrera, T., and Aptsiauri, N. (2010). "Hard" and "soft" lesions underlying the HLA class I alterations in cancer cells: implications for immunotherapy. Int. J. Cancer *127*, 249–256.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307–315.

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. *21*, 938–945.

Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature *569*, 503–508.

Glass, J.L., Hassane, D., Wouters, B.J., Kunimoto, H., Avellino, R., Garrett-Bakelman, F.E., Guryanova, O.A., Bowman, R., Redlich, S., Intlekofer, A.M., et al. (2017). Epigenetic identity in AML depends on disruption of nonpromoter regulatory elements and is affected by antagonistic effects of mutations in epigenetic modifiers. Cancer Discov. *7*, 868–883.

Gordon, S. (2003). Alternative activation of macrophages. Nat. Rev. Immunol. *3*, 23–35.

Green, M.R., Kihira, S., Liu, C.L., Nair, R.V., Salari, R., Gentles, A.J., Irish, J., Stehr, H., Vicente-Dueñas, C., Romero-Camarero, I., et al. (2015). Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation. Proc. Natl. Acad. Sci. U S A *112*, E1116–E1125.

Groom, J.R., and Luster, A.D. (2011). CXCR3 ligands: redundant, collaborative and antagonistic functions. Immunol. Cell Biol. *89*, 207–215.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science *348*, 648–660.

Gu, Z., Churchman, M.L., Roberts, K.G., Moore, I., Zhou, X., Nakitandwe, J., Hagiwara, K., Pelletier, S., Gingras, S., Berns, H., et al. (2019). PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. Nat. Genet. *51*, 296–307.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics *14*, 7.

Hashwah, H., Schmid, C.A., Kasser, S., Bertram, K., Stelling, A., Manz, M.G., and Müller, A. (2017). Inactivation of CREBBP expands the germinal center B cell compartment, down-regulates MHCII expression and promotes DLBCL growth. Proc. Natl. Acad. Sci. U S A *114*, 9701–9706.

Hay, S.B., Ferchen, K., Chetal, K., Grimes, H.L., and Salomonis, N. (2018). The Human Cell Atlas bone marrow single-cell interactive web portal. Exp. Hematol. *68*, 51–61.

Hmeljak, J., Sanchez-Vega, F., Hoadley, K.A., Shih, J., Stewart, C., Heiman, D., Tarpey, P., Danilova, L., Drill, E., Gibb, E.A., et al. (2018). Integrative molecular characterization of malignant pleural mesothelioma. Cancer Discov. *8*, 1548–1565.

Hohtari, H., Brück, O., Blom, S., Turkki, R., Sinisalo, M., Kovanen, P.E., Kallioniemi, O., Pellinen, T., Porkka, K., and Mustjoki, S. (2019). Immune cell constitution in bone marrow microenvironment predicts outcome in adult ALL. Leukemia *33*, 1570–1582.

Holling, T.M., Schooten, E., Langerak, A.W., and van den Elsen, P.J. (2004). Regulation of MHC class II expression in human T-cell malignancies. Blood *103*, 1438–1444.

Hsiau, T., Conant, D., Rossi, N., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., Enzmann, B.L., et al. (2019). Inference of CRISPR edits from Sanger trace data. BioRxiv. https://doi.org/10.1101/251082.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. *31*, e15.

Izawa, K., Martin, E., Soudais, C., Bruneau, J., Boutboul, D., Rodriguez, R., Lenoir, C., Hislop, A.D., Besson, C., Touzot, F., et al. (2017). Inherited CD70 deficiency in humans reveals a critical role for the CD70–CD27 pathway in immunity to Epstein-Barr virus infection. J. Exp. Med. *214*, 73–89.

Jiang, Y., Ortega-Molina, A., Geng, H., Ying, H.-Y., Hatzi, K., Parsa, S., McNally, D., Wang, L., Doane, A.S., Agirre, X., et al. (2017). CREBBP inactivation promotes the development of HDAC3-dependent lymphomas. Cancer Discov. *7*, 38–53.

Joyce, J.A., and Fearon, D.T. (2015). T cell exclusion, immune privilege, and the tumor microenvironment. Science *348*, 74–80.

Jutz, S., Leitner, J., Schmetterer, K., Doel-Perez, I., Majdic, O., Grabmeier-Pfistershammer, K., Paster, W., Huppa, J.B., and Steinberger, P. (2016). Assessment of costimulation and coinhibition in a triple parameter T cell reporter line: simultaneous measurement of NF-κB, NFAT and AP-1. J. Immunol. Methods *430*, 10–20.

Kline, J., Godfrey, J., and Ansell, S.M. (2020). The immune landscape and response to immune checkpoint blockade therapy in lymphoma. Blood *135*, 523–533.

Knaus, H.A., Berglund, S., Hackl, H., Blackford, A.L., Zeidner, J.F., Montiel-Esparza, R., Mukhopadhyay, R., Vanura, K., Blazar, B.R., Karp, J.E., et al. (2018). Signatures of $CD8^+$ T cell dysfunction in AML patients and their reversibility with response to chemotherapy. JCI Insight *3*, https://doi.org/10.1172/jci.insight.120974.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Kumar, A., Kankainen, M., Parsons, A., Kallioniemi, O., Mattila, P., and Heckman, C.A. (2017). The impact of RNA sequence library construction protocols on transcriptomic profiling of leukemia. BMC Genomics *18*, 629.

Leitner, J., Kuschei, W., Grabmeier-Pfistershammer, K., Woitek, R., Kriehuber, E., Majdic, O., Zlabinger, G., Pickl, W.F., and Steinberger, P. (2010). T cell stimulator cells, an efficient and versatile cellular system to assess the role of costimulatory ligands in the activation of human T cells. J. Immunol. Methods *362*, 131–141.

Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., Jiang, P., Shen, H., Aster, J.C., Rodig, S., et al. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. *17*, 174.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.

van der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. J. Mach. Learn. Res. *15*, 3221–3245.

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Manojlovic, Z., Christofferson, A., Liang, W.S., Aldrich, J., Washington, M., Wong, S., Rohrer, D., Jewell, S., Kittles, R.A., Derome, M., et al. (2017). Comprehensive molecular profiling of 718 multiple myelomas reveals significant differences in mutation frequencies between African and European descent cases. PLoS Genet. *13*, e1007087.

Mateos, M.-V., Blacklock, H., Schjesvold, F., Oriol, A., Simpson, D., George, A., Goldschmidt, H., Larocca, A., Chanan-Khan, A., Sherbenou, D., et al. (2019). Pembrolizumab plus pomalidomide and dexamethasone for patients with relapsed or refractory multiple myeloma (KEYNOTE-183): a randomised, open-label, phase 3 trial. Lancet Haematol. *6*, e459–e469.

Maude, S.L., Frey, N., Shaw, P.A., Aplenc, R., Barrett, D.M., Bunin, N.J., Chew, A., Gonzalez, V.E., Zheng, Z., Lacey, S.F., et al. (2014). Chimeric antigen receptor T cells for sustained remissions in leukemia. N. Engl. J. Med. *371*, 1507–1517.

Mehtonen, J., Pölönen, P., Häyrynen, S., Dufva, O., Lin, J., Liuksiala, T., Granberg, K., Lohi, O., Hautamäki, V., Nykter, M., et al. (2019). Data-driven characterization of molecular phenotypes across heterogeneous sample collections. Nucleic Acids Res. *47*, e76.

Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al. (2015). The human transcriptome across tissues and individuals. Science *348*, 660–665.

Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess, G.T., Han, K., Jeng, E.E., et al. (2017). Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. Nat. Commun. *8*, 1–8.

Morimoto, Y., Toyota, M., Satoh, A., Murai, M., Mita, H., Suzuki, H., Takamura, Y., Ikeda, H., Ishida, T., Sato, N., et al. (2004). Inactivation of class II transactivator by DNA methylation and histone deacetylation associated with absence of HLA-DR induction by interferon-γ in haematopoietic tumour cells. Br. J. Cancer *90*, 844–852.

Morvan, M.G., and Lanier, L.L. (2016). NK cells and cancer: you can teach innate cells new tricks. Nat. Rev. Cancer *16*, 7–19.

Muhlethaler-Mottet, A., Otten, L.A., Steimle, V., and Mach, B. (1997). Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. EMBO J. *16*, 2851–2860.

Munn, D.H., and Mellor, A.L. (2016). Ido in the tumor microenvironment: inflammation, counter-regulation, and tolerance. Trends Immunol. *37*, 193–207.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Nordlund, J., Bäcklin, C.L., Wahlberg, P., Busche, S., Berglund, E.C., Eloranta, M.-L., Flaegstad, T., Forestier, E., Frost, B.-M., Harila-Saari, A., et al. (2013). Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome Biol *14*, r105.

Park, Y., Figueroa, M.E., Rozek, L.S., and Sartor, M.A. (2014). MethylSig: a whole genome DNA methylation analysis pipeline. Bioinformatics *30*, 2414–2422.

Petti, A.A., Williams, S.R., Miller, C.A., Fiddes, I.T., Srivatsan, S.N., Chen, D.Y., Fronick, C.C., Fulton, R.S., Church, D.M., and Ley, T.J. (2019). A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. Nat. Commun. *10*, 1–16.

Pollari, M., Brück, O., Pellinen, T., Vähämurto, P., Karjalainen-Lindsberg, M.-L., Mannisto, S., Kallioniemi, O., Kellokumpu-Lehtinen, P.-L., Mustjoki, S., Leivonen, S.-K., et al. (2018). PD-L1+ tumor-associated macrophages and PD-1+ tumor-infiltrating lymphocytes predict survival in primary testicular lymphoma. Haematologica *103*, 1908–1914.

Pölönen, P., Mehtonen, J., Lin, J., Liuksiala, T., Häyrynen, S., Teppo, S., Mäkinen, A., Kumar, A., Malani, D., Pohjolainen, V., et al. (2019). Hemap: an interactive online resource for characterizing molecular phenotypes across hematologic malignancies. Cancer Res. *79*, 2466–2479.

Ratzinger, F., Haslacher, H., Poeppl, W., Hoermann, G., Kovarik, J.J., Jutz, S., Steinberger, P., Burgmann, H., Pickl, W.F., and Schmetterer, K.G. (2014). Azithromycin suppresses CD4+ T-cell activation by direct modulation of mTOR activity. Sci. Rep. *4*, 1–10.

Reddy, A., Zhang, J., Davis, N.S., Moffitt, A.B., Love, C.L., Waldrop, A., Leppa, S., Pasanen, A., Meriranta, L., Karjalainen-Lindsberg, M.-L., et al. (2017). Genetic and functional drivers of diffuse large B cell lymphoma. Cell *171*, 481–494.e15.

Rendeiro, A.F., Krausgruber, T., Fortelny, N., Zhao, F., Penz, T., Farlik, M., Schuster, L.C., Nemc, A., Tasnády, S., Réti, M., et al. (2020). Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. Nat. Commun. *11*, 1–14.

Ribas, A., and Wolchok, J.D. (2018). Cancer immunotherapy using checkpoint blockade. Science *359*, 1350–1355.

Rimsza, L.M., Roberts, R.A., Miller, T.P., Unger, J.M., LeBlanc, M., Braziel, R.M., Weisenberger, D.D., Chan, W.C., Muller-Hermelink, H.K., Jaffe, E.S., et al. (2004). Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: a follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. Blood *103*, 4251–4258.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roemer, M.G.M., Redd, R.A., Cader, F.Z., Pak, C.J., Abdelrahman, S., Ouyang, J., Sasse, S., Younes, A., Fanale, M., Santoro, A., et al. (2018). Major histocompatibility complex class II and programmed death ligand 1 expression predict outcome after programmed death 1 blockade in classic Hodgkin lymphoma. J. Clin. Oncol. *36*, 942–950.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.

Royston, P., and Altman, D.G. (2013). External validation of a Cox prognostic model: principles and methods. BMC Med. Res. Methodol. *13*, 33.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schuster, S.J., Svoboda, J., Chong, E.A., Nasta, S.D., Mato, A.R., Anak, Ö., Brogdon, J.L., Pruteanu-Malinici, I., Bhoj, V., Landsburg, D., et al. (2017). Chimeric antigen receptor T cells in refractory B-cell lymphomas. N. Engl. J. Med. *377*, 2545–2554.

Scott, D.W., and Gascoyne, R.D. (2014). The tumour microenvironment in B cell lymphomas. Nat. Rev. Cancer *14*, 517–534.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J *8*, 289–317.

Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. *33*, 1152–1158.

Silacci, P., Mottet, A., Steimle, V., Reith, W., and Mach, B. (1994). Developmental extinction of major histocompatibility complex class II gene expression in plasmocytes is mediated by silencing of the transactivator gene CIITA. J. Exp. Med. *180*, 1329–1336.

Silva, F.P.G., Swagemakers, S.M.A., Erpelinck-Verschueren, C., Wouters, B.J., Delwel, R., Vrieling, H., van der Spek, P., Valk, P.J.M., and Giphart-Gassler, M. (2009). Gene expression profiling of minimally differentiated acute myeloid leukemia: M0 is a distinct entity subdivided by RUNX1 mutation status. Blood *114*, 3001–3007.

Simpson, A.J.G., Caballero, O.L., Jungbluth, A., Chen, Y.-T., and Old, L.J. (2005). Cancer/testis antigens, gametogenesis and cancer. Nat. Rev. Cancer *5*, 615–625.

Spranger, S., Spaapen, R.M., Zha, Y., Williams, J., Meng, Y., Ha, T.T., and Gajewski, T.F. (2013). Up-regulation of PD-L1, Ido, and tregs in the melanoma tumor microenvironment is driven by CD8+ T cells. Sci. Transl. Med. *5*, 200ra116.

Steimle, V., Siegrist, C.A., Mottet, A., Lisowska-Grospierre, B., and Mach, B. (1994). Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. Science *265*, 106–109.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S.Y., Hahn, W.C., Sharp, P.A., et al. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA *9*, 493–501.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

Sturm, G., Finotello, F., Petitprez, F., Zhang, J.D., Baumbach, J., Fridman, W.H., List, M., and Aneichyk, T. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics *35*, i436–i445.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. *102*, 15545–15550.

Szabo, P.A., Levitin, H.M., Miron, M., Snyder, M.E., Senda, T., Yuan, J., Cheng, Y.L., Bush, E.C., Dogra, P., Thapa, P., et al. (2019). Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. Nat. Commun. *10*, 1–16.

The Cancer Genome Atlas Research Network (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. *368*, 2059–2074.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.-H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. Immunity *48*, 812–830.e14.

Toffalori, C., Zito, L., Gambacorta, V., Riba, M., Oliveira, G., Bucci, G., Barcella, M., Spinelli, O., Greco, R., Crucitti, L., et al. (2019). Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation. Nat. Med. *25*, 603–611.

Tumeh, P.C., Harview, C.L., Yearley, J.H., Shintaku, I.P., Taylor, E.J.M., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature *515*, 568–571.

Tyner, J.W., Tognon, C.E., Bottomly, D., Wilmot, B., Kurtz, S.E., Savage, S.L., Long, N., Schultz, A.R., Traer, E., Abel, M., et al. (2018). Functional genomic landscape of acute myeloid leukaemia. Nature *562*, 526–531.

Vago, L., Perna, S.K., Zanussi, M., Mazzi, B., Barlassina, C., Stanghellini, M.T.L., Perrelli, N.F., Cosentino, C., Torri, F., Angius, A., et al. (2009). Loss of mismatched HLA in leukemia after stem-cell transplantation. N. Engl. J. Med. *361*, 478–488.

Vasconcelos, Y., De Vos, J., Vallat, L., Rème, T., Lalanne, A.I., Wanherdrick, K., Michel, A., Nguyen-Khac, F., Oppezzo, P., Magnac, C., et al. (2005). Gene expression profiling of chronic lymphocytic leukemia can discriminate cases with stable disease and mutated Ig genes from those with progressive disease and unmutated Ig genes. Leukemia *19*, 2002–2005.

Wang, L., Jia, B., Claxton, D.F., Ehmann, W.C., Rybka, W.B., Mineishi, S., Naik, S., Khawaja, M.R., Sivik, J., Han, J., et al. (2018). VISTA is highly expressed on MDSCs and mediates an inhibition of T cell response in patients with AML. OncoImmunology *7*, e1469594.

Wellenstein, M.D., and de Visser, K.E. (2018). Cancer-cell-intrinsic mechanisms shaping the tumor immune landscape. Immunity *48*, 399–416.

Wetzler, M., McElwain, B.K., Stewart, C.C., Blumenson, L., Mortazavi, A., Ford, L.A., Slack, J.L., Barcos, M., Ferrone, S., and Baer, M.R. (2003). HLA-DR antigen-negative acute myeloid leukemia. Leukemia *17*, 707–715.

Yang, C., Siebert, J.R., Burns, R., Gerbec, Z.J., Bonacci, B., Rymaszewski, A., Rau, M., Riese, M.J., Rao, S., Carlson, K.-S., et al. (2019). Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome. Nat. Commun. *10*, 1–16.

Younes, A., Brody, J., Carpio, C., Lopez-Guillermo, A., Ben-Yehuda, D., Ferhanoglu, B., Nagler, A., Ozcan, M., Avivi, I., Bosch, F., et al. (2019). Safety and activity of ibrutinib in combination with nivolumab in patients with relapsed non-Hodgkin lymphoma or chronic lymphocytic leukaemia: a phase 1/2a study. Lancet Haematol. *6*, e67–e78.

Zachariadis, V., Cheng, H., Andrews, N.J., and Enge, M. (2020). A highly scalable method for joint whole genome sequencing and gene expression profiling of single cells. BioRxiv. https://doi.org/10.1101/2020.03.04.976530.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD2 FITC (clone S5.2)	BD	Cat#: 347404
CD3 PerCP-Cy5.5 (clone SK7)	BD	Cat#: 332771
CD10 PE (clone HI10a)	BD	Cat#: 332776
CD19 PE-Cy7 (clone SJ25C1)	BD	Cat#: 341113
CD22 APC (clone S-HCL-1)	BD	Cat#: 333145
CD45 APC-Cy7 (clone 2D1)	BD	Cat#: 348814
CD56 FITC (clone NCAM16.2)	BD	Cat#: 345811
CD13 PE (clone L138)	BD	Cat#: 347406; RRID: AB_2732011
CD34 PerCP-Cy5.5 (clone 8G12)	BD	Cat#: 347222
CD117 PC7 (clone 104D2D1)	Beckman Coulter	Cat#: IM3698; RRID: AB_131184
CD33 APC (clone P67.6)	BD	Cat#: 345800
CD11b APC-Alexa Fluor 750 (clone Bear 1)	Beckman Coulter	Cat#: B36295
HLA-DR V450 (clone L243)	BD	Cat#: 655874; RRID: AB_2716783
CD45 KromeOrange (clone J.33)	Beckman Coulter	Cat#: A96416
CD14 FITC (clone MP9)	BD	Cat#: 345784
CD64 PE (clone 10.1)	Bio-Rad (Formerly AbD Serotec)	Cat#: MCA756PE; RRID: AB_321800
CD33 PE-Cy7 (clone P67.6)	BD	Cat#: 333952; RRID: AB_2713932
HLA-DR APC (clone L243)	BD	Cat#: 347403
HLA-DR FITC (clone G46-6)	BD	Cat#: 555811; RRID: AB_396145
IgG2a κ isotype control FITC (clone G155-178)	BD	Cat#: 555573; RRID: AB_395952
CD45 PerCP (clone 2D1)	BD	Cat#: 345809
CD45.2 APC (clone 104)	BioLegend	Cat#: 109814; RRID: AB_389211
CD19 APC-Cy7 (clone SJ25C1)	BD	Cat#: 557791; RRID: AB_396873
CD8 (clone C8/144B)	BioSB	Cat#: BSB 5172
CD11b (clone EP45)	BioSB	Cat#: BSB 6439
CD34 (clone QBEnd 10)	Dako	Cat#: M716501-2; RRID: AB_2750581
CD68 (clone KP1)	Abcam	Cat#: ab955; RRID: AB_307338
IDO (clone D5J4E)	Cell Signaling Technology	Cat#: 86630; RRID: AB_2636818
VISTA (clone D1L2G)	Cell Signaling Technology	Cat#: 64953
CXCL9 (polyclonal)	Thermo Fisher Scientific	Cat#: PA5-34743; RRID: AB_2552095
CD14 (clone D7A2T)	Cell Signaling Technology	Cat#: 75181
DAPI (4′,6-Diamidine-2′-phenylindole dihydrochloride)	Sigma-Aldrich	Cat#: 10236276001
Alexa Fluor 647 Goat Anti-Mouse IgG (H+L)	Thermo Fisher Scientific	Cat#: A-21236; RRID: AB_2535805
Alexa Fluor 647 Goat Anti-Rabbit IgG (H+L)	Thermo Fisher Scientific	Cat#: A-21245; RRID: AB_2535813
Alexa Fluor 750 Goat Anti-Mouse IgG (H+L)	Thermo Fisher Scientific	Cat#: A-21037; RRID: AB_2535708
Alexa Fluor 750 Goat Anti-Rabbit IgG (H+L)	Thermo Fisher Scientific	Cat#: A-21039; RRID: AB_2535710
BrightVision Poly-HRP Goat Anti-Mouse	Immunologic	VWR Cat#: VWRKDPVM55HRP
BrightVision Poly-HRP Goat anti-Rabbit	Immunologic	VWR Cat#: VWRKDPVR55HRP
Biological Samples		
AML patient bone marrow aspirates	This study; Helsinki University Hospital (Helsinki, Finland)	N/A
AML patient bone marrow biopsy tissue microarray	(Brück et al., 2020)	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CML patient bone marrow biopsy tissue microarray	(Brück et al., 2018)	N/A
ALL patient bone marrow biopsy tissue microarray	(Hohtari et al., 2019)	N/A
Healthy donor bone marrow biopsy tissue microarray	(Brück et al., 2018)	N/A
DLBCL patient tumor tissue microarray		N/A
Testicular DLBCL patient tumor tissue microarray	(Pollari et al., 2018)	N/A
Oligonucleotides		
CIITA sgRNA: 5'-GATATTGGCATAAGC CTCCC-3'	(Deuse et al., 2019)	N/A
CD70 sgRNA 1: 5'- GGCTACGTATCCATC GTGA-3'	(Izawa et al., 2017)	N/A
CD70 sgRNA 2: 5'- GTACACATCCA GGTGACGC-3'	(Izawa et al., 2017)	N/A
CD70 sgRNA 3: 5'- GCAGGCTGATGCT ACGGG-3'	(Izawa et al., 2017)	N/A
Control sgRNA: 5'- GGGTAGAATACCC CCATT-3'	(Morgens et al., 2017)	N/A
CIITA target site primer F: CTTAAC AGCGATGCTGACCCC	(Deuse et al., 2019)	N/A
CIITA target site primer R: TGGCCTCC ATCTCCCCTCTCTT	(Deuse et al., 2019)	N/A
Recombinant DNA		
lentiCRISPRv2	(Sanjana et al., 2014)	RRID: Addgene_52961
psPAX2	N/A	RRID: Addgene_12260
pCMV-VSV-G	(Stewart et al., 2003)	RRID: Addgene_8454
pLEX_307	N/A	RRID: Addgene_41392
CD27 cDNA	N/A	ORFeome
Chemicals, Peptides, and Recombinant Proteins		
Decitabine	Selleck Chemicals	Cat#: S1200; CAS: 2353-33-5
Recombinant human interferon gamma	PeproTech	Cat#: 300-02
DAPI (4′,6-Diamidine-2′-phenylindole dihydrochloride)	Sigma-Aldrich	Cat#: 10236276001
ProLong Gold	Thermo Fisher Scientific	Cat#: P36934; RRID: SCR_015961
Alexa Fluor 488 Tyramide Reagent	Thermo Fisher Scientific	Cat#: B40953
Alexa Fluor 555 Tyramide SuperBoost Kit	Thermo Fisher Scientific	Cat#: B40933
Critical Commercial Assays		
miRNeasy kit	Qiagen	Cat#: 217004
Ribo-Zero™ rRNA Removal Kit	Illumina (formerly Epicentre)	Cat#: MRZH11124
SuperScript™ Double-Stranded cDNA Synthesis Kit	Thermo Fisher Scientific	Cat#: 11917010
ScriptSeq v2™ Complete kit	Illumina	Cat#: BHMR1224
Agencourt AMPure XP PCR purification system	Beckman Coulter	Cat#: A63881
Chromium Single Cell 30 Library & Gel Bead Kit v2	10x Genomics	Cat#: PN-120237
Chromium Single Cell 5' Library & Gel Bead Kit	10x Genomics	Cat#: PN-1000006

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and processed clinical, array, and sequence data (Hemap)	(Pölönen et al., 2019)	hemap.uta.fi
AML ERRBS methylation, gene expression, genetic alteration and clinical data	(Glass et al., 2017)	GEO: GSE86952
BeatAML gene expression, genetic alteration and clinical data	(Tyner et al., 2018)	dbGaP: phs001657.v1.p1
TCGA AML and DLBCL	TCGA Research Network	dbGaP: phs000178.v8.p7
CCLE RNA-seq, RRBS methylation data, and cell line annotations	(Barretina et al., 2012; Ghandi et al., 2019)	RRID: SCR_013836; https://portals.broadinstitute.org/ccle/data
DLBCL gene expression, genetic alteration, and clinical data	(Chapuy et al., 2018)	GEO: GSE98588
GTEx gene expression data	(GTEx Consortium, 2015)	RRID: SCR_013042; https://gtexportal.org/home/datasets
pre-B-ALL and T-ALL processed methylation data	(Nordlund et al., 2013)	GEO: GSE49031
CoMMpass MM gene expression, genetic alteration, and clinical data	(Manojlovic et al., 2017)	https://research.themmrf.org/rp/download
PanALL	(Gu et al., 2019)	https://pecan.stjude.cloud/proteinpaint/study/PanALL
DLBCL gene expression, genetic alteration, and clinical data (Reddy et al.)	(Reddy et al., 2017)	EGA: EGAS00001002606
MOLM13 and MV411 RNA-seq	This study	GEO: GSE150374
FIMM AML 8 scRNA-seq	This study	EGA: EGAS00001004444
Human cell atlas scRNA	(Hay et al., 2018)	https://preview.data.humancellatlas.org/
AML scRNA-seq data (van Galen et al.)	(van Galen et al., 2019)	GEO: GSE116256
CLL scRNA-seq data (Rendeiro et al.)	(Rendeiro et al., 2020)	GEO: GSE111014
Normal NK cell scRNA-seq data (Yang et al.)	(Yang et al., 2019)	GEO: GSE130430
Experimental Models: Cell Lines		
293FT	Thermo Fisher Scientific	Cat#: R70007; RRID: CVCL_6911
HL60	ATCC	Cat#: ATCC® CCL-240; RRID: CVCL_0002
MOLM13	DSMZ	Cat#: ACC-554; RRID: CVCL_2119
MV411	DSMZ	RRID: CVCL_0064
MONOMAC1	DSMZ	Cat#: ACC 252; RRID: CVCL_1425
NB4	DSMZ	Cat#: ACC 207; RRID: CVCL_0005
OCIAML2	DSMZ	Cat#: ACC 99; RRID: CVCL_1619
OCIAML3	DSMZ	Cat#: ACC 582; RRID: CVCL_1844
PL21	DSMZ	Cat#: ACC 536; RRID: CVCL_2161
SUDHL5	Leppä lab, University of Helsinki, Helsinki, Finland	RRID: CVCL_1735
Software and Algorithms		
affy 1.52.0	(Gautier et al., 2004)	RRID: SCR_012835; http://www.bioconductor.org/packages/release/bioc/html/affy.html
Bioconductor	N/A	RRID: SCR_006442; http://www.bioconductor.org/
CellProfiler 2.1.2	(Carpenter et al., 2006)	RRID: SCR_007358; https://cellprofiler.org/
CIBERSORT	(Newman et al., 2015)	https://cibersort.stanford.edu/
ComplexHeatmap	(Gu et al., 2016)	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
edgeR	(Robinson et al., 2010)	RRID: SCR_012802; http://bioconductor.org/packages/edgeR/
Fiji	(Schindelin et al., 2012)	RRID: SCR_002285; http://fiji.sc
FlowJo 10.0.8r1	Tree Star	RRID: SCR_008520
ggplot2	N/A	RRID: SCR_014601; https://ggplot2.tidyverse.org
glmnet 2.0.16	(Friedman et al., 2010)	RRID: SCR_015505; https://cran.r-project.org/web/packages/glmnet/index.html
GSEA	(Subramanian et al., 2005)	RRID: SCR_003199; http://software.broadinstitute.org/gsea
GSVA 1.24.0	(Hänzelmann et al., 2013)	https://bioconductor.org/packages/release/bioc/html/GSVA.html
FDb.InfiniumMethylation.hg19	N/A	http://bioconductor.org/packages/release/data/annotation/html/FDb.InfiniumMethylation.hg19.html
limma	(Ritchie et al., 2015)	RRID: SCR_010943; http://bioconductor.org/packages/release/bioc/html/limma.html
mclust 5.4	(Scrucca et al., 2016)	https://cran.r-project.org/web/packages/mclust/index.html
MCPcounter	(Becht et al., 2016)	https://github.com/ebecht/MCPcounter
methylSig	(Park et al., 2014)	http://sartorlab.ccmb.med.umich.edu/software
MSigDB	(Subramanian et al., 2005)	RRID: SCR_016863; http://software.broadinstitute.org/gsea/msigdb
Picard	http://broadinstitute.github.io/picard/	RRID: SCR_006525; http://broadinstitute.github.io/picard/
R	R Core Team	https://www.r-project.orgs
STAR	(Dobin et al., 2013)	RRID: SCR_015899; https://github.com/alexdobin/STAR
Subread	(Liao et al., 2013)	RRID: SCR_009803; http://subread.sourceforge.net/
survival 2.42-4	N/A	https://cran.r-project.org/web/packages/survival/index.html
Trimmomatic	(Bolger et al., 2014)	RRID: SCR_011848; http://www.usadellab.org/cms/index.php?page=trimmomatic
Seurat 3.1	(Stuart et al., 2019)	RRID: SCR_016341; https://cran.r-project.org/web/packages/Seurat/index.html
batchelor 1.01	(Haghverdi et al., 2018)	https://bioconductor.org/packages/release/bioc/html/batchelor.html
SingleR 1.01	(Aran et al., 2019)	https://github.com/dviraran/SingleR
CellPhoneDB	(Efremova et al., 2020)	RRID: SCR_017054;https://github.com/Teichlab/cellphonedb
ICE	(Hsiau et al., 2019)	https://ice.synthego.com
TrimGalore	N/A	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Bismark 0.19.1	(Krueger and Andrews, 2011)	https://www.bioinformatics.babraham.ac.uk/projects/bismark/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Satu Mustjoki (satu.mustjoki@helsinki.fi).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The accession numbers for the raw data generated in this study are GEO: GSE150374 (RNA-seq and RRBS data) and EGA: EGAS00001004444 (scRNA-seq data). Processed data, including reanalysis of published datasets are available at Synapse DOI: 10.7303/syn21991014. Scripts used to generate results are available at github: https://github.com/systemsgenomics/ImmunogenomicLandscape-BloodCancers. Hemap data can be queried and visualized at http://hemap.uta.fi.

Accessions for multi-omics data sets used in this study: CoMMpass: https://research.themmrf.org/rp/download (Manojlovic et al., 2017), BeatAML dbGaP: phs001657.v1.p1 (Supplemental Information from publication) (Tyner et al., 2018), TCGA AML and DLBCL dbGaP: phs000178.v8.p7, DLBCL GEO: GSE98588 (Chapuy et al., 2018), DLBCL EGA: EGAS00001002606 (Supplemental Information from publication and requested from authors) (Reddy et al., 2017), PanALL https://pecan.stjude.cloud/proteinpaint/study/PanALL (Gu et al., 2019), CCLE https://portals.broadinstitute.org/ccle/data (Barretina et al., 2012; Ghandi et al., 2019), GTEx https://gtexportal.org/home/datasets (GTEx Consortium, 2015), AML ERRBS GEO: GSE86952 (Glass et al., 2017), pre-B-ALL and T-ALL GEO: GSE49031 (Nordlund et al., 2013). Accession for scRNA data sets used in this study: AML GEO: GSE116256 (van Galen et al., 2019), CLL GEO: GSE111014 (Rendeiro et al., 2020), Normal NK cell GEO: GSE130430 (Yang et al., 2019), Human Cell Atlas https://preview.data.humancellatlas.org (Hay et al., 2018).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patients

RNA Sequencing, Flow Cytometry, and Cell Culture Experiments

Bone marrow (BM) aspirates from 133 AML patients for flow cytometry (of which 37 for paired RNA-seq) and 8 patients for scRNA-seq were collected at diagnosis after signed informed consent from each patient (permit numbers 239/13/03/00/2010, 303/13/03/01/2011, Helsinki University Hospital (HUH) Ethics Committee) in accordance with the Declaration of Helsinki. Characteristics of these patients are listed in Tables S1 (flow cytometry) and S3 (scRNA-seq).

Tissue Microarrays (TMA)

Tissue microarrays of CML, ALL, and AML cohorts have been previously described (Brück et al., 2018; Hohtari et al., 2019; Brück et al., 2020). These include diagnostic BM biopsies from AML (n=66), B-ALL (n=54), T-ALL (n=14), and CML (n=62) patients treated in the Department of Hematology, HUH between 2005-2015, and DLBCL biopsies treated at the Department of Oncology, HUH (n=233). In addition, BM biopsies taken in 2010 from subjects due to persistent abnormal leukocyte, erythrocyte, or platelet count and without diagnosis of hematological malignancy, chronic infection, nor autoimmune disorder in six years of follow-up were included as controls (n=11). Study subjects gave written informed research consent to the study and to the Finnish Hematology Registry. The study complied with the Declaration of Helsinki and the HUH ethics committee (permit number 303/13/03/01/2011). Fresh BM biopsies and lymphoma samples were formalin-fixed and paraffin-embedded (FFPE) in the Department of Pathology, HUSLAB and stored at the Helsinki Biobank at HUH. TMA blocks were constructed from up to four 1 mm cores from representative regions of tumor samples. RNA sequencing-defined molecular subtypes were available from a subset of DLBCL TMA patients (Reddy et al., 2017). Additionally, mIHC data from a previously published cohort of 74 primary testicular DLBCL patients were used (Pollari et al., 2018). Characteristics of the TMA patients are listed in Table S5.

Cell Lines

The MOLM13, MONOMAC1, MV411, OCIAML2, OCIAML3, NB4, and PL21 AML cell lines were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) and the HL60 cell line from the American Type Culture Collection (ATCC). SUDHL5 DLBCL cells were obtained from prof. Sirpa Leppä (University of Helsinki, Finland). Cells were cultured in RPMI-1640 (Lonza) with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin (R10). The cell lines were authenticated using GenePrint10 System (Promega) and confirmed to have an overall identity estimate of > 80% at the 18 tested alleles. The 293FT cell line was obtained from Thermo Fisher Scientific and cultured in DMEM (Lonza) with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin (Gibco) (D10).

The Jurkat triple parameter reporter (Jurkat-TPR) cell line containing NFAT-EGFP, NF-κB-ECFP, and AP-1-mCherry reporters (Jutz et al., 2016) and the Jurkat-AP-1-EGFP reporter cell line (Ratzinger et al., 2014) were obtained from Dr. Peter Steinberger and cultured in IMDM (Gibco/Thermo Fisher Scientific) with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin. T cell stimulator (TCS) cells and TCS-CD80 engineered to constitutively express CD80 (Leitner et al., 2010) were obtained from Dr. Peter Steinberger and cultured in R10.

METHOD DETAILS

Processing of Genome-Wide Multilevel Data

Each dataset and sample numbers used in the analysis are listed in Figure S1.

Hemap

The Hemap dataset includes 9,544 gene expression profiles collected across several studies from the Gene Expression Omnibus (GEO) database described in (Pölönen et al., 2019). The data, curated sample annotations and disease categories are available at http://hemap.uta.fi. Briefly, these data represent microarray data from the commonly used hgu133Plus2 platform that were processed using the RMA probe summarization algorithm (Irizarry et al., 2003) with probe mapping to Entrez Gene IDs (from BrainArray version 18.0.0, ENTREZG) and a bias-correction method (Eklund and Szallasi, 2008) to generate gene expression signal levels. To ensure representative samples for the immunological analyses, we performed filtering on the original dataset. Treatments that induced cell differentiation or activation of normal cells were kept, while other *ex vivo* treatments (617 samples), non-malignant cells from patients (417), or clusters representing only a single study (>90% samples with the same study ID, 38 samples) were excluded from the analysis performed here, resulting in a dataset of 8,472 samples. We also distinguished sorted (using FACS, isolated by magnetic beads using either positive or negative selection, or microdissected), unsorted, and cell line samples based on sample descriptions. Finally, clinical information was added when available (survival from GEO: GSE10358, GEO: GSE10846, GEO: GSE11877, GEO: GSE12417, and GEO: GSE14468; progression-free survival, sex, age, race, and tumor cell contents from GEO: GSE13314, GEO: GSE10846, GEO: GSE24080, and GEO: GSE19784). All annotations for the samples used are provided in Table S1.

TCGA

Processed data were retrieved into feature matrices for AML (The Cancer Genome Atlas Research Network, 2013) and DLBCL, containing expression, mutation, CNV, methylation, and clinical data, using TCGA Feature Matrix Pipeline and fmx-construction.sh command, available at https://github.com/cancerregulome/gidget. For representing the data at individual gene loci, level 3 RSEM RNA-seq and methylation data for each TCGA AML and DLBCL sample was obtained from firehose GDAC, run stddata__2015_11_01. Methylation data were annotated using FDb.InfiniumMethylation.hg19 R package to assign probes at TSS.

CCLE

RNA-seq read counts (RPKM) dated 2018.05.02, RRBS methylation data for CpG islands and TSS 1 kb dated 2018.06.14, and cell line annotations dated 2012.10.18 were downloaded from: https://portals.broadinstitute.org/ccle/data.

Other Multilevel Datasets

Affymetrix Human Genome U133 Plus 2.0 gene expression datasets (DLBCL: GEO: GSE98588 and AML: GEO: GSE6891) were normalized using affy 1.52.0 (Gautier et al., 2004) RMA and gene expression values obtained using Brainarray v18 probe mapping. Mutations, chromosomal rearrangements, and clinical and sample characteristics were obtained from Tables S2, S3, S4, and S5 for the DLBCL study (Chapuy et al., 2018). Clinical data, mutations, and sample annotations were obtained from Tables S1 and S2 for AML (Glass et al., 2017).

Genomic alteration and subtype assignments for DLBCL with 1,001 samples were downloaded from Tables S1 and S2 (Reddy et al., 2017). Quantile normalized FPKM gene expression data was obtained from the authors, resulting in a dataset of 624 samples with multilevel data.

Methylation data from pre-B-ALL and T-ALL samples in GEO: GSE49031 (processed beta values for each probe) were used in the analysis of *CIITA* expression and methylation (*CIITA* probe cg04945379).

BeatAML mutation, clinical, and sample annotation data were downloaded from source data (from Tables S3) (Tyner et al., 2018). The RNA-seq count matrix was obtained from the authors. Genes with log2 counts per million (CPM) level > 1 in over 1% of samples were voom transformed and quantile normalized using limma (Ritchie et al., 2015). Mutation status was assigned based on exome sequencing and clinical sequencing data. Only bone marrow samples were used in the statistical analysis of immunological features.

CoMMpass gene expression, mutations, clinical and sample annotation data were downloaded from the CoMMpass data portal (https://research.themmrf.org/rp/download). HTSeq gene counts were filtered to contain log2 CPM level > 1 in over 2.5 % of the samples and subsequently library sizes were normalized using calcNormFactors and expression values were voom transformed using limma. Testis antigen expression was called if log2 expression was above 3.

PanALL study HTSeq raw read counts and gene expression signatures for each pre-B-ALL subtypes were downloaded from: https://pecan.stjude.cloud/proteinpaint/study/PanALL (Gu et al., 2019). Library sizes were normalized using calcNormFactors and expression values were voom transformed using limma. Limma removeBatchEffect was used for removing batch effects related to RNA-seq library and chemistry used.

Sample Stratification Based on Gene Expression Profiles

Molecular subtypes were identified from the Hemap, TCGA AML, BeatAML datasets using an data-driven approach as previously described (Mehtonen et al., 2019). Briefly, the Barnes-Hut approximated version of t-SNE implementation (BH-SNE) (van der Maaten, 2014; van der Maaten and Hinton, 2008) was used with 15% most variable genes and perplexity set to 30 to perform dimensionality reduction. Kernel density-based clustering algorithm known as mean-shift clustering (Cheng, 1995) with bandwidth (BW) parameter set to 1.5 (subsets of data, one cancer type) or 2.5 (all data) was used (LPCM package in R) to cluster the data following the dimensionality reduction. To identify corresponding clusters in different datasets of the same cancer type, similarity in sample clustering between datasets was evaluated in a data-driven manner using GSVA (Hänzelmann et al., 2013) enrichment scores as previously described (Mehtonen et al., 2019). Briefly, top 20 positively and negatively correlated genes per cluster were used to identify similar clusters in a new dataset with significant enrichment for cluster specific genes.

Similarly, Hemap MM (GEO: GSE16716, GEO: GSE24080) and pre-B-ALL clustering and BH-SNE projection was performed using 12.5 % and 5 % variable genes and perplexity set to 30 using density BW parameter 1.5 and 1.25, respectively. 835 samples used for

subtype signature identification were used in the analysis. Gene sets for each subtypes were defined as having log2 FC >1 and FDR <0.001.

CoMMpass clustering and BH-SNE projection was performed for 12.5 % variable genes, perplexity set to 30 using two clustering density BW parameters (1 and 2) to identify larger sample groups and clusters with more resolution. Initial run revealed clustering by gender and therefore, genes with significantly higher expression by gender were removed prior to BH-SNE and clustering. This resulted in separation of molecular subtypes, characterized by fusion genes, hyperdiploidy and distinct mutation patterns. Molecular subtypes were validated using the Hemap MM subtypes using the gene set approach from (Mehtonen et al., 2019).

Statistical Analysis Using Discrete Gene Expression Features

For individual genes, discrete categories (high, low, and not detected) were assigned based on mixture model fit as described previously for Hemap and DLBCL GEO: GSE98588 datasets (Pölönen et al., 2019). Briefly, Gaussian finite mixture models were fitted by expectation-maximization algorithm (R package mclust version 4.3) (Scrucca et al., 2016). The model was chosen by fitting both equal and variable variance models and ultimately choosing the model which achieved a higher Bayesian Information Criterion (BIC) to avoid overfitting. To assure minimal amount of misclassifications of data samples to discrete categories, three additional rules were implemented. First, if the uncertainty value from the model was above 0.1, value of 0 was assigned to denote low level. Secondly, log2 expression values lower than 4 or higher than 10 were assigned to a value -1 and 1, respectively. Thirdly, genes without clear background distribution (gene is always expressed), or if over 90% of the samples had uncertain expression based on the model classification, categories were re-evaluated. If >60% of the uncertain samples had expression above or below 6, categories were assigned as 1, and -1, respectively. For binary classification, values -1 and 0 were merged as low/not detected expression and 1 as expressed for statistical evaluation.

Sample Group Specificity

Hypergeometric tests followed by Bonferroni adjustment of p values were used to estimate statistical enrichment of gene expression in a particular sample group. Secondly, a one-tailed Wilcoxon rank sum test was performed to test whether gene is expressed at a higher level. P value was corrected using Bonferroni if multiple groups were compared. Thirdly, fold change was computed between the tested groups.

Development of Immunological Scores

Cytolytic Score

To find genes most specifically expressed in $CD8^+$ T cells/NK cells, specificity to these cell types was evaluated with the sample group specificity tests as described above (hypergeometric test adjusted p value < 1e-5, fold change > 1.5, Wilcoxon rank sum test adjusted p value < 0.01) using Hemap samples. Genes with significantly higher expression compared to B cells, HSC, erythroid cells, macrophage, monocyte, and dendritic cells were kept, resulting in 46 $CD8^+$/NK cell-specific marker genes (Figure S1B). Known $CD8^+$ T cell/NK cell-specific genes *GZMA*, *GZMB*, *PRF1*, *GNLY*, *GZMH*, and *GZMM* were chosen for further exploration, as they are directly related to cytolytic activity of T/NK cells.

Sorted BCL and AML samples and cell lines from all cancers, excluding T cell-like TCL lymphomas (because of transcriptomic similarity of TCL cancer cells and T/NK cells), were used to check whether these genes are expressed in pure cancer cell populations by requiring that expression values from pure samples (probeset noise distribution) were well separated from T/NK cells with high expression of the genes. As a second criteria, unsorted BCL and AML populations were compared to sorted BCL and AML samples to inspect if unsorted populations have samples with higher expression of the gene, indicating that increased signal is coming from T/NK cells. *GZMB* was filtered out, as it was highly expressed in a subset of pure samples. Geometric mean of gene expression for *GZMA*, *PRF1*, *GNLY*, *GZMH*, *GZMM* was computed followed by log2 transformation to be used as a proxy of cytolytic activity in hematological cancers.

HLA Scores

To find genes related to HLA II antigen presentation, normal non-APC cells including CD4/CD8/regulatory/gamma-delta T cells, NK cells, erythroid lineage cells, and neutrophils were compared to normal APC cells including DCs, B cells, and macrophages to find HLA II genes that are expressed highly in APC cells, but not in any non-APC cells. Wilcoxon rank sum test was computed with option "greater" to find genes with higher expression in APC cells. Adjusted p value cutoff 0.001 and fold change cutoff > 4 were set to find significant genes, resulting in 350 genes. Genes were further filtered by computing fold change between each individual APC and non-APC cells to ensure genes are expressed higher in each APC cell type (median fold change > 2 to non-APC) resulting in total of 66 genes (Figure S4A and Table S4). Pairwise Spearman correlation of the genes overexpressed in APC was used to identify HLA II genes whose expression most highly correlated with each other using Hemap cancer samples (Figure S4B). The geometric mean of the HLA II genes *HLA-DMA*, *HLA-DMB*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*, and *HLA-DRB1* was defined as HLA II score.

Due to the ubiquitous expression of HLA I on all cell types, no filtering was necessary and the geometric mean of known HLA I genes *B2M*, *HLA-A*, *HLA-B*, and *HLA-C* was used to detect HLA I expression in the samples.

Validation of Immunological Scores

Flow Cytometry of AML Patient Samples

To analyze T and NK cell percentages in 37 AML BM samples, fresh diagnostic-phase BM aspirates in EDTA tubes were used. Antibodies according to Tables S1 and S4 were added to 50 μl of the BM sample, mixed and incubated for 15 min, washed with PBS +

0,1% NaN3, centrifuged, and the supernatant was discarded. Erythrocytes were lysed by incubating the sample in FACS Lysing Solution (BD) and the sample was washed with PBS + 0,1% NaN3, centrifuged, and the supernatant was discarded. The sample was resuspended into 0.5 ml FACSFlow Sheath Fluid (BD) and 200,000 events were acquired with FACSCanto (BD Pharmingen). Data were analyzed using FlowJo (10.0.8r1). For quantification of T/NK cells for cytolytic score validation, cell debris was excluded based on low forward scatter (FSC), lymphocytes were identified as CD45highSSClow cells, and T cells gated as $CD3^+$ and NK cells as $CD3^-CD2^+$ out of lymphocytes (Figure S1D). Percentage of the sum of T and NK cells was calculated out of all non-debris cells. For quantification of HLA-DR^+ AML blasts for HLA II score validation, cell debris was excluded based on low forward scatter (FSC), blasts were identified based on CD45 and SSC, and gated for HLA-DR positivity.

RNA Sequencing of AML Patient Samples and Cell Lines

RNA sequencing was performed from the same 37 AML patient samples. Briefly, total RNA (2.5-5 μg) was extracted from BM mononuclear cells obtained by Ficoll-Paque gradient centrifugation using the miRNeasy kit (Qiagen) and depleted of ribosomal-RNA (Ribo-Zero™ rRNA Removal Kit, Epicentre) after purification, then reverse transcribed to double stranded cDNA (SuperScript™ Double-Stranded cDNA Synthesis Kit, Thermo Fisher Scientific). Sequencing libraries were prepared with Illumina compatible Epicentre Nextera™ Technology and ScriptSeq v2™ Complete kit (Illumina) and were purified with SPRI beads (Agencourt AMPure XP, Beckman Coulter) and library QC was evaluated on High Sensitivity chips by Agilent Bioanalyzer (Agilent). Paired-end sequencing with 100 bp read length was performed using Illumina HiSeq 2000. The reads were preprocessed as described previously (Kumar et al., 2017). Briefly, Trimmomatic (Bolger et al., 2014) was used to correct read data for low quality, Illumina adapters, and short read-length. Filtered paired-end reads were aligned to the human genome (GRCh38) using STAR (Dobin et al., 2013) with the guidance of EnsEMBL v82 gene models. Default 2-pass per-sample parameters were used, except that the overhang on each side of the splice junctions was set to 99. The alignments were then sorted and PCR duplicates were marked using Picard, feature counts were computed using SubRead (Liao et al., 2013), feature counts were converted to expression estimates using Trimmed Mean of M-values (TMM) normalization (Robinson and Oshlack, 2010) in edgeR (Robinson et al., 2010), and lowly expressed genomic features with CPM value ≤ 1.00 were removed. Default parameters were used, with exception that reads were allowed to be assigned to overlapping genome features in the feature counting. The immunological scores were calculated from TMM- normalized expression estimates based on geometric mean of selected genes, as described above.

Validation Samples from Hemap

We used GEO: GSE13314 MALT lymphoma pathologic data from Chng et al. (Chng et al., 2009) (Table S1) to validate cytolytic score in lymphoma samples. GSMids with MALT immunohistochemistry data scored by a pathologist was compared to cytolytic score using Spearman's correlation.

Comparison to Other Immunological Scores

CIBERSORT (Newman et al., 2015) was computed for Hemap dataset with parameters perm=100 and QN=F. Similarly, R package "MCPcounter" command MCPcounter.estimate was used to infer immunology scores in Hemap dataset. GSVA was used to compute enrichment of Bindea et al. gene sets (Bindea et al., 2013) for Hemap data using parameter tau=0.25. Cytolytic score was correlated to CIBERSORT, MCP-counter, and Bindea et al. gene set scores using Pearson's correlation in Hemap data.

Gene Set Enrichment Analysis

Gene set enrichment analysis was computed using the command line version of GSEA (Subramanian et al., 2005). A total of 1645 genesets from MsigDB V5 c2 category gene sets (BIOCARTA, KEGG, REACTOME, SA, SIG, ST), MsigDB HALLMARKS, version 4 of NCI NATURE Pathway Interaction Database, and WIKIPW (6.2015) were used for enrichment analysis. Immunological scores were used as a continuous phenotype to rank genes using Pearson correlation as metric for ranking. Sample permutation and multiple hypothesis testing correction was done to obtain FDR for each gene set. Gene sets were limited to contain between 5 to 500 genes per gene set. Single sample enrichment score was assigned to significant pathways for visualization and for Bindea gene sets using GSVA package 1.24.0 (Hänzelmann et al., 2013) in R.

Analysis of Microenvironment Genes Correlated to Cytolytic Score

Spearman correlation between gene expression level and cytolytic score was computed for Hemap dataset using unsorted samples (DLBCL, CHL, MCL, FL, CLL, CML, AML, MDS, pre-B-ALL, T-ALL). FC between pure CTL/NK cells and unsorted cancers were used as a metric to distinguish genes that are likely to be CTL/NK cell-expressed, expressed in the stroma/microenvironment, or in cancer. Genes with FDR < 0.01, FC > 2 and Spearman's R > 0.2 were assigned as significant CTL/NK genes. Genes with FDR < 0.01, FC < -2 and Spearman's R ± 0.2 were assigned as significant stromal/cancer genes. Significant genes and assigned categories are shown in Table S2.

Average FC and average correlation were computed to aggregate disease level results for BCL (DLBCL, CHL, MCL, FL). FDR values (for correlation) were combined using Stouffer's method. Genes with FDR < 0.01, FC > 1.5 and Spearman's R > 0.2 were assigned as significant CTL/NK genes. Genes with FDR < 0.01, FC < -1.5 and Spearman's R ± 0.2 were assigned as significant stromal/cancer genes.

Analysis of Immunomodulatory Genes

Wilcoxon rank sum test was used for determining differential expression of immunomodulatory genes for each disease subtype and for each cohort (Table S5). Results were aggregated for visualization, by computing average FC and by combining FDR values for each gene using Stouffer's method. FDR values were discretized to five categories based on significance cutoffs for visualization (0.05, 0.01, 0.001, 1e-5, 1e-16).

Multiplexed Immunohistochemistry (mIHC)

General

TMA blocks were sliced in 3.5 μm sections on Superfrost objective slides. We used 0.1% Tween-20 diluted in 10 mM Tris-HCL buffered saline pH 7.4 as washing buffer.

Tissue Preparation

Tissue deparaffinization and rehydration was performed in xylene and graded ethanol series. Then, heat-induced epitope retrieval (HIER) was carried out in 10 mM Tris-HCl - 1 mM EDTA buffer (pH 9) in +99°C for 20 min (PT Module; Thermo Fisher Scientific). Tissue peroxide quenching with 0.9% H2O2 for 15 min was followed by protein blocking with 10% normal goat serum (TBS-NGS) for 15 min.

Staining

Primary antibody diluted in protein blocking solution as in (Tables S2 and S5) and secondary anti-mouse or anti-rabbit horseradish peroxidase-conjugated (HRP) antibody (Immunologic) diluted 1:1 in washing buffer were applied for 1h45min and 45 min, respectively. Tyramide signal was amplified (TSA; PerkinElmer) for 10 min. Primary antibodies and HRP activity were inactivated with HIER, followed by peroxide and protein block steps as described above. The second primary antibody with its matching HRP-conjugated secondary antibody diluted 1:5 in washing buffer were added and TSA signal amplified. We repeated HIER, peroxide block and protein block and applied two additional primary antibodies immunized in different species overnight in +4°C. AlexaFluor647 and AlexaFluor750 fluorochrome-conjugated secondary antibodies (Thermo Fisher Scientific) diluted 1:150 in washing buffer (45 min) and 4′,6-diamidino-2-phenylindole counterstain (DAPI; Roche/Sigma-Aldrich) diluted 1:250 in TBS (15 min) were added. Last, ProLong Gold (Thermo Fisher Scientific) was used to mount slides.

Imaging

Fluorescent images were acquired with the AxioImager.Z2 (Zeiss) microscope equipped with Zeiss Plan-Apochromat 20x objective (NA 0.8). Scanned images were acquired and converted to JPEG2000 format (95% quality). For representative images shown in figures, image channels were recolored using Fiji (Schindelin et al., 2012; Schneider et al., 2012), brightness and contrast were adjusted using identical parameters for images acquired using the same antibody panel to maintain comparability and representative regions of the images were selected.

Image Analysis

Unfocused images were eliminated from the analysis. Cell segmentation and intensity measurements were computed based on adaptive Otsu thresholding and gradient intracellular intensity of grayscaled DAPI staining with the image analysis platform CellProfiler 2.1.2 (Carpenter et al., 2006). Cores with fewer than 1500 cells were eliminated from analysis. Cutoffs for marker positivity were based on staining intensity patterns of pooled cells of all samples and were confirmed visually (Tables S2 and S5, 'mIHC panel'). Counts of all cells and cells positive for marker combinations were averaged across multiple cores of each patient when available. Immune cells were quantified as proportion of positive cells to all cells (e.g. proportion of $CD68^+$ cells to total core cell count).

Genomic Correlations with Immunological Features

Feature Generation from Multilevel Data

Feature matrix generation, pairwise analysis run, and feature-specific filtering was based on the TCGA featurematrix pipeline available in https://github.com/cancerregulome/gidget/tree/master/commands/feature_matrix_construction. Continuous/discrete numeric data matrices were generated for the analyzed datasets Hemap, TCGA AML and DLBCL, Chapuy et al. DLBCL, and Glass *et al.* AML including clinical, genomic, and immunologic features. In case of categorical features, binary indicator features were generated to compare i) each categorical feature to the rest of the samples, ii) two categorical features to rest of the samples and iii) comparing two categorical features against each other. Feature types (gene expression, protein expression, clinical, methylation, CNV, mutations, sample annotation) were distinguished from each other. Missing values were assigned as NA. To account for differential methylation within the same locus, probes associated with each gene were first correlated to gene expression using Spearman's correlation ($P < 0.05$) and divided into positive and negatively correlated sets. Probes with standard deviation below 0.1 were removed. Mean methylation for these sets were computed to obtain two methylation features per gene, with positive and negative association to gene expression.

Feature Statistical Analysis

Spearman correlation was used for numeric-numeric feature pairs (in R use="pairwise.complete.obs"), while one-tailed Fisher's exact test of co-occurrence was used for binary-binary feature analysis. Wilcoxon rank sum test was used for numeric-binary variables. To keep the number of comparisons smaller and statistically reliable, only features with at least 5 observations were used in the analysis.

For p value adjustment, the number of features and intrinsic correlation are very different for different data pairs. Therefore, we performed separate statistical tests: The first test was to find whether features of a gene (methylation, mutation, CNV, gene expression) are correlated with each other (to identify alterations of the gene itself associated with each other), followed by Benjamini-Hochberg (BH) correction of the obtained p values. The second test was to assess whether features of a gene are correlated to features of

other genes (to identify e.g. driver alterations associated with immunological features) and included additional feature types, such as clinical variables and sample annotations, followed by BH correction of p values. Multiple hypothesis testing was performed separately for the correlation, Wilcoxon test and Fisher's test results, as these produce different p value distributions. Similarly, different significance level cutoffs were used for different data pairs. Methylation-methylation and CNV-CNV pairs were omitted. FDR cutoff was set to 0.1, except for mutations FDR < 0.25 was permitted. Further filtering criteria are specified in the result tables. This procedure allowed us to identify most relevant correlations and to filter out correlations difficult to interpret.

Single-cell RNA Sequencing

FIMM AML

Frozen Ficoll-purified bone marrow mononuclear cells from AML patients at diagnosis were thawed and prepared for scRNA-seq either directly or after staining for CD45 (CD45-PerCP clone 2D1, BD BioSciences), and sorting CD45-positive cells using an Influx cell sorter (BD Biosciences) to remove cell debris. The gel beads in emulsion (GEM) generation, cDNA amplification, and library preparation were performed according to the manufacturer's instructions using Chromium Single Cell 3' v2 Reagent Kit (samples 5238, 5750, 6187, 6333) and Chromium Single Cell V(D)J v1 Reagent Kit (samples 3667, 5897, 6386) (10x Genomics) with target loading of 7,000-25,000 cells per Chromium chip lane. 3' gene expression libraries were sequenced using Illumina HiSeq 2500 Rapid mode with paired-end reads 26 bp and 98 bp. 5' gene expression and VDJ libraries were sequenced using Illumina NovaSeq 6000 system with paired-end reads 26 bp and 91 bp or 150 bp and 150 bp, respectively.

The Cell Ranger v1.3 and v3.0.1 mkfastq, count and vdj analysis pipelines (10x Genomics) were used to demultiplex and convert Chromium single-cell sequencing barcode and read data to FASTQ files and to align reads and generate gene-cell matrices. The raw data were aligned to the GRCh38 reference genome. UMI counts were quantified using the 10X genomics Cell Ranger pipeline (v2.1.1 for 3' samples, v3.1 for 5' samples). R package HGNChelper was used for mapping the gene symbols to most recent symbols for both platforms, resulting in a total of 22,763 genes. Seurat (Stuart et al., 2019) R package v3 and R version 3.6 were used for scRNA data processing. Cells with > 10% mitochondrial gene counts, or less than 200 or more than 4,000 detected genes were filtered out. SCTransform with 3,000 variable features was used for data normalization. Batch correction for removing technical variation between patients was performed using fastMNN available in the batchelor 1.01 R package (Haghverdi et al., 2018). 50 PCs from fastMNN run were used for UMAP projection and clustering using the Seurat implementation of the Louvain's algorithm, using resolution set to 0.5. SingleR 1.0.1 (Aran et al., 2019) was used for the automated cell type annotation, using ENCODE/Blueprint annotations available in the package. Altogether, we analyzed a total of 30,579 AML BM cells, resulting in 20 distinct clusters. Clusters, UMAP coordinates, SingleR annotations, and cytolytic cell type annotations are included in Table S3.

HCA Normal Bone Marrow Cells

Raw fastq files were downloaded from https://preview.data.humancellatlas.org (Hay et al., 2018). Data were aligned with cellranger 3.0.2 with default parameters and the filtered count matrix was used for downstream analysis. Genes were filtered to include only genes present in more than 100 cells. Bad quality cells were removed if i) UMIs from mitochondrial genes in a cell was more than 10%, ii) total number of UMIs was 50,000 or more, or iii) number of genes expressed in a cell was 6,000 or more. Then, genes were once more filtered to include only those expressed in minimum in 400 cells.

Galen et al. AML Cells

scRNA profiles for 16 AML samples were downloaded from GEO: GSE116256 (van Galen et al., 2019). Data was processed similarly as FIMM AML. Cells with > 10% mitochondrial gene counts, or less than 100 or more than 3000 detected genes were filtered out. 30 PCs from fastMNN were used and data clustered using resolution set to 0.5.

Rendeiro et al. CLL scRNA

ScRNA profiles for 16 AML samples were downloaded from GEO: GSE111014 (Rendeiro et al., 2020). Data was processed similarly as FIMM AML. Cells with > 10% mitochondrial gene counts, or less than 200 or more than 4000 detected genes were filtered out. 22 PCs from fastMNN were used and data clustered using resolution set to 0.5.

Yang et al. Normal Bone Marrow NK Cells

Bone marrow NK cell count matrices (Yang et al., 2019) were downloaded from GEO: GSE130430. Data was processed similarly as FIMM AML data set. Cells with > 5% mitochondrial gene counts, or less than 200 or more than 2,500 detected genes were filtered out. 50 PCs from fastMNN were used and data clustered using resolution set to 0.8.

Integrated Analysis of T and NK Cells HCA and FIMM AML Cohorts

For comparison between AML and normal T cells, we generated a reference data set from normal BM T cells using 8 HCA samples and integrated these cells with T cells from AML BM, resulting in a total of 52,909 profiled cells (Figure 3I). fastMNN was used for data integration by using custom merge order of each sample. T cells from HCA samples were merged first, followed by incorporating T cells from AML samples. Similarly, for comparison between AML and normal NK cells, we integrated NK cells from HCA BM, normal BM sorted NK cells from 6 donors from Yang et al., and AML resulting in 26,601 NK cells. Respective data sets were merged and normalized together for the comparative analysis of NK and T cells, using log normalization and scaling to 10,000 counts per cell.

Statistical Analysis

Seurat FindAllMarkers and FindMarkers tools and Wilcoxon rank sum test and minimum log fold change >0.1 were used to compute the statistical significance between a group and rest of the cells or between groups.

scRNA Gene Set Analysis

Cytolytic and HLA scores, or NK and T cell identity related gene sets from Szabo et al. or Yang et al., respectively, were summarized using the geometric mean of the gene set genes and by excluding zero counts from the calculations. Differentially expressed genes for each AML scRNA cluster were associated with pathway profiles using Fisher's exact test.

CellPhoneDB

CellPhoneDB V2 (Efremova et al., 2020) was used for identifying potential ligand-receptor pairs between cytolytic cells and AML blasts for MDS-like samples with high cytolytic cell infiltration (5897, 3667). Each sample was analyzed individually using the clusters from the integrated FIMM AML map. Significant ligand-receptor pairs identified from both samples, with adjusted p value < 0.05 were extracted, requiring the ligand and receptor to be expressed in at least 10% of the cells.

Differential Methylation Analysis

ERRBS (Glass et al. AML)

GEO: GSE86952 raw aligned ERRBS AML methylation data were analyzed using the methylSig (Park et al., 2014) R package to identify methylation changes in patients with low HLA II expression or PML-RARA mutation compared to rest of the samples. Intersection of samples with both methylation, gene expression, and mutation data was 106 samples. Data files were read in R using the methylSigReadData command with parameters context="CpG" and destranded=TRUE. methylSigCalc command was run to obtain differentially methylated CpGs with parameters min.per.group=5 to require a minimum of 5 CpGs per group for calling differential methylation. All significant DMCs with FDR < 0.05 and absolute methylation change > 25% for *CIITA* genomic locus were obtained.

Illumina 450k (TCGA)

For comparison of differentially methylated immune checkpoint genes between TCGA AML and DLBCL samples, beta values of Illumina 450k probes within 1 kb of the transcription start site of the genes were averaged to obtain a single value representing methylation of the gene promoter area. M values were calculated from mean beta values using log2(beta/(1-beta)) and differential methylation analysis was performed using limma. For the corresponding differential gene expression analysis, RNA-seq read counts were converted to CPM and normalized using limma voom with quantile normalization, and differential expression analysis was performed using limma.

CCLE For correlation of gene expression to methylation in CCLE RRBS data, CpG and TSS 1 kb methylation values of each gene were averaged and the resulting mean methylation value was compared with gene expression using Spearman correlation.

Decitabine and IFNγ Treatment Experiments

Flow Cytometry

To assess the effect of decitabine or IFNγ treatment on HLA-DR expression by flow cytometry, cells were plated on flat-bottom 96-well plates at 50,000 cells/well (100,000 cell/well for primary AML cells) in a volume of 100 μl. For cell lines, the experiments were performed in R10 and for primary AML cells, StemSpan SFEM II (STEMCELL Technologies) supplemented with IL-6 (20 ng/ml), IL-3 (10 ng/ml), GM-CSF (20 ng/ml), G-CSF (20 ng/ml), SCF (50 ng/ml), IL-1b (10 ng/ml), FLT3L (50 ng/ml) (all from Peprotech), Stemregenin 1 (0.5 μM), and UM729 (1 μM) was used. Cells were treated with 10, 100, or 1,000 nM decitabine (Selleck) or DMSO as a control, both in the presence or absence of 10 or 100 ng/ml recombinant human interferon gamma (Peprotech), all conditions in triplicate wells. After 3 days, 25 μl of cell suspension (40 μl for primary AML cells) from each well was washed with 100 μl PBS-EDTA and stained with 5 μl of HLA-DR-FITC (clone G46-6, BD BioSciences) or isotype control (clone G155-178, BD BioSciences) antibodies in a total volume of 25 μl PBS-EDTA and antibodies. For primary AML cells, additionally 0.5 μl CD45-PerCP (clone 2D1, BD BioSciences) was used. Cells were then washed with 100 μl PBS-EDTA, resuspended to 50 μl PBS-EDTA, and 10,000 events were acquired on a FACSVerse flow cytometer (BD BioSciences). Flow cytometry data were analyzed using FlowJo 10.0.8r1. Viable cells were gated using forward (FSC-A) and side scatter (SSC-A), followed by gating for singlets using FSC-A and FSC-H. For primary AML cells, blasts were gated based on CD45 and SSC. HLA-DR$^+$ cells were gated such that untreated isotype control-stained cells contained < 1% positive cells. Final HLA-DR$^+$ cell percentages were obtained by subtracting isotype control-stained HLA-DR$^+$ cell percentages (mean of triplicate wells) from HLA-DR antibody-stained HLA-DR$^+$ cell percentages at each treatment condition.

RNA-seq and RRBS

To profile cells treated with decitabine or IFNγ using RNA sequencing or RRBS, cells were plated on 6-well plates at 1.5 million cells/well in a volume of 3 ml. Cells were treated with 100 nM decitabine or DMSO as a control, both in the presence or absence of 100 ng/ml recombinant human interferon gamma. After 3 days, the cultures were washed with PBS and split to two equal fractions. One fraction was pelleted and stored in -70°C for later DNA extraction using NucleoSpin Tissue kit (Macherey-Nagel) and the other fraction was used for RNA extraction using miRNeasy Mini kit (Qiagen). Three replicate experiments were performed on different days. RNA sequencing (lncRNA-seq) and RRBS were performed at Novogene.

RNA-seq Data Analysis

RNA-seq data were preprocessed as described above for AML patient samples. For differential expression analyses, lowly expressed genes with CPM value $\leq$ 1.00 in more than 75% of samples were removed. Read counts were normalized by Trimmed Mean of M-values (TMM) normalization (Robinson and Oshlack, 2010) and differential expression analysis was performed using the glmQLFit and glmQLTest functions in edgeR (Robinson et al., 2010).

RRBS Data Analysis

Poor quality and adapter containing bases were trimmed using TrimGalore in RRBS mode using stringency 3. Quality processed reads were mapped to the human genome (hg38 v2) using Bismark (Krueger and Andrews, 2011). Mapped reads were processed with Bismark methylation caller to estimate the methylation status of each cytosine in every mapped read. RnBeads version 2.2.0 was used for RRBS data quality control, preprocessing, and differential methylation analysis. Coverage threshold was set to 10 and the *t*-test was used for differential methylation analysis, using the built-in annotations for hg38. In-addition, CpG islands were expanded to +-2000 bp for CpG island shores and additional +-2000 bp for CpG island shelves as in methylSig (Park et al., 2014).

CRISPR-Cas9-Mediated Gene Silencing

Constructs encoding guide RNAs targeting *CIITA* (Deuse et al., 2019) and *CD70* (Izawa et al., 2017) and a control sgRNA targeting a safe genomic region (Morgens et al., 2017) were cloned into lentiCRISPRv2 (a gift from Feng Zhang, Addgene # 52961) (Sanjana et al., 2014). To produce lentivirus, 10 μg of the plasmid was transfected into 293FT cells seeded on the previous day at 3.8 million cells/T-75 flask together with 7.5 μg of psPAX2 (a gift from Didier Trono, Addgene plasmid # 12260), and 2.5 μg of pCMV-VSV-G (a gift from Bob Weinberg, Addgene plasmid # 8454) (Stewart et al., 2003) using 40 μl Lipofectamine 2000 (Thermo Fisher Scientific) according to manufacturer's instructions. After 6 h incubation, the culture medium was replaced with D10 containing 1% bovine serum albumin (BSA). After 60 h, viral supernatant was harvested, filtered using a 0.45 μm filter and stored in -70°C. For primary cell transductions, the lentiviral supernatant was concentrated by overnight centrifugation at 5,000 g. To transduce cells, 1.5 million cells (3 million primary AML cells) were suspended in 500 μl viral supernatant (50 μl concentrated virus with 450 μl culture medium for primary AML cells) and 8 μg/ml Polybrene in a 24-well plate well, centrifuged at room temperature at 800 g for 2 h, after which virus was washed away. Cells were selected using 0.5 μg/ml puromycin for prior to experiments.

For analysis of Cas9-induced mutations in *CIITA*, 1 million cells were harvested and genomic DNA was extracted using NucleoSpin Tissue kit (Macherey-Nagel). PCR was prepared using primers flanking the sgRNA target site in a volume of 20 μl containing 100 ng of sample DNA, 4 μl of 5× HF buffer (Thermo Fisher Scientific), and final concentrations of 0.02 U/μl Phusion polymerase (Thermo Fisher Scientific), dNTP at 200 μM each and 0.5 μM of each locus-specific primer, brought to a final volume with water. Samples were amplified by Applied Biosystems 2720 Thermal Cycler using the following program: initial denaturation at 98°C 30 s, 30 cycles at 98°C for 10 s, at 62°C for 20 s, and at 72°C for 15 s, and the final extension at 72°C for 5 minutes. Sanger sequencing was performed and editing efficiency was analyzed using ICE (Hsiau et al., 2019) on the sequence trace files. This resulted in 45-44% editing efficiency (25-28% frameshift frequency) in MOLM13 and 34-41% editing efficiency (28-35% frameshift frequency) in MV411, depending on the sequencing direction used for analysis.

Immune Checkpoint Gene List Curation

A list of co-stimulatory and co-inhibitory ligands and immunomodulatory enzymes known to be expressed on APCs or T/NK cell target cells was curated based on the literature. T cell ligands were based largely on a comprehensive review on T cell co-stimulation (Chen and Flies, 2013), and the list was supplemented with NK cell receptor ligands and immunomodulatory enzymes. References for the receptor-ligand interactions and their stimulatory or inhibitory effect on T/NK cells are listed in Table S5.

TCR Signaling Reporter Experiments

A CD27-expressing Jurkat-TPR cell line (Jurkat-TPR-CD27) was generated by transducing Jurkat-TPR cells as described above with lentivirus containing CD27 cDNA (ORFeome Library; Genome Biology Unit supported by HiLIFE and the Faculty of Medicine, University of Helsinki, and Biocenter Finland) in the pLEX_307 vector. For the stimulation experiments with SUDHL5 cells performed on a flat-bottom 96-well plate in 100 μl R10, Jurkat-TPR-CD27 cells were plated at 50,000 cells/well either alone or with 15,000 TCS-Ctrl or TCS-CD80 cells, and additionally with or without 50,000 SUDHL5 cells transduced with CD70-targeting or control CRISPR constructs. After 24 h incubation, 50 μl of cell suspension from each well was washed with 100 μl PBS-EDTA and stained with CD45.2-APC (clone 104, BioLegend, 0.1 μl/well) and CD19-APC-Cy7 (clone SJ25C1, BD BioSciences, 0.5 μl/well) in a total volume of 25 μl PBS-EDTA and antibodies. Cells were then washed with 100 μl PBS-EDTA, resuspended to 50 μl PBS-EDTA, and samples were acquired on a iQue Screener PLUS flow cytometer (Sartorius). Viable cells were gated based on FSC/SSC, followed by gating singlets based on FSC-A and FSC-H and excluding TCS and SUDHL5 cells based on negativity for CD45.2 and CD19. Cells positive for NFAT-EGFP, NFκB-ECFP, and double positive cells were then gated, and cells positive for NFκB-ECFP were visualized in Figure 5G.

Antigen Expression Analysis

Genes expressed only in normal or cancer cells were identified using the sample group specificity tests were performed as described above (hypergeometric test adjusted p value < 1e-2, fold change > 1.25, Wilcoxon rank sum test adjusted p value < 1e-5) using Hemap samples. Genes expressed in cancer were required to be expressed highly in > 5% of the patients in at least one disease based on mixture model categories (high vs. low expressed/not detected) and not expressed in normal cells. GTEx database (GTEx Consortium, 2015) V6 RNA-seq gene median RPKM values for each tissue were used to find genes specific to testis when compared to other tissues (excluding ovary). Testis genes were defined to have < 0.25 RPKM expression in all other tissues, resulting in a total of 1,563 genes common between Hemap and GTEx datasets. This list of genes was filtered to contain only coding genes expressed in Hemap cancer samples, resulting in 59 genes. CCLE cell line data for hematological cancers was also used to filter out genes not

expressed > 0.5 RPKM levels in at least 5 cell lines, resulting in a final antigen genelist containing 27 genes. Number of expressed CGA genes was computed for each sample using mixture model-discretized gene expression values for Hemap data, and using a cutoff of 0.25 RPKM for CCLE data. CGA number was used for survival analysis and discrete categories for different numbers of CGAs expressed were generated for categorical analysis. For Chapuy et al. DLBCL, patient LS2208 with testicular DLBCL was omitted from the analysis.

Survival Analysis

Univariate Analysis

Cox regression available in R package 'survival' version 2.42-4 was applied for univariate analyses of numeric immunology scores, including HLA I score, HLA II score, and CGA number in myeloma, AML, and DLBCL datasets, and cytolytic score in AML and DLBCL where unsorted samples were available (for myeloma GEO: GSE19784, GEO: GSE16716, GEO: GSE24080, CoMMpass for DLBCL GEO: GSE10846, GEO: GSE11318, GEO: GSE17372, GEO: GSE98588, Reddy et al. and for AML GEO: GSE10358, GEO: GSE12662, GEO: GSE12417, GEO: GSE14468, GEO: GSE6891, BeatAML, TCGA AML). Furthermore, all co-stimulatory genes and individual CGAs were included in the analysis. Additionally, stroma/microenvironment and $CD8^+$/NK cell expressed genes (as identified in the analysis of microenvironment genes correlated to cytolytic score in Figures 2 and S2 related analysis) and MDS-signature genes (Figure 3 and S3 related analysis) were included in the AML univariate analysis. Similarly, significant microenvironment related genes (Rho >0) were included in DLBCL analysis. Well-known prognostic markers, including ISS 1-3 for myeloma and IPI 0-5 and ABC/GCB subtypes for DLBCL and age in AML were also included in the analysis.

Multivariable Analysis

Features associated with overall survival in univariate analysis were selected for multiple regression analyses to evaluate their combinatorial prognostic effect in MM, DLBCL and AML datasets. Features were filtered using an adjusted p value cutoff 0.2 to reduce the number of features for the multiple regression analysis and to decrease the false discovery rate. Only significant features with similarly signed coefficients in at least two cohorts were used in the analysis. Regularized Cox regression model available in glmnet 2.0.16 R package (Friedman et al., 2010) was used to fit the Cox model. L1 and L2 norm ratio (alpha parameter) was optimized using 10-fold cross-validation and alpha values from 0 to 0.2 with 0.05 increments. To reduce variability in the model, cross-validation for each alpha value was iterated 100 times. Alpha and lambda with the lowest mean fitting error were used for the final model fitting. Independent test datasets were used for model validation. Hemap RCHOP-treated samples from GEO: GSE10846 and GEO: GSE17372 were used for training and GEO: GSE98588 for model validation in DLBCL, and for myeloma GEO: GSE19874 and GEO: GSE24080 were used for model training and CoMMpass cohort was used for model validation. Prognostic index (PI) was computed for each sample as in (Royston and Altman, 2013) for training and validation sets. Univariate Cox proportional hazards model and Kaplan-Meier plots were used to compare model performance between training and test sets using the immunological PI. PI with a similar hazard ratio and a low overall p value were used to verify that a set of distinct immunology features could be used to distinguish different patient outcomes independently of the dataset where the model was generated. Model performance was also evaluated by assessing the models ability to further stratify patients within known prognostic subgroups, including the IPI categories 1-5 and ABC/GCB subtypes for DLBCL, ISS 1-3 risk groups for MM and ELN2017 AML risk classification available in BeatAML cohort.

Data Visualization

R package "ComplexHeatmap" (Gu et al., 2016) was used for drawing heatmaps and oncoprints and "ggplot2" for drawing boxplots, barplots, and dot plots. Gene expression Z-scores were used for t-SNE map visualization to denote samples with low and high expression (low: < -2 to -1 and high 1 to > 2). For Hemap dataset, e-staining was used for gene expression visualization for mixture model components (not detected, low, and high) as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical details of all experiments are reported in the text, figure legends, and figures, including statistical significance, and counts. The Spearman's correlation and the correlation test for significance were used for the statistical analysis between two numeric variables. The non-parametric two-sided Wilcoxon rank sum test was used for statistical analysis between two groups. For survival analysis, statistical analysis between Kaplan-Meier curves was performed using the log-rank test and the log-rank score test was used for Cox proportional hazards models. If other statistical tests were used, it is reported in the figure legends. Significance codes correspond to p values or FDR as follows: ns $\geq$ 0.05, * < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001. In boxplots, horizontal line indicates the median, boxes indicate the interquartile range, and whiskers extend from the hinge to the smallest/largest value at most 1.5 * IQR of the hinge.

Cancer Cell, Volume *38*

Supplemental Information

Immunogenomic Landscape of Hematological Malignancies

Olli Dufva, Petri Pölönen, Oscar Brück, Mikko A.I. Keränen, Jay Klievink, Juha Mehtonen, Jani Huuhtanen, Ashwini Kumar, Disha Malani, Sanna Siitonen, Matti Kankainen, Bishwa Ghimire, Jenni Lahtela, Pirkko Mattila, Markus Vähä-Koskela, Krister Wennerberg, Kirsi Granberg, Suvi-Katri Leivonen, Leo Meriranta, Caroline Heckman, Sirpa Leppä, Matti Nykter, Olli Lohi, Merja Heinäniemi, and Satu Mustjoki

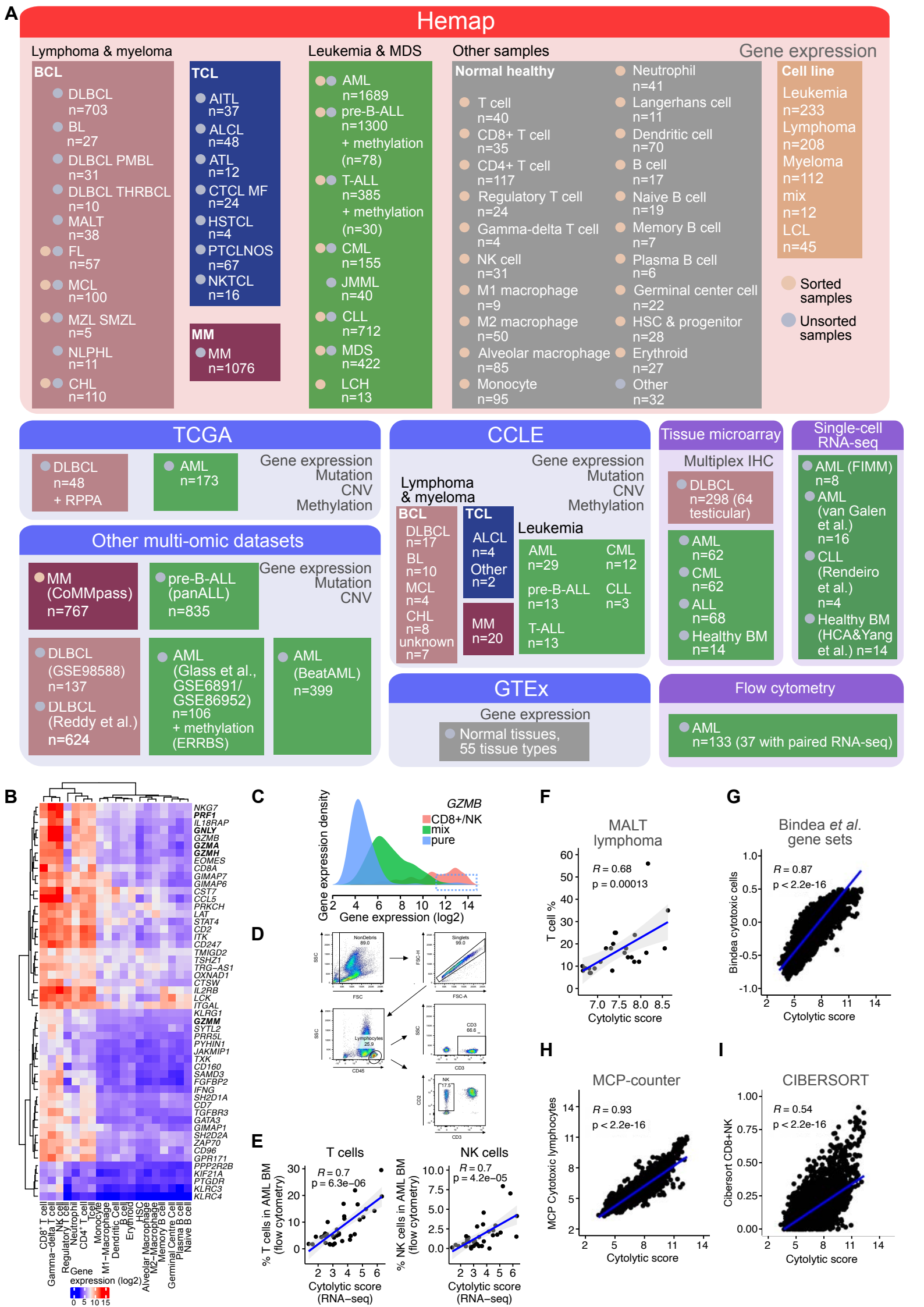

A
Hemap
Lymphoma & myeloma
BCL
DLBCL n=703
BL n=27
DLBCL PMBL n=31
DLBCL THRBCL n=10
MALT n=38
FL n=57
MCL n=100
MZL SMZL n=5
NLPHL n=11
CHL n=110
TCL
AITL n=37
ALCL n=48
ATL n=12
CTCL MF n=24
HSTCL n=4
PTCLNOS n=67
NKTCL n=16
MM
MM n=1076
Leukemia & MDS
AML n=1689
pre-B-ALL n=1300 + methylation (n=78)
T-ALL n=385 + methylation (n=30)
CML n=155
JMML n=40
CLL n=712
MDS n=422
LCH n=13
Other samples
Normal healthy
T cell n=40
CD8+ T cell n=35
CD4+ T cell n=117
Regulatory T cell n=24
Gamma-delta T cell n=4
NK cell n=31
M1 macrophage n=9
M2 macrophage n=50
Alveolar macrophage n=85
Monocyte n=95
Neutrophil n=41
Langerhans cell n=11
Dendritic cell n=70
B cell n=17
Naive B cell n=19
Memory B cell n=7
Plasma B cell n=6
Germinal center cell n=22
HSC & progenitor n=28
Erythroid n=27
Other n=32
Gene expression
Cell line
Leukemia n=233
Lymphoma n=208
Myeloma n=112
mix n=12
LCL n=45
Sorted samples
Unsorted samples
TCGA
DLBCL n=48 + RPPA
AML n=173
Gene expression
Mutation
CNV
Methylation
Other multi-omic datasets
MM (CoMMpass) n=767
pre-B-ALL (panALL) n=835
Gene expression
Mutation
CNV
DLBCL (GSE98588) n=137
DLBCL (Reddy et al.) n=624
AML (Glass et al., GSE6891/ GSE86952) n=106 + methylation (ERRBS)
AML (BeatAML) n=399
CCLE
Gene expression
Mutation
CNV
Methylation
Lymphoma & myeloma
BCL
DLBCL n=17
BL n=10
MCL n=4
CHL n=8
unknown n=7
TCL
ALCL n=4
Other n=2
MM n=20
Leukemia
AML n=29
CML n=12
pre-B-ALL n=13
CLL n=3
T-ALL n=13
Tissue microarray
Multiplex IHC
DLBCL n=298 (64 testicular)
AML n=62
CML n=62
ALL n=68
Healthy BM n=14
Single-cell RNA-seq
AML (FIMM) n=8
AML (van Galen et al.) n=16
CLL (Rendeiro et al.) n=4
Healthy BM (HCA&Yang et al.) n=14
GTEx
Gene expression
Normal tissues, 55 tissue types
Flow cytometry
AML n=133 (37 with paired RNA-seq)
B
NKG7
PRF1
IL18RAP
GNLY
GZMB
GZMA
GZMH
EOMES
CD8A
GIMAP7
GIMAP6
CST7
CCL5
PRKCH
LAT
STAT4
CD2
ITK
CD247
TMIGD2
TSHZ1
TRG-AS1
OXNAD1
CTSW
IL2RB
LCK
ITGAL
KLRG1
GZMM
SYTL2
PRR5L
PYHIN1
JAKMIP1
TXK
CD160
SAMD3
FGFBP2
IFNG
SH2D1A
CD7
TGFBR3
GATA3
GIMAP1
SH2D2A
ZAP70
CD96
GPR171
PPP2R2B
KIF21A
PTGDR
KLRC3
KLRC4
CD8+ T cell
Gamma−delta T cell
NK cell
Regulatory T cell
Neutrophil
CD4+ T cell
Monocyte
M1−Macrophage
Dendritic Cell
B cell
Erythroid
HSC
Alveolar Macrophage
M2−Macrophage
Memory B cell
Germinal Centre Cell
Plasma B cell
Naive B cell
Gene expression (log2)
0 5 10 15
C
GZMB
CD8+/NK
mix
pure
Gene expression density
Gene expression (log2)
D
NonDebris
Singlets
Lymphocytes
CD45
CD3
NK
CD2
E
T cells
NK cells
% T cells in AML BM (flow cytometry)
% NK cells in AML BM (flow cytometry)
R = 0.7
p = 6.3e−06
R = 0.7
p = 4.2e−05
Cytolytic score (RNA−seq)
F
MALT lymphoma
R = 0.68
p = 0.00013
T cell %
Cytolytic score
G
Bindea et al. gene sets
R = 0.87
p < 2.2e-16
Bindea cytotoxic cells
Cytolytic score
H
MCP-counter
R = 0.93
p < 2.2e-16
MCP Cytotoxic lymphocytes
Cytolytic score
I
CIBERSORT
R = 0.54
p < 2.2e-16
Cibersort CD8+NK
Cytolytic score

Figure S1. Schematic of samples included in the study and assessment of cytotoxic lymphocyte infiltration in hematological malignancies using gene expression data. Related to Figure 1.

(A) Schematic of all the datasets, sample numbers, and data types used in the analyses. Names of different multi-omics datasets are shown in the figure.

(B) Median expression of genes most enriched in normal CTLs and NK cells shown as a heatmap. Rows and columns are clustered using Euclidean distance and complete linkage.

(C) Density plot of *GZMB* gene expression in purified cancer cells, unsorted cancer samples, and $CD8^+$ T cell/NK cell populations. The plot indicates a subset of purified cancer samples highly expressing *GZMB*, demonstrating ectopic expression of *GZMB* leading to its exclusion as a cytolytic marker.

(D) Flow cytometry gating strategy to identify percentages of T cells ($CD3^+$) and NK cells for comparison with RNA-seq data (related to Figure 1C).

(E) Validation of cytolytic score in AML. Correlation (Spearman) of cytolytic score obtained from RNA-seq and T or NK cell fraction obtained by flow cytometry from 35 AML BM samples. The gray shading represents the 95% confidence interval.

(F) Cytolytic score validation in Hemap MALT lymphoma. Correlation (Spearman) between cytolytic score and reported T cell percentage based on pathological analysis. The gray shading represents the 95% confidence interval.

(G-I) Scatterplots comparing cytolytic score and Bindea *et al.* "Cytotoxic cells" gene set score, MCP-counter "Cytotoxic lymphocytes", and CIBERSORT sum of "T cells CD8", "NK cells resting", and "NK cells activated" fractions. Spearman correlation between each comparison is shown.

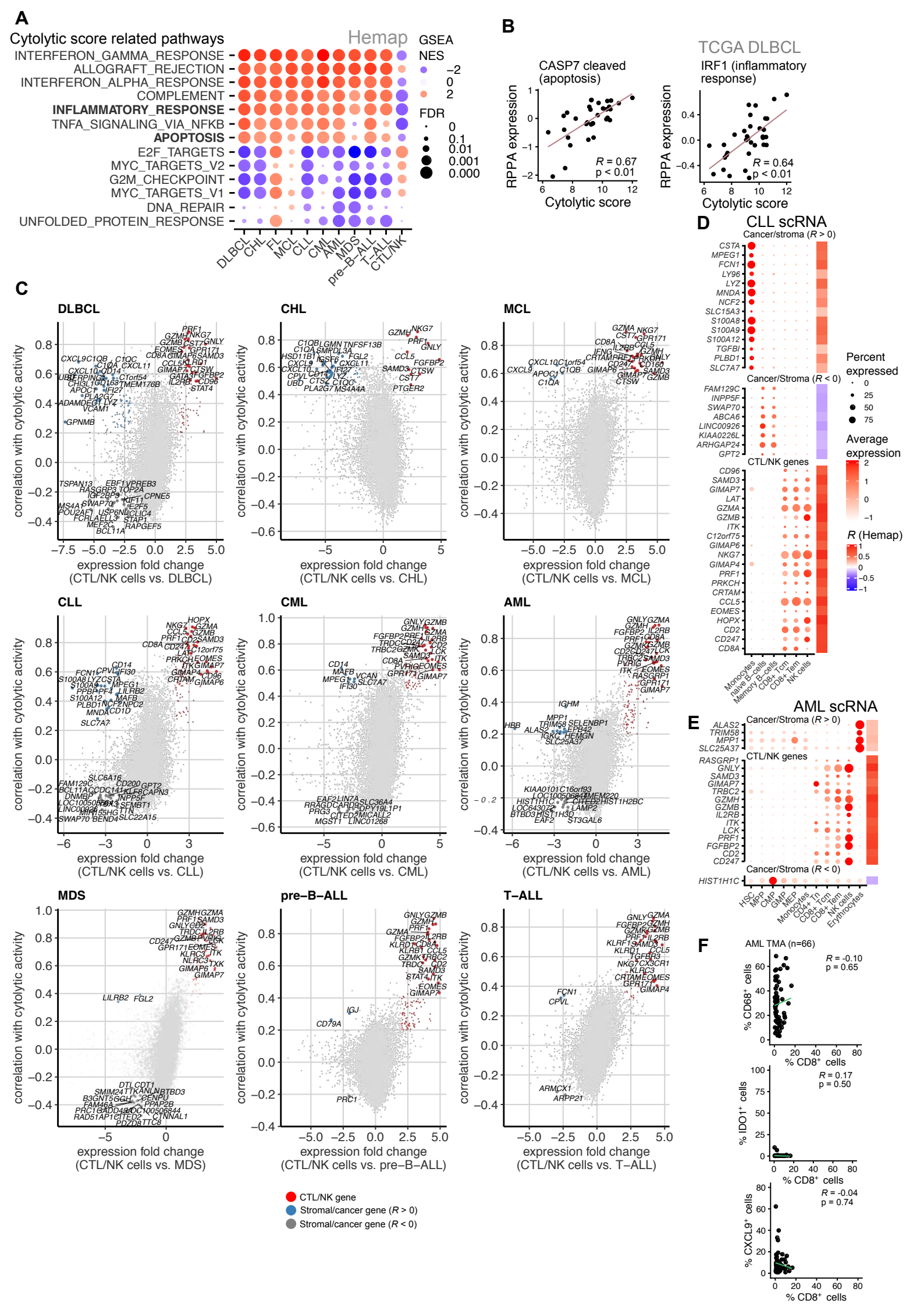

A
Cytolytic score related pathways
Hemap
GSEA NES
FDR
INTERFERON_GAMMA_RESPONSE
ALLOGRAFT_REJECTION
INTERFERON_ALPHA_RESPONSE
COMPLEMENT
INFLAMMATORY_RESPONSE
TNFA_SIGNALING_VIA_NFKB
APOPTOSIS
E2F_TARGETS
MYC_TARGETS_V2
G2M_CHECKPOINT
MYC_TARGETS_V1
DNA_REPAIR
UNFOLDED_PROTEIN_RESPONSE
DLBCL CHL FL MCL CLL CML AML MDS pre–B–ALL T–ALL CTL/NK
B
TCGA DLBCL
CASP7 cleaved (apoptosis)
IRF1 (inflammatory response)
RPPA expression
Cytolytic score
R = 0.67
p < 0.01
R = 0.64
p < 0.01
C
DLBCL
CHL
MCL
CLL
CML
AML
MDS
pre–B–ALL
T–ALL
correlation with cytolytic activity
expression fold change (CTL/NK cells vs. DLBCL)
expression fold change (CTL/NK cells vs. CHL)
expression fold change (CTL/NK cells vs. MCL)
expression fold change (CTL/NK cells vs. CLL)
expression fold change (CTL/NK cells vs. CML)
expression fold change (CTL/NK cells vs. AML)
expression fold change (CTL/NK cells vs. MDS)
expression fold change (CTL/NK cells vs. pre–B–ALL)
expression fold change (CTL/NK cells vs. T–ALL)
CTL/NK gene
Stromal/cancer gene (R > 0)
Stromal/cancer gene (R < 0)
D
CLL scRNA
Cancer/stroma (R > 0)
Cancer/Stroma (R < 0)
CTL/NK genes
Percent expressed
Average expression
R (Hemap)
Monocytes
naive B-cells
Memory B-cells
CD8+ Tcm
CD8+ Tem
NK cells
E
AML scRNA
Cancer/Stroma (R > 0)
CTL/NK genes
Cancer/Stroma (R < 0)
HSC
MPP
CMP
GMP
MEP
Monocytes
CD4+ Tn
CD8+ Tcm
CD8+ Tem
NK cells
Erythrocytes
F
AML TMA (n=66)
% CD68+ cells
% CD8+ cells
R = -0.10
p = 0.65
% IDO1+ cells
R = 0.17
p = 0.50
% CXCL9+ cells
R = -0.04
p = 0.74

Figure S2. Gene expression correlations with cytolytic score across hematological malignancies. Related to Figure 2.

(A) Dot plot of gene set enrichment analysis (GSEA) normalized enrichment scores (NES, point color) and FDR (point size) for selected significantly enriched MsigDB HALLMARKS pathways correlated with cytolytic activity across cancers. Pathway enrichment scores between pure CTL/NK and other normal cells are also shown. Pathways related to panel B are bolded.

(B) TCGA DLBCL RPPA protein expression correlations with cytolytic score to validate pathway enrichment profiles highlighted in A for DLBCL.

(C) Scatter plots of significant (FDR < 0.01, absolute FC > 2, absolute R > 0.2) genes correlated with cytolytic score in each cancer type. Maximum of 25 genes from each category are shown. CTL/NK genes (red), cancer/stroma R > 0 (blue), cancer/stroma R < 0 (dark grey). Full list of significant genes is shown in Table S2.

(D) Dot plot of cell type assignment for CLL scRNA expressed genes as identified in C. Dot plot showing the average expression (in color) and percentage of cells with detectable expression (size of dot). Only genes significantly differentially expressed in any cell type are shown. Genes are classified into groups as in C. Correlation coefficient in Hemap bulk data set is shown on the right.

(E) Dotplot as in D for AML scRNA.

(F) Scatter plots comparing the percentage of $CD8^+$ cells (CTLs) out of total cells with the percentages of $CD68^+$ cells (macrophages), $IDO1^+$ cells, and $CXCL9^+$ cells out of total cells in the AML IHC cohort (n = 62). Spearman correlation coefficients and p values adjusted using the BH method are shown.

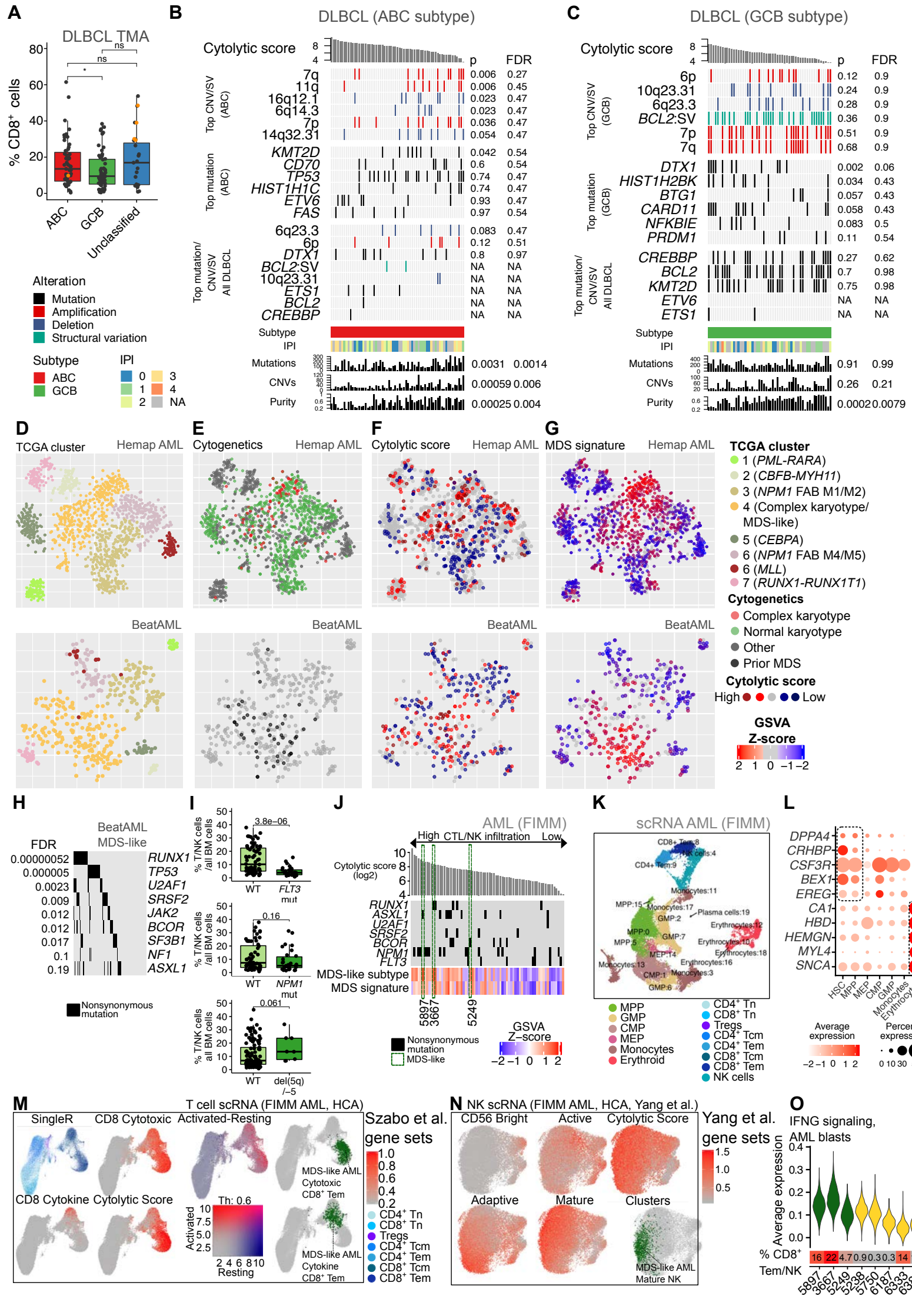

A
DLBCL TMA
% CD8+ cells
ns
ns
*
ABC
GCB
Unclassified
Alteration
Mutation
Amplification
Deletion
Structural variation
Subtype
ABC
GCB
IPI
0
1
2
3
4
NA
B
DLBCL (ABC subtype)
Cytolytic score
p
FDR
Top CNV/SV (ABC)
7q 0.006 0.27
11q 0.006 0.45
16q12.1 0.023 0.47
6q14.3 0.023 0.47
7p 0.036 0.47
14q32.31 0.054 0.47
Top mutation (ABC)
KMT2D 0.042 0.54
CD70 0.6 0.54
TP53 0.74 0.47
HIST1H1C 0.74 0.47
ETV6 0.93 0.47
FAS 0.97 0.54
Top mutation/ CNV/SV All DLBCL
6q23.3 0.083 0.47
6p 0.12 0.51
DTX1 0.8 0.97
BCL2:SV NA NA
10q23.31 NA NA
ETS1 NA NA
BCL2 NA NA
CREBBP NA NA
Subtype
IPI
Mutations 0.0031 0.0014
CNVs 0.00059 0.006
Purity 0.00025 0.004
C
DLBCL (GCB subtype)
Cytolytic score
p
FDR
Top CNV/SV (GCB)
6p 0.12 0.9
10q23.31 0.24 0.9
6q23.3 0.28 0.9
BCL2:SV 0.36 0.9
7p 0.51 0.9
7q 0.68 0.9
Top mutation (GCB)
DTX1 0.002 0.06
HIST1H2BK 0.034 0.43
BTG1 0.057 0.43
CARD11 0.058 0.43
NFKBIE 0.083 0.5
PRDM1 0.11 0.54
Top mutation/ CNV/SV All DLBCL
CREBBP 0.27 0.62
BCL2 0.7 0.98
KMT2D 0.75 0.98
ETV6 NA NA
ETS1 NA NA
Subtype
IPI
Mutations 0.91 0.99
CNVs 0.26 0.21
Purity 0.0002 0.0079
D
TCGA cluster
Hemap AML
BeatAML
E
Cytogenetics
Hemap AML
BeatAML
F
Cytolytic score
Hemap AML
BeatAML
G
MDS signature
Hemap AML
BeatAML
TCGA cluster
1 (PML-RARA)
2 (CBFB-MYH11)
3 (NPM1 FAB M1/M2)
4 (Complex karyotype/ MDS-like)
5 (CEBPA)
6 (NPM1 FAB M4/M5)
6 (MLL)
7 (RUNX1-RUNX1T1)
Cytogenetics
Complex karyotype
Normal karyotype
Other
Prior MDS
Cytolytic score
High Low
GSVA Z-score
2 1 0 −1 −2
H
BeatAML
MDS-like
FDR
0.00000052 RUNX1
0.000005 TP53
0.0023 U2AF1
0.009 SRSF2
0.012 JAK2
0.012 BCOR
0.017 SF3B1
0.1 NF1
0.19 ASXL1
Nonsynonymous mutation
I
% T/NK cells /all BM cells
3.8e−06
WT
FLT3 mut
0.16
WT
NPM1 mut
0.061
WT
del(5q) /−5
J
AML (FIMM)
High CTL/NK infiltration Low
Cytolytic score (log2)
RUNX1
ASXL1
U2AF1
SRSF2
BCOR
NPM1
FLT3
MDS-like subtype
MDS signature
5897
3667
5249
Nonsynonymous mutation
MDS-like
GSVA Z-score
−2 −1 0 1 2
K
scRNA AML (FIMM)
MPP
GMP
CMP
MEP
Monocytes
Erythroid
CD4+ Tn
CD8+ Tn
Tregs
CD4+ Tcm
CD4+ Tem
CD8+ Tcm
CD8+ Tem
NK cells
L
DPPA4
CRHBP
CSF3R
BEX1
EREG
CA1
HBD
HEMGN
MYL4
SNCA
HSC
MPP
MEP
CMP
GMP
Monocytes
Erythrocytes
Average expression
−2 −1 0 1 2
Percent expressed
0 10 30 50
M
T cell scRNA (FIMM AML, HCA)
SingleR
CD8 Cytotoxic
Activated-Resting
CD8 Cytokine
Cytolytic Score
Th: 0.6
Activated
Resting
MDS-like AML Cytotoxic CD8+ Tem
MDS-like AML Cytokine CD8+ Tem
Szabo et al. gene sets
CD4+ Tn
CD8+ Tn
Tregs
CD4+ Tcm
CD4+ Tem
CD8+ Tcm
CD8+ Tem
N
NK scRNA (FIMM AML, HCA, Yang et al.)
CD56 Bright
Active
Cytolytic Score
Adaptive
Mature
Clusters
MDS-like AML Mature NK
Yang et al. gene sets
O
IFNG signaling, AML blasts
Average expression
% CD8+ Tem/NK
16 22 4.7 0.9 0.3 0.3 14 1.7
5897
3667
5249
5238
5150
6187
6333
6386

Figure S3. Genetic alterations and molecular subtypes associated with cytolytic score. Related to Figure 3.

(A) Comparison of $CD8^+$ cell percentages out of all cells between DLBCL molecular subtypes based on quantitative multiplex immunohistochemistry performed on tissue microarrays. ABC and GCB subtypes are compared using a one-tailed Wilcoxon rank sum test. THRBCL cases are colored in yellow.

(B) Genetic alterations associated with cytolytic score in DLBCL patients of ABC subtype (n = 63, GSE98588) is shown as an oncoprint where columns corresponding to a sample are ranked by cytolytic score and mutations, copy number variations (CNVs), structural variations (SVs), cell of origin, IPI risk score, mutation burden, CNV number, and sample purity are plotted on rows. Discrete state/class is indicated as color (genetic aberrations and sample categories) and continuous values are represented as barcharts (score or percentage values). P values and FDR for correlations between cytolytic score and genetic aberrations are shown. Top CNVs/SVs correlated with cytolytic score in ABC are shown at the top, top mutations correlated with cytolytic score in ABC in the middle, and mutations and CNVs correlated with cytolytic score in all DLBCLs from Figure 3E are shown at the bottom.

(C) Genetic alterations associated with cytolytic score in DLBCL patients of GCB subtype (n = 54, GSE98588) are shown as in B.

(D-G) Visualization of Hemap AML and BeatAML samples using a t-SNE representation. (D) Clusters, (E) cytogenetics in Hemap AML and MDS status in BeatAML (F) cytolytic score and (G) MDS signature are colored on the t-SNE maps. Key characteristics of the clusters are annotated.

(H) MDS subtype-associated mutations are shown in the BeatAML dataset. FDR (Fisher's exact test) is shown on the left and genes are sorted based on significance.

(I) Combined flow cytometry-based percentage of T and NK cells (defined as $CD45^+CD3^+$ and $CD45^+CD3^-CD2^+$, respectively) out of all BM cells of 125 AML patients stratified by *FLT3* and *NPM1* mutations and deletion of 5q or chromosome 5.

(J) Cytolytic score, MDS-characteristic genetic alterations and MDS-like subtype related gene signature GSVA scores are shown for AML patients (FIMM AML cohort). MDS-like samples selected for scRNA-seq are marked as dark green rectangles.

(K) UMAP plot demonstrating 20 clusters identified in 8 AML scRNA-seq samples (FIMM AML). Main cell type comprising each cluster based on data-driven (SingleR) annotation is shown as color and text labels. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte–erythroid progenitor.

(L) Dot plot showing the average expression (in color) and percentage of cells with detectable expression (size of dot) of top MDS signature genes associated with HSC/MPP or erythroid lineages identified by SingleR in AML scRNA-seq samples.

(M) UMAP plots of T cells combined across scRNA-seq studies colored based on data-driven (SingleR) annotation (left), Szabo et al. gene set (Szabo et al., 2019) average expression (grey indicating low score and individual gene set scores visualized in tones of red; comparison of activated-resting as blended heatmap where activation score in tones of red and resting score in blue), single-cell cytolytic score, and locations of MDS-like AML $CD8^+$ activation states (in green) after clustering.

(N) UMAP plots of NK cells combined across scRNA-seq studies showing their characteristics colored based on gene set scores (red tones indicate high score) (Yang et al., 2019), single-cell cytolytic score, and position of MDS-like AML mature NK cells (in green) after clustering.

(O) Interferon gamma signaling gene set average expression in each scRNA-profiled AML sample. Average gene expression was computed using only HSC/MPP/GMP/CMP annotated cells corresponding to AML blasts. Percentages of $CD8^+$ Tem or NK cells out of all cells in the samples are shown below. MDS-like AML cases are shown in green. The width of a violin plot indicates the kernel density of the average expression values and the height the min and max values.

Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 * IQR from the quartiles.

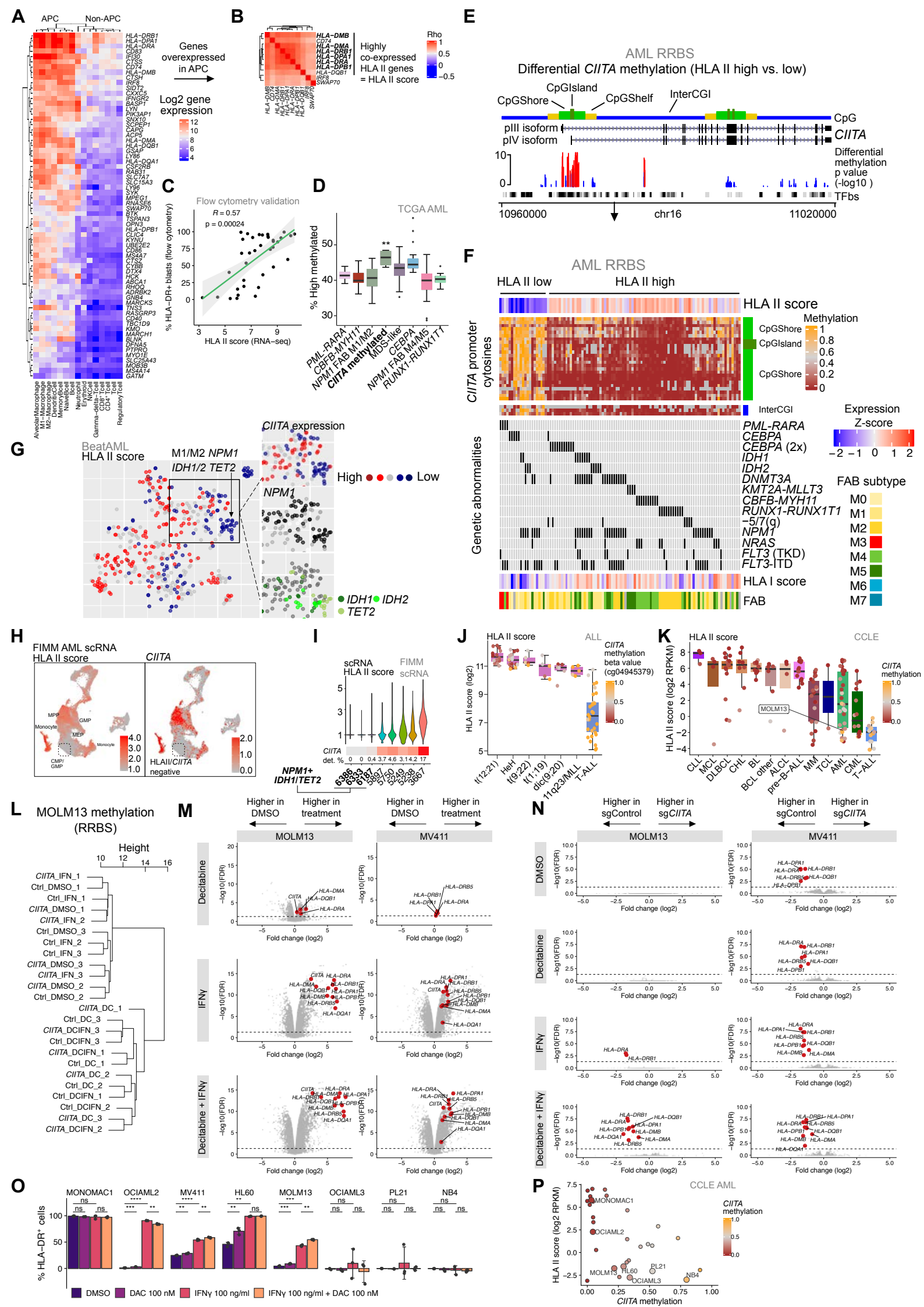

A
APC
Non-APC
Genes overexpressed in APC
Log2 gene expression
B
Highly co-expressed HLA II genes = HLA II score
Rho
C
Flow cytometry validation
R = 0.57
p = 0.00024
% HLA−DR+ blasts (flow cytometry)
HLA II score (RNA−seq)
D
TCGA AML
% High methylated
PML−RARA
CBFB−MYH11
NPM1 FAB M1/M2
CIITA methylated
MDS−like
CEBPA
NPM1 FAB M4/M5
RUNX1−RUNX1T1
E
AML RRBS
Differential CIITA methylation (HLA II high vs. low)
CpGIsland
CpGShore
CpGShelf
InterCGI
CpG
pIII isoform
pIV isoform
CIITA
Differential methylation p value (-log10)
TFbs
10960000
chr16
11020000
F
AML RRBS
HLA II low
HLA II high
HLA II score
CIITA promoter cytosines
Methylation
CpGShore
CpGIsland
CpGShore
InterCGI
Expression Z-score
Genetic abnormalities
PML-RARA
CEBPA
CEBPA (2x)
IDH1
IDH2
DNMT3A
KMT2A-MLLT3
CBFB-MYH11
RUNX1-RUNX1T1
−5/7(q)
NPM1
NRAS
FLT3 (TKD)
FLT3-ITD
FAB subtype
M0
M1
M2
M3
M4
M5
M6
M7
HLA I score
FAB
G
BeatAML
HLA II score
M1/M2 NPM1
IDH1/2 TET2
CIITA expression
High
Low
NPM1
IDH1
IDH2
TET2
H
FIMM AML scRNA
HLA II score
CIITA
MPP
GMP
Monocyte
MEP
Monocyte
CMP/ GMP
HLAII/CIITA negative
I
scRNA HLA II score
FIMM scRNA
CIITA det. %
NPM1+ IDH1/TET2
6386
6333
6187
5897
5750
5249
5208
3867
J
HLA II score
ALL
CIITA methylation beta value (cg04945379)
HLA II score (log2)
t(12;21)
HeH
t(9;22)
t(1;19)
dic(9;20)
11q23/MLL
T-ALL
K
HLA II score
CCLE
CIITA methylation
HLA II score (log2 RPKM)
MOLM13
CLL
MCL
DLBCL
CHL
BL
BCL other
ALCL
pre-B-ALL
MM
TCL
AML
CML
T-ALL
L
MOLM13 methylation (RRBS)
Height
CIITA_IFN_1
Ctrl_DMSO_1
Ctrl_IFN_1
CIITA_DMSO_1
CIITA_IFN_2
Ctrl_DMSO_3
Ctrl_IFN_2
Ctrl_IFN_3
CIITA_DMSO_3
CIITA_IFN_3
CIITA_DMSO_2
Ctrl_DMSO_2
CIITA_DC_1
Ctrl_DC_3
CIITA_DCIFN_3
Ctrl_DCIFN_3
CIITA_DCIFN_1
Ctrl_DC_1
CIITA_DC_2
Ctrl_DC_2
Ctrl_DCIFN_1
Ctrl_DCIFN_2
CIITA_DC_3
CIITA_DCIFN_2
M
Higher in DMSO
Higher in treatment
MOLM13
MV411
Decitabine
IFNγ
Decitabine + IFNγ
−log10(FDR)
Fold change (log2)
N
Higher in sgControl
Higher in sgCIITA
MOLM13
MV411
DMSO
Decitabine
IFNγ
Decitabine + IFNγ
−log10(FDR)
Fold change (log2)
O
MONOMAC1
OCIAML2
MV411
HL60
MOLM13
OCIAML3
PL21
NB4
% HLA−DR+ cells
DMSO
DAC 100 nM
IFNγ 100 ng/ml
IFNγ 100 ng/ml + DAC 100 nM
P
CCLE AML
CIITA methylation
HLA II score (log2 RPKM)
MONOMAC1
OCIAML2
MOLM13
HL60
PL21
OCIAML3
NB4
CIITA methylation

Figure S4. Genetic and epigenetic alterations associated with HLA gene expression. Related to Figure 4.

(A) Heatmap of log2 average expression of genes significantly (FDR < 0.01 and FC > 2) differentially expressed between APC and non-APC normal cell types in Hemap. Rows and columns are clustered using Euclidean distance and complete linkage.

(B) Pairwise gene correlations (Spearman) for highly correlated genes from A in Hemap tumor samples are shown as a correlation matrix heatmap. HLA II genes included in the HLA II score based on their high correlation with each other are indicated in bold.

(C) Comparison of HLA II score and HLA II surface protein expression level in blasts in a validation cohort of AML BM samples (n = 37) profiled using both RNA-seq and flow cytometry for HLA-DR. The gray shading represents the 95% confidence interval.

(D) Percentages of highly methylated samples in transcriptomic subtypes of TCGA AML patients. Bolded (*CIITA* methylated) group consists of patients with *NPM1+IDH1/IDH2/TET2* mutation, methylated *CIITA*, and low HLA II expression corresponding to the sub-cluster highlighted in Figure 4E.

(E) Differentially methylated cytosines (DMCs) in the *CIITA* region between AML patient samples with low and high HLA II score in GSE86952 ERRBS dataset. Histogram indicates the negative log10 p value of differential methylation at each cytosine, with red and blue colors indicating hypermethylated and hypomethylated cytosines in HLA II low samples, respectively. CpG areas, including CpG islands, CpG shores (< 2 kb flanking CpG islands), and CpG shelves (< 2 kb flanking outwards from CpG shores) are shown above *CIITA* exons belonging to isoforms pIII (lymphoid) and pIV (IFNγ-inducible). Transcription factor binding sites (TFbs) are shown below.

(F) Heatmap showing methylation of cytosines at *CIITA* regions significantly hypermethylated in the HLA II low group compared to high in the AML GSE86952 ERRBS dataset. 0 indicates no methylation and 1 indicates complete methylation. Patients (columns) are grouped by HLA II score and PML-RARA status. Rows correspond to cytosines at the CpG island, shores, and inter-CGI area (> 4 kb outwards from a CpG island) shown on the right.

(G) HLA II score colored on the BeatAML t-SNE map. CIITA expression, and *NPM1* and *IDH1*, *IDH2,* and *TET2* mutation status is labeled for the cluster corresponding to that in Figure 4E (TCGA AML). The sub-cluster with lowest HLA II score and IDH1/IDH2/TET2 mutations is indicated by an arrow.

(H) UMAP plot of FIMM AML scRNA-seq samples (n=8) with HLA II score and *CIITA* expression colored (red tones correspond to high level). Myeloid cell types are labeled based on data-driven (SingleR) annotation. HSC, hematopoietic stem cell; MPP, multipotent progenitor, CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte–erythroid progenitor.

(I) Violin plots of HLA II score in blast clusters comparing the FIMM AML scRNA-seq samples. Percentage of cells in which *CIITA* expression is detected is indicated below as a heatmap and numeric value. Patients harboring *NPM1* and *IDH1/TET2* mutations are labeled in bold.

(J) HLA II score and *CIITA* methylation for pre-B-ALL and T-ALL subtypes is shown for Hemap GSE49031 dataset. Dot color represents methylation of the cg04945379 *CIITA* probe.

(K) HLA II score of CCLE hematological cell lines. Dot color represents mean *CIITA* CpG methylation. MOLM13 AML cell line used in the follow-up experiments (panels L-N) is highlighted (data from MV411 cell line not available in CCLE dataset).

(L) Cluster dendrogram of MOLM13 AML samples based on global methylation using Euclidean distance and Ward's linkage. Ctrl/CIITA = Control/*CIITA* sgRNA, DMSO = dimethyl sulfoxide, IFN = interferon gamma, DC = decitabine, 1-3 = replicate number.

(M) Volcano plots of MOLM13 and MV411 AML cells treated with decitabine (top), IFNγ (middle), and decitabine + IFNγ (bottom) compared to DMSO control. Significantly differentially expressed HLA II genes and *CIITA* are colored with red.

(N) Volcano plots of MOLM13 and MV411 AML cells expressing *CIITA* sgRNAs compared to control sgRNA-expressing cells. Comparisons under different treatments are shown. Significantly differentially expressed HLA II genes and *CIITA* are colored with red.

(O) Percentages of HLA-DR$^+$ cells of AML cell lines treated with DMSO, decitabine, IFNγ, or decitabine + IFNγ assessed by flow cytometry. P values obtained using Welch's *t*-test.
Scatter plot of HLA II score and *CIITA* methylation (mean within 1 kb of promoter) in CCLE AML cell lines. Cell lines used in the experiment shown in O are represented by larger dots and labeled.

Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 * IQR from the quartiles.

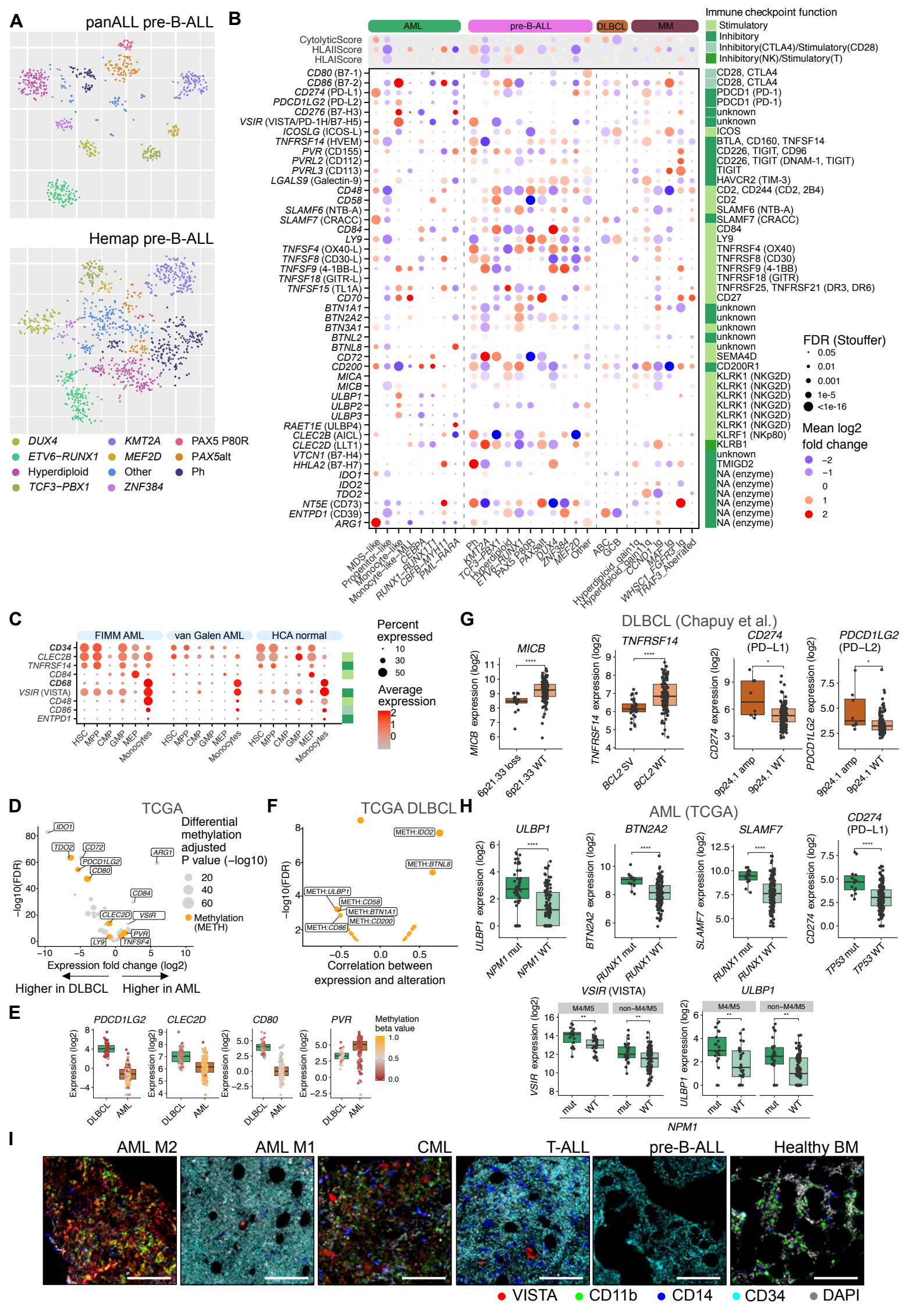

A
panALL pre-B-ALL
Hemap pre-B-ALL
DUX4
ETV6−RUNX1
Hyperdiploid
TCF3−PBX1
KMT2A
MEF2D
Other
ZNF384
PAX5 P80R
PAX5alt
Ph
B
AML
pre-B-ALL
DLBCL
MM
CytolyticScore
HLAIIScore
HLAIScore
CD80 (B7-1)
CD86 (B7-2)
CD274 (PD-L1)
PDCD1LG2 (PD-L2)
CD276 (B7-H3)
VSIR (VISTA/PD-1H/B7-H5)
ICOSLG (ICOS-L)
TNFRSF14 (HVEM)
PVR (CD155)
PVRL2 (CD112)
PVRL3 (CD113)
LGALS9 (Galectin-9)
CD48
CD58
SLAMF6 (NTB-A)
SLAMF7 (CRACC)
CD84
LY9
TNFSF4 (OX40-L)
TNFSF8 (CD30-L)
TNFSF9 (4-1BB-L)
TNFSF18 (GITR-L)
TNFSF15 (TL1A)
CD70
BTN1A1
BTN2A2
BTN3A1
BTNL2
BTNL8
CD72
CD200
MICA
MICB
ULBP1
ULBP2
ULBP3
RAET1E (ULBP4)
CLEC2B (AICL)
CLEC2D (LLT1)
VTCN1 (B7-H4)
HHLA2 (B7-H7)
IDO1
IDO2
TDO2
NT5E (CD73)
ENTPD1 (CD39)
ARG1
MDS−like
Progenitor−like
Monocyte−like
Monocyte−like_MLL
RUNX1−RUNX1T1
CBFB−MYH11
PML−RARA
Ph
KMT2A
TCF3−PBX1
Hyperdiploid
ETV6−RUNX1
PAX5 P80R
PAX5alt
DUX4
ZNF384
MEF2D
Other
ABC
GCB
Hyperdiploid_gain1q
Hyperdiploid_gain1q
CCND1 lg
MAF lg
WHSC1_FGFR3 lg
TRAF3_Aberrated
Immune checkpoint function
Stimulatory
Inhibitory
Inhibitory(CTLA4)/Stimulatory(CD28)
Inhibitory(NK)/Stimulatory(T)
CD28, CTLA4
CD28, CTLA4
PDCD1 (PD-1)
PDCD1 (PD-1)
unknown
unknown
ICOS
BTLA, CD160, TNFSF14
CD226, TIGIT, CD96
CD226, TIGIT (DNAM-1, TIGIT)
TIGIT
HAVCR2 (TIM-3)
CD2, CD244 (CD2, 2B4)
CD2
SLAMF6 (NTB-A)
SLAMF7 (CRACC)
CD84
LY9
TNFRSF4 (OX40)
TNFRSF8 (CD30)
TNFRSF9 (4-1BB)
TNFRSF18 (GITR)
TNFRSF25, TNFRSF21 (DR3, DR6)
CD27
unknown
unknown
unknown
unknown
unknown
SEMA4D
CD200R1
KLRK1 (NKG2D)
KLRK1 (NKG2D)
KLRK1 (NKG2D)
KLRK1 (NKG2D)
KLRK1 (NKG2D)
KLRK1 (NKG2D)
KLRF1 (NKp80)
KLRB1
unknown
TMIGD2
NA (enzyme)
NA (enzyme)
NA (enzyme)
NA (enzyme)
NA (enzyme)
NA (enzyme)
FDR (Stouffer)
0.05
0.01
0.001
1e-5
<1e-16
Mean log2 fold change
−2
−1
0
1
2
C
FIMM AML
van Galen AML
HCA normal
CD34
CLEC2B
TNFRSF14
CD84
CD68
VSIR (VISTA)
CD48
CD86
ENTPD1
HSC
MPP
CMP
GMP
MEP
Monocytes
Percent expressed
10
30
50
Average expression
2
1
0
G
DLBCL (Chapuy et al.)
MICB
TNFRSF14
CD274 (PD−L1)
PDCD1LG2 (PD−L2)
MICB expression (log2)
TNFRSF14 expression (log2)
CD274 expression (log2)
PDCD1LG2 expression (log2)
6p21.33 loss
6p21.33 WT
BCL2 SV
BCL2 WT
9p24.1 amp
9p24.1 WT
D
TCGA
Differential methylation adjusted P value (−log10)
20
40
60
Methylation (METH)
IDO1
TDO2
CD72
PDCD1LG2
CD80
ARG1
CD84
CLEC2D
VSIR
PVR
LY9
TNFSF4
−log10(FDR)
Expression fold change (log2)
Higher in DLBCL
Higher in AML
F
TCGA DLBCL
METH:IDO2
METH:BTNL8
METH:ULBP1
METH:CD58
METH:BTN1A1
METH:CD200
METH:CD86
−log10(FDR)
Correlation between expression and alteration
H
AML (TCGA)
ULBP1
BTN2A2
SLAMF7
CD274 (PD−L1)
ULBP1 expression (log2)
BTN2A2 expression (log2)
SLAMF7 expression (log2)
CD274 expression (log2)
NPM1 mut
NPM1 WT
RUNX1 mut
RUNX1 WT
TP53 mut
TP53 WT
VSIR (VISTA)
ULBP1
M4/M5
non−M4/M5
VSIR expression (log2)
ULBP1 expression (log2)
mut
WT
NPM1
E
PDCD1LG2
CLEC2D
CD80
PVR
Expression (log2)
Methylation beta value
1.0
0.5
0.0
DLBCL
AML
I
AML M2
AML M1
CML
T-ALL
pre-B-ALL
Healthy BM
VISTA
CD11b
CD14
CD34
DAPI

Figure S5. Genetic and epigenetic alterations and cancer subtypes associated with expression of immunomodulatory genes. Related to Figure 5.

(A) Pre-B-ALL subtypes colored on a t-SNE maps are shown for PanALL (Gu et al.) and Hemap.

(B) Dot plot of immunomodulatory genes in molecular subtypes of AML, pre-B-ALL, DLBCL, and MM. Dot color represents the mean log2 fold change (all cohorts per disease) between one subtype vs. other subtypes. Log2 fold change is limited to range -2 and 2. Dot size represents combined FDR values obtained using Stouffer's method for each comparison per disease. FDR values are discretized to five categories based on significance cutoffs (0.05, 0.01, 0.001, 1e-5, 1e-16).

(C) Dot plot of selected immunomodulatory genes in different myeloid and progenitor cell populations in AML (FIMM and van Galen et al. datasets) (van Galen et al., 2019) and normal bone marrow (HCA) based on scRNA-seq. Dot size represents the percentage of cells expressing the gene and color represents expression level. The lineage markers *CD34* (stem/progenitor cells) and *CD68* (monocytes) are shown in bold.

(D) Volcano plot of differentially expressed immunomodulatory genes between TCGA AML and DLBCL. Point size indicates negative log10 of p value of differential methylation of the gene between the two cancer types. Genes with methylation log2 fold change > 1.5 are colored in yellow.

(E) Comparison of *PDCD1LG2*, *CLEC2D, CD80,* and *PVR* gene expression levels between TCGA DLBCL and TCGA AML samples. Dots indicating individual patients are colored by average methylation within 1 kb of the transcription start site.

(F) Volcano plot of correlations (Spearman) of immunomodulatory gene expression with promoter methylation in TCGA DLBCL. Point size is proportional to the adjusted p value.

(G) Selected immunomodulatory genes whose expression is correlated with genetic alterations in DLBCL (GSE98588) are shown as boxplots comparing cases with alteration (mut/SV/amp) and wild-type (WT) cases. P values obtained using two-sided Wilcoxon rank sum test are shown.

(H) Selected immunomodulatory genes whose expression is correlated with genetic alterations in TCGA AML are shown as boxplots comparing cases with alteration (mut) and wild-type (WT) cases. P values obtained using two-sided Wilcoxon rank sum test are shown. For *C10orf54 (VISTA)* and *ULBP1*, expression stratified both by *NPM1* mutation and M4/M5 (myelomonocytic/monocytic) status is shown.

(I) Multiplex immunohistochemistry of BM biopsies for VISTA, CD11b, CD14, CD34 and DAPI from representative patients of each profiled cancer type and healthy controls. Scale bars, 100 μm.

Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 * IQR from the quartiles.

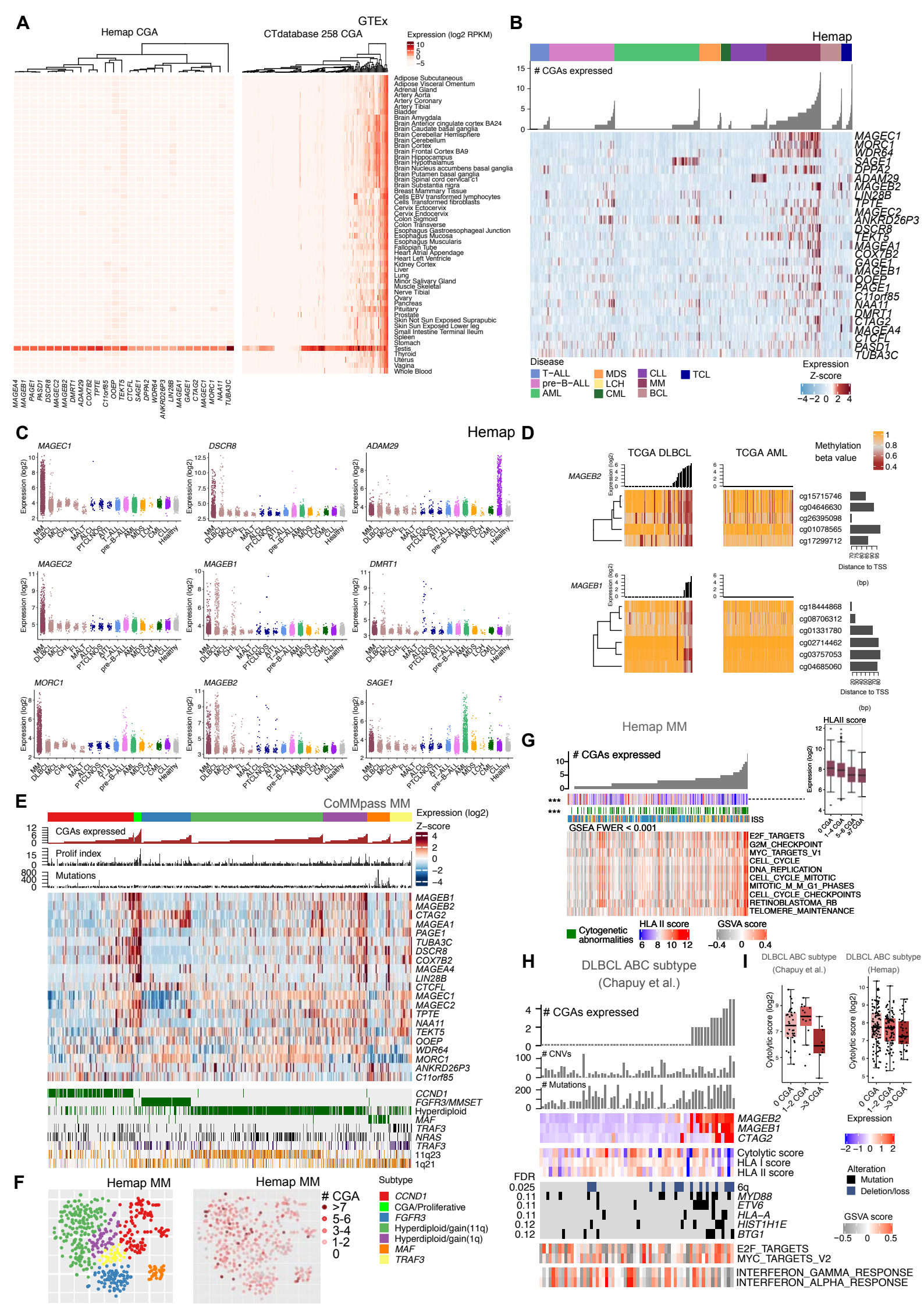

A
Hemap CGA
GTEx
CTdatabase 258 CGA
Expression (log2 RPKM)
10
5
0
−5
Adipose Subcutaneous
Adipose Visceral Omentum
Adrenal Gland
Artery Aorta
Artery Coronary
Artery Tibial
Bladder
Brain Amygdala
Brain Anterior cingulate cortex BA24
Brain Caudate basal ganglia
Brain Cerebellar Hemisphere
Brain Cerebellum
Brain Cortex
Brain Frontal Cortex BA9
Brain Hippocampus
Brain Hypothalamus
Brain Nucleus accumbens basal ganglia
Brain Putamen basal ganglia
Brain Spinal cord cervical c1
Brain Substantia nigra
Breast Mammary Tissue
Cells EBV transformed lymphocytes
Cells Transformed fibroblasts
Cervix Ectocervix
Cervix Endocervix
Colon Sigmoid
Colon Transverse
Esophagus Gastroesophageal Junction
Esophagus Mucosa
Esophagus Muscularis
Fallopian Tube
Heart Atrial Appendage
Heart Left Ventricle
Kidney Cortex
Liver
Lung
Minor Salivary Gland
Muscle Skeletal
Nerve Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin Not Sun Exposed Suprapubic
Skin Sun Exposed Lower leg
Small Intestine Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood
MAGEA4
MAGEB1
PAGE1
PASD1
DSCR8
MAGEC2
MAGEB2
DMRT1
ADAM29
COX7B2
TPTE
C11orf85
OOEP
TEKT5
CTCFL
SAGE1
DPPA2
WDR64
ANKRD26P3
LIN28B
MAGEA1
GAGE1
CTAG2
MAGEC1
MORC1
NAA11
TUBA3C
B
Hemap
CGAs expressed
15
10
5
0
MAGEC1
MORC1
WDR64
SAGE1
DPPA2
ADAM29
MAGEB2
LIN28B
TPTE
MAGEC2
ANKRD26P3
DSCR8
TEKT5
MAGEA1
COX7B2
GAGE1
MAGEB1
OOEP
PAGE1
C11orf85
NAA11
DMRT1
CTAG2
MAGEA4
CTCFL
PASD1
TUBA3C
Disease
T−ALL
pre−B−ALL
AML
MDS
LCH
CML
CLL
MM
BCL
TCL
Expression
Z-score
−4 −2 0 2 4
C
Hemap
MAGEC1
DSCR8
ADAM29
MAGEC2
MAGEB1
DMRT1
MORC1
MAGEB2
SAGE1
Expression (log2)
MM
DLBCL
MCL
CHL
FL
MALT
ALCL
PTCLNOS
AITL
T-ALL
pre−B−ALL
AML
MDS
LCH
CML
CLL
Healthy
D
TCGA DLBCL
TCGA AML
Methylation
beta value
1
0.8
0.6
0.4
MAGEB2
cg15715746
cg04646630
cg26395098
cg01078565
cg17299712
Distance to TSS
(bp)
MAGEB1
cg18444868
cg08706312
cg01331780
cg02714462
cg03757053
cg04685060
Distance to TSS
(bp)
E
CoMMpass MM
Expression (log2)
Z−score
4
2
0
−2
−4
CGAs expressed
Prolif index
Mutations
MAGEB1
MAGEB2
CTAG2
MAGEA1
PAGE1
TUBA3C
DSCR8
COX7B2
MAGEA4
LIN28B
CTCFL
MAGEC1
MAGEC2
TPTE
NAA11
TEKT5
OOEP
WDR64
MORC1
ANKRD26P3
C11orf85
CCND1
FGFR3/MMSET
Hyperdiploid
MAF
TRAF3
NRAS
TRAF3
11q23
1q21
F
Hemap MM
Hemap MM
CGA
>7
5-6
3-4
1-2
0
Subtype
CCND1
CGA/Proliferative
FGFR3
Hyperdiploid/gain(11q)
Hyperdiploid/gain(1q)
MAF
TRAF3
G
Hemap MM
CGAs expressed
ISS
HLAII score
Expression (log2)
0 CGA
1-2 CGA
3-4 CGA
≥5 CGA
GSEA FWER < 0.001
E2F_TARGETS
G2M_CHECKPOINT
MYC_TARGETS_V1
CELL_CYCLE
DNA_REPLICATION
CELL_CYCLE_MITOTIC
MITOTIC_M_M_G1_PHASES
CELL_CYCLE_CHECKPOINTS
RETINOBLASTOMA_RB
TELOMERE_MAINTENANCE
Cytogenetic abnormalities
HLA II score
6 8 10 12
GSVA score
−0.4 0 0.4
H
DLBCL ABC subtype
(Chapuy et al.)
CGAs expressed
CNVs
Mutations
MAGEB2
MAGEB1
CTAG2
Cytolytic score
HLA I score
HLA II score
FDR
0.025
0.11
0.11
0.11
0.12
0.12
6q
MYD88
ETV6
HLA−A
HIST1H1E
BTG1
E2F_TARGETS
MYC_TARGETS_V2
INTERFERON_GAMMA_RESPONSE
INTERFERON_ALPHA_RESPONSE
I
DLBCL ABC subtype
(Chapuy et al.)
DLBCL ABC subtype
(Hemap)
Cytolytic score (log2)
0 CGA
1-2 CGA
>3 CGA
Expression
−2 −1 0 1 2
Alteration
Mutation
Deletion/loss
GSVA score
−0.5 0 0.5

Figure S6. CGA expression in hematological malignancies. Related to Figure 6.

(A) GTEx dataset log2 RPKM expression of 27 CGAs found expressed in hematological malignancies are shown as a heatmap. On the right expression of CTdatabase CGA genes (total 258) are shown.
(B) Expression of CGAs across hematological malignancies in Hemap are shown as a heatmap. Total number of expressed CGAs per patient are shown as a barplot.
(C) Log2 expression for selected CGAs for Hemap cancers and healthy controls.
(D) *MAGEB2* and *MAGEB1* methylation values are shown for each probe associated with the gene for TCGA DLBCL and AML data sets. Samples are ranked by gene expression indicated above the heatmap. Probe distances (bp) from TSS are shown as a barplot on the right.
(E) Expression of CGAs in MM (CoMMpass) are shown as a heatmap. Samples are grouped based on molecular subtype identified by clustering and ordered by total number of expressed CGAs per patient shown as a barplot. Genetic alterations enriched in the molecular subtypes are shown below as an oncoprint.
(F) Identified MM subtypes and number of expressed CGAs are shown for Hemap MM as a t-SNE map.
(G) GSVA scores of pathway gene sets and clinical features correlated with the number of expressed CGAs in Hemap MM shown as an oncoprint.
(H) Genetic alterations and gene sets correlated with the number of expressed CGAs in the ABC subtype of DLBCL (Chapuy et al., 2018). Mutation and CNV load, immunological scores, and expression of 3 most commonly expressed CGAs in DLBCL are shown.
(I) Cytolytic score stratified by CGA number in ABC DLBCL (Chapuy et al. and Hemap).

Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the min and max values at most 1.5 * IQR from the quartiles.

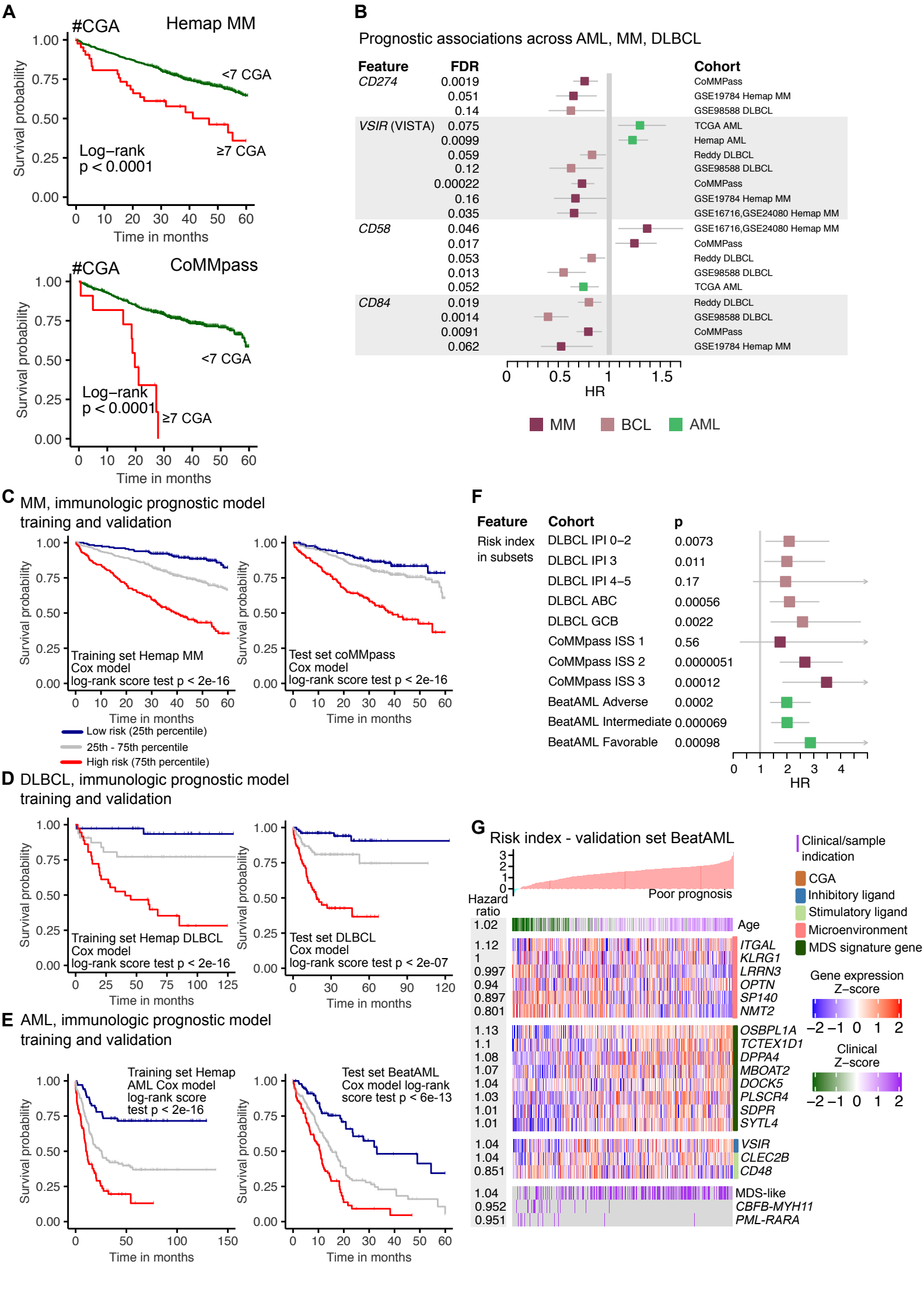

A
#CGA
Hemap MM
Survival probability
1.00
0.75
0.50
0.25
0.00
<7 CGA
≥7 CGA
Log−rank
p < 0.0001
0 10 20 30 40 50 60
Time in months
#CGA
CoMMpass
<7 CGA
≥7 CGA
Log−rank
p < 0.0001
Time in months
B
Prognostic associations across AML, MM, DLBCL
Feature
FDR
Cohort
CD274 0.0019 CoMMPass
0.051 GSE19784 Hemap MM
0.14 GSE98588 DLBCL
VSIR (VISTA) 0.075 TCGA AML
0.0099 Hemap AML
0.059 Reddy DLBCL
0.12 GSE98588 DLBCL
0.00022 CoMMPass
0.16 GSE19784 Hemap MM
0.035 GSE16716,GSE24080 Hemap MM
CD58 0.046 GSE16716,GSE24080 Hemap MM
0.017 CoMMPass
0.053 Reddy DLBCL
0.013 GSE98588 DLBCL
0.052 TCGA AML
CD84 0.019 Reddy DLBCL
0.0014 GSE98588 DLBCL
0.0091 CoMMPass
0.062 GSE19784 Hemap MM
0 0.5 1 1.5
HR
MM
BCL
AML
C
MM, immunologic prognostic model training and validation
Survival probability
Training set Hemap MM
Cox model
log-rank score test p < 2e-16
Test set coMMpass
Cox model
log-rank score test p < 2e-16
Time in months
Low risk (25th percentile)
25th - 75th percentile
High risk (75th percentile)
D
DLBCL, immunologic prognostic model training and validation
Training set Hemap DLBCL
Cox model
log-rank score test p < 2e-16
Test set DLBCL
Cox model
log-rank score test p < 2e-07
E
AML, immunologic prognostic model training and validation
Training set Hemap AML Cox model log-rank score test p < 2e-16
Test set BeatAML Cox model log-rank score test p < 6e-13
F
Feature
Cohort
p
Risk index in subsets
DLBCL IPI 0–2 0.0073
DLBCL IPI 3 0.011
DLBCL IPI 4–5 0.17
DLBCL ABC 0.00056
DLBCL GCB 0.0022
CoMMpass ISS 1 0.56
CoMMpass ISS 2 0.0000051
CoMMpass ISS 3 0.00012
BeatAML Adverse 0.0002
BeatAML Intermediate 0.000069
BeatAML Favorable 0.00098
0 1 2 3 4
HR
G
Risk index - validation set BeatAML
Hazard ratio
Poor prognosis
Clinical/sample indication
CGA
Inhibitory ligand
Stimulatory ligand
Microenvironment
MDS signature gene
Gene expression Z−score
Clinical Z−score
Age
ITGAL
KLRG1
LRRN3
OPTN
SP140
NMT2
OSBPL1A
TCTEX1D1
DPPA4
MBOAT2
DOCK5
PLSCR4
SDPR
SYTL4
VSIR
CLEC2B
CD48
MDS-like
CBFB-MYH11
PML-RARA

Figure S7. Survival associations of immunological features. Related to Figure 7.

(A) Kaplan-Meier curves of samples with frequent CGA expression (samples with ≥7CGA expressed) in Hemap and CoMMpass MM cohorts (related to Figure 7A). P value was computed using the log-rank test.

(B) Forest plot of univariate Cox proportional hazards model coefficients and confidence intervals and FDR values for AML, MM, and DLBCL cohorts with significant survival association for selected features (FDR<0.2 in at least two cohorts and two distinct diseases). Full list of univariate results is shown in Table S7.

(C) Kaplan-Meier curves of overall survival for MM patients stratified by the risk index of the immunological risk model in the training set (Hemap MM) and in the validation set (CoMMpass MM). Patients were divided into three groups based on 75th and 25th percentiles to divide patients into high, intermediate, and low risk groups and p value was obtained using the score (log-rank) test.

(D) As in C for the DLBCL training set (Hemap R-CHOP treated GSE10846,GSE17372) and validation set (GSE98588).

(E) As in C for the AML training set (Hemap AML) and validation set (BeatAML).

(F) Forest plot of univariate Cox proportional hazard model coefficients and p values depicting the immunological risk model performance in subgroups of the survival cohorts, stratified by existing prognostic risk groups, including the IPI categories 0-5 and ABC/GCB subtypes for DLBCL, ISS 1-3 risk groups for MM, and ELN2017 AML risk groups containing adverse (contains also the adverse-intermediate group), intermediate, and favorable (contains also the intermediate-favorable groups) risk groups. Most prognostic risk groups could be further stratified by the immunological risk model.

(G) Heatmap depicting features included in the AML immunological risk model in the validation set (BeatAML). Patients are ordered by the risk index and hazard ratio (HR) for each feature is shown on the left. Features are grouped based on predefined categories and features are ordered within category by the HR.