



**HAL**  
open science

## Leukemia relapse via genetic immune escape after allogeneic hematopoietic cell transplantation

Simona Pagliuca, Carmelo Gurnari, Colin Hercus, Sébastien Hergalant, Sanghee Hong, Adele Dhuyser, Maud D'aveni, Alice Aarnink, Marie Thérèse Rubio, Pierre Feugier, et al.

### ► To cite this version:

Simona Pagliuca, Carmelo Gurnari, Colin Hercus, Sébastien Hergalant, Sanghee Hong, et al.. Leukemia relapse via genetic immune escape after allogeneic hematopoietic cell transplantation. Nature Communications, 2023, 14 (1), pp.3153. 10.1038/s41467-023-38113-4 . hal-04162108

**HAL Id: hal-04162108**

**<https://hal.univ-lorraine.fr/hal-04162108>**

Submitted on 14 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Leukemia relapse via genetic immune escape after allogeneic hematopoietic cell transplantation

Received: 9 October 2022

Accepted: 13 April 2023

Published online: 31 May 2023

 Check for updates

Simona Pagliuca <sup>1,2,3,13</sup>, Carmelo Gurnari <sup>1,4,13</sup>, Colin Hercus<sup>5</sup>, Sébastien Hergalant <sup>6</sup>, Sanghee Hong<sup>7</sup>, Adele Dhuyser<sup>3,8</sup>, Maud D'Aveni<sup>2,3</sup>, Alice Aarnink<sup>3,8</sup>, Marie Thérèse Rubio<sup>2,3</sup>, Pierre Feugier<sup>2</sup>, Francesca Ferraro <sup>9</sup>, Hetty E. Carraway <sup>10</sup>, Ronald Sobecks<sup>11</sup>, Betty K. Hamilton <sup>11</sup>, Navneet S. Majhail<sup>12</sup>, Valeria Visconte <sup>1</sup> & Jaroslaw P. Maciejewski <sup>1</sup> 


Graft-versus-leukemia (GvL) reactions are responsible for the effectiveness of allogeneic hematopoietic cell transplantation as a treatment modality for myeloid neoplasia, whereby donor T- effector cells recognize leukemia neoantigens. However, a substantial fraction of patients experiences relapses because of the failure of the immunological responses to control leukemic outgrowth. Here, through a broad immunogenetic study, we demonstrate that germline and somatic reduction of human leucocyte antigen (HLA) heterogeneity enhances the risk of leukemic recurrence. We show that preexistent germline-encoded low evolutionary divergence of class II HLA genotypes constitutes an independent factor associated with disease relapse and that acquisition of clonal somatic defects in HLA alleles may lead to escape from GvL control. Both class I and II HLA genes are targeted by somatic mutations as clonal selection factors potentially impairing cellular immune responses and response to immunomodulatory strategies. These findings define key molecular modes of post-transplant leukemia escape contributing to relapse.

Hematopoietic cell transplantation (allo-HCT) holds a pivotal place in the therapeutic algorithm of myeloid neoplasms (MN)<sup>1,2</sup>. Either as an upfront treatment in diseases with low blast burden, after disease control in myelodysplastic syndromes (MDS) or following consolidation therapy in intermediate and high-risk acute myeloid leukemia (AML), its therapeutic index is principally related to the immunogenic

graft-versus-leukemia (GvL) effect, exerted by donor-derived immune effectors on leukemic cells<sup>3</sup>. Relapse following allo-HCT may be mechanistically multifactorial and includes re-expansion of the residual leukemia with or without acquisition of novel driver lesions along with the development of various immune escape mechanisms contributing to the evasion from GvL effect<sup>4,5</sup>.

<sup>1</sup>Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>2</sup>Department of Hematology, CHRU de Nancy, Vandœuvre-lès-Nancy, France. <sup>3</sup>CNRS UMR 7365, IMoPA, Biopole of University of Lorraine, Vandœuvre-lès-Nancy, France.

<sup>4</sup>Department of Biomedicine and Prevention, PhD in Immunology, Molecular Medicine and Applied Biotechnology, University of Rome Tor Vergata, Rome, Italy. <sup>5</sup>Novocraft Technologies Sdn Bhd, Kuala Lumpur, Malaysia. <sup>6</sup>Inserm UMR-S 1256 Nutrition-Genetics-Environmental Risk Exposure, University of Lorraine, 54500 Vandœuvre-lès-Nancy, France. <sup>7</sup>Division of Hematologic Malignancies and Cellular Therapy, Department of Medicine, Duke University School of Medicine, Durham, NC, USA. <sup>8</sup>Histocompatibility Department, CHRU de Nancy, Vandœuvre-lès-Nancy, France. <sup>9</sup>Division of Oncology, Department of Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO, USA. <sup>10</sup>Leukemia Program, Hematology Department, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>11</sup>Blood and Marrow Transplant Program, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>12</sup>Sarah Cannon Transplant and Cellular Therapy Network, Nashville, TN, USA. <sup>13</sup>These authors contributed equally: Simona Pagliuca, Carmelo Gurnari.

 e-mail: [maciejj@ccf.org](mailto:maciejj@ccf.org)

Post-transplant alloreactivity ideally acts via class I and II human leukocyte antigen (HLA)-restricted mechanisms. Therefore, one of the biological modes of relapse after allo-HCT is the loss of expression of disparate immunodominant HLA alleles via loss of heterozygosity (LOH) or epigenetic down-regulation, which have been demonstrated in haploidentical, mismatched unrelated and matched related settings<sup>6–8</sup>. While large genomic aberrations including the loss or the duplication of an entire haplotype or of a single allele have been noted across various studies<sup>6,9–11</sup>, more granular lesions of the HLA region including microdeletions and point mutations, remain underestimated in allo-HCT. Such lesions, ultimately resulting in phenotypes of immune resistance, have been reported in the context of clonal adaptive recovery in immune-mediated aplastic anemia (AA) or as immune escape drivers in solid tumors, together with up-modulation of negative immune checkpoint regulators<sup>12–14</sup>.

HLA heterogeneity and the type of HLA polymorphisms in certain loci have also been found as factors influencing the direction of alloreactive responses<sup>15–19</sup>. HLA evolutionary divergence (HED), a metric reflecting the immunopeptidome diversity, has been recently linked to post-immunotherapy outcomes in solid tumors<sup>20</sup>. Furthermore, we and others demonstrated that recipient HED correlates with the efficiency of immune reconstitution after allo-HCT<sup>21,22</sup>. Since this score mirrors the antigenic spectrum capacity, it may be considered a surrogate of the individuals' ability to present leukemia associated antigens (LAA) in transplant recipients<sup>20,23,24</sup>.

In analogy with solid cancer immunotherapy, we hypothesized that, also in the context of allo-HCT, a reduced variability of HLA genotypes as well as genomic mechanisms of escape can converge in overlapping immune-escape phenotypes.

Here we investigate relapsed leukemia in allo-HCT recipients for the presence of non-driving genomic lesions in immune elements as part of the mutational landscape of recurrent leukemia. To unravel the genomic dysfunctions leading to immune-escape, we study HED and both genetic and transcriptomic inactivation of HLA via somatic mutations or down-modulation as determinants of leukemic relapse in the setting of allo-HCT.

We demonstrate how both germline and somatic dysfunction of HLA heterogeneity provides an immunogenetic framework prone to post-transplant leukemia immune escape and relapse.

## Results

### Dysfunction of germline HLA heterogeneity as immunogenetic driver of post allo-HCT relapse

We started from the concept that a more structurally diverse HLA allele combination could widen the spectrum of antigen presentation, thereby enhancing the GvL effect. We thus analyzed a cohort of 494 patients (median follow up 27.5 [14–45] months) who received allo-HCT for MN and had available high-resolution HLA genotype (Table 1; Fig. 1; Data Source 1). For each patient we calculated per-locus, per-class and global HED scores. This parameter quantitates the physiochemical divergence existing between two amino acids within the peptide binding cleft of two homologous HLA alleles. To investigate the impact of HED distribution on clinical outcomes, we categorized each value using a cutoff that could be standardized and clinically meaningful. Since specific polymorphisms in different HLA loci have been previously associated with AML susceptibility or relapse<sup>25–27</sup> thereby potentially biasing global HED scores, we did not base our study on cutoff values computed on this patient population. Instead, for the assessment of this metric we used cutoffs defined in healthy controls (HC) (Supplementary Fig. 1; Supplementary Data 1) and considered as a benchmark the 50th percentile of these HC HED scores (locus-specific and global), as we previously described<sup>28</sup>.

In the matched allo-HCT context ( $N = 393$ ), we found that patients with high class II HED had a lower relapse rate as compared to those with low class II HED values (hazard ratio [HR]: 0.65, 95% CI: 0.45–0.958,

**Table 1 | Patient and transplant characteristics of the whole cohort ( $N = 494$ )**

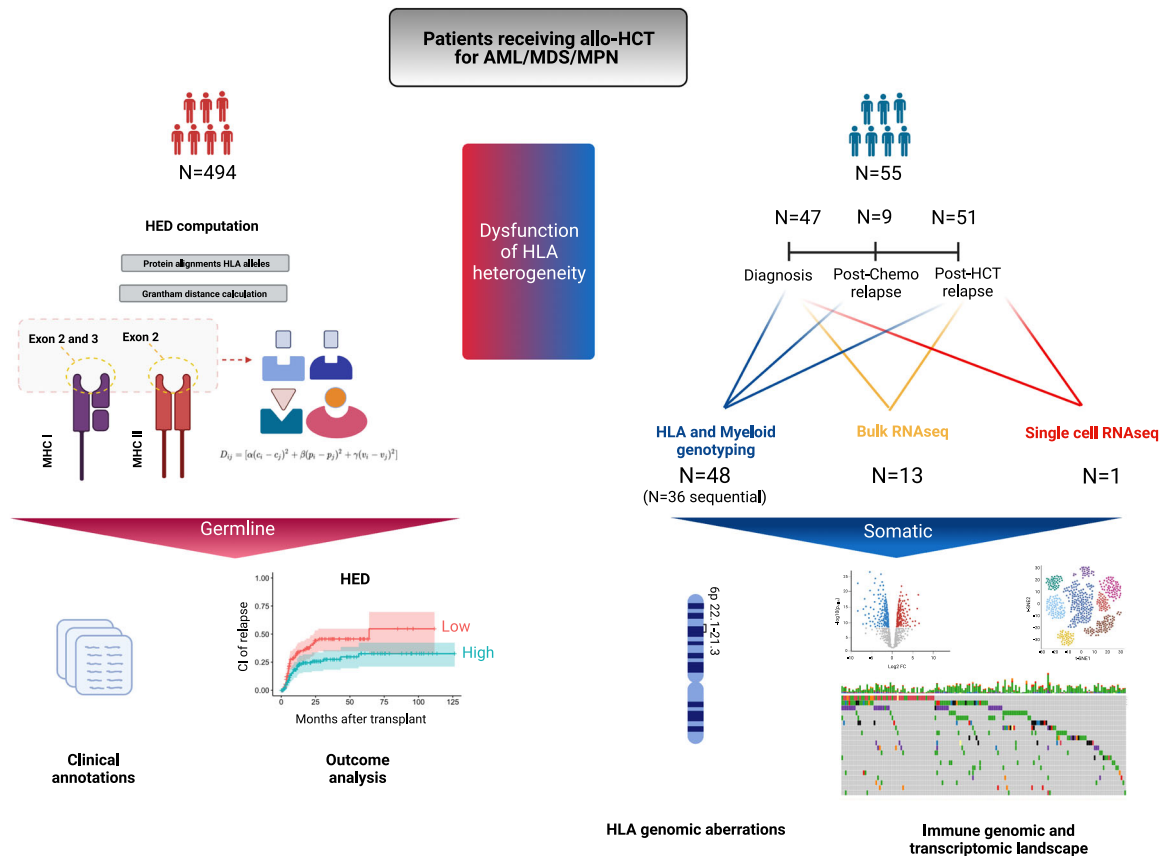
	N (% or IQR)
Whole cohort	$N = 494$
Median Age at HCT (years)	59 (52–65)
Disease	
AML	294 (59)
MDS	125 (26)
MPN	75 (15)
Donor type	
MRD	130 (26)
MUD	261 (53)
MMUD	6 (<1)
Haplo	97 (20)
Graft type	
BM	153(31)
PBSC	341 (69)
Conditioning regimen	
MAC	176 (36)
RIC	318 (64)
Post-transplant relapse	
Yes	142 (29)
No	352 (71)
Median time to relapse (days)	173 (100–319)
Median follow-up (months)	27.5 (14–45.3)

IQR interquartile range, HCT hematopoietic cell transplantation, AML acute myeloid leukemia, MDS myelodysplastic syndrome, MPN myeloproliferative neoplasms, MRD matched related donor, MUD matched unrelated donor, MMUD mismatched unrelated donor, Haplo haploidentical donor, BM bone marrow, PBSC peripheral blood stem cells, MAC myeloablative, RIC reduced intensity.

$p = 0.021$ ). A breakdown of the specific contribution of the class II loci, showed that HLA-DQB1 and HLA-DPB1 impacted the most on this result ( $p = 0.04$ , Fig. 2a, b). Among class I loci, the only impact was seen for HLA-C, for which high HED values were associated with lower cumulative incidence (CI) of relapse ( $p = 0.041$ ). Class II, but not class I HED, affected the probability of relapse also when considered in cause-specific multivariable regression models (HR: 0.54, 95% CI: 0.36–0.81,  $p = 0.003$ ), adjusted for age, type of donor, conditioning regimen, disease type, stem cell source, HCT-CI score, disease risk and year of transplant (Fig. 2c, d). In a similar fashion, high class II HED values were associated with better overall survival (OS HR: 0.63, 95% CI: 0.46–0.86,  $p = 0.004$ ), whereas no correlation was observed with acute and chronic GvHD (Supplementary Fig. 2).

For the haploidentical and mismatched unrelated population ( $N = 100$ ), neither class I, nor class II or global HED were found to affect disease recurrence (Supplementary Fig. 3), implying that a modulation of the immunogenetic risk by other transplant-related factors may influence the GvL effect.

To link immune control dysfunction with abnormal HLA heterogeneity, we performed deep targeted next generation sequencing (NGS) for T cell receptor (TCR) V $\beta$  complementarity determining region 3 (CDR3) after 3–6 months following allo-HCT ( $N = 24$ ) and characterized the specificity of TCR repertoire using previously described bioinformatic frameworks<sup>21</sup>. Overall, patients with higher class I and II HED were characterized by a less expanded repertoire together with increased cancer-related clonotypes and less clonally-polarized features compared to those with lower divergence. This phenomenon was also recorded for pathogen-recognizing clonotypes, underlying a general imprinting for a more diverse and variegated antigenic response (Supplementary Fig. 4). We asked whether the differences in repertoire expansion could be related to a higher frequency



**Fig. 1 | Study design.** In total 494 patients receiving a first allogeneic hematopoietic cell transplantation (HCT) for a myeloid disorder were included in this study. For the estimation of the dysfunction of the germline HLA diversity, we calculated the HLA evolutionary divergence, a metric able to quantitate the structural diversity at the peptide binding site of each homologous HLA molecule (LEFT panel). This metric was correlated in univariate and multivariate analysis with the cumulative incidence of relapse and other clinical outcomes. To characterize the somatic

dysfunction of HLA diversity (RIGHT panel) we performed a longitudinal immunogenomic and transcriptomic analysis in a subgroup of patients relapsing after HCT ( $N=55$ ). Classical HLA loci and myeloid genes were genotyped, and their mutational status was analyzed. In addition, bulk and single cell RNA sequencing were also performed. Of note is that 7 samples used for the bulk RNA sequencing were previously published (Christopher et al. NEJM 2018).

of CMV reactivation in relation to HED values; but no difference in terms of CMV infection was observed between patients with low vs high class I and II HED ( $p=0.07$  for class I and  $p=0.691$  for class II, Supplementary Data 2). Longitudinal clonotype tracking analysis showed that some of these clones were present in pre-transplant or donor samples predating post-transplant hyper-expansion (Supplementary Fig. 5).

### HLA and myeloid landscape

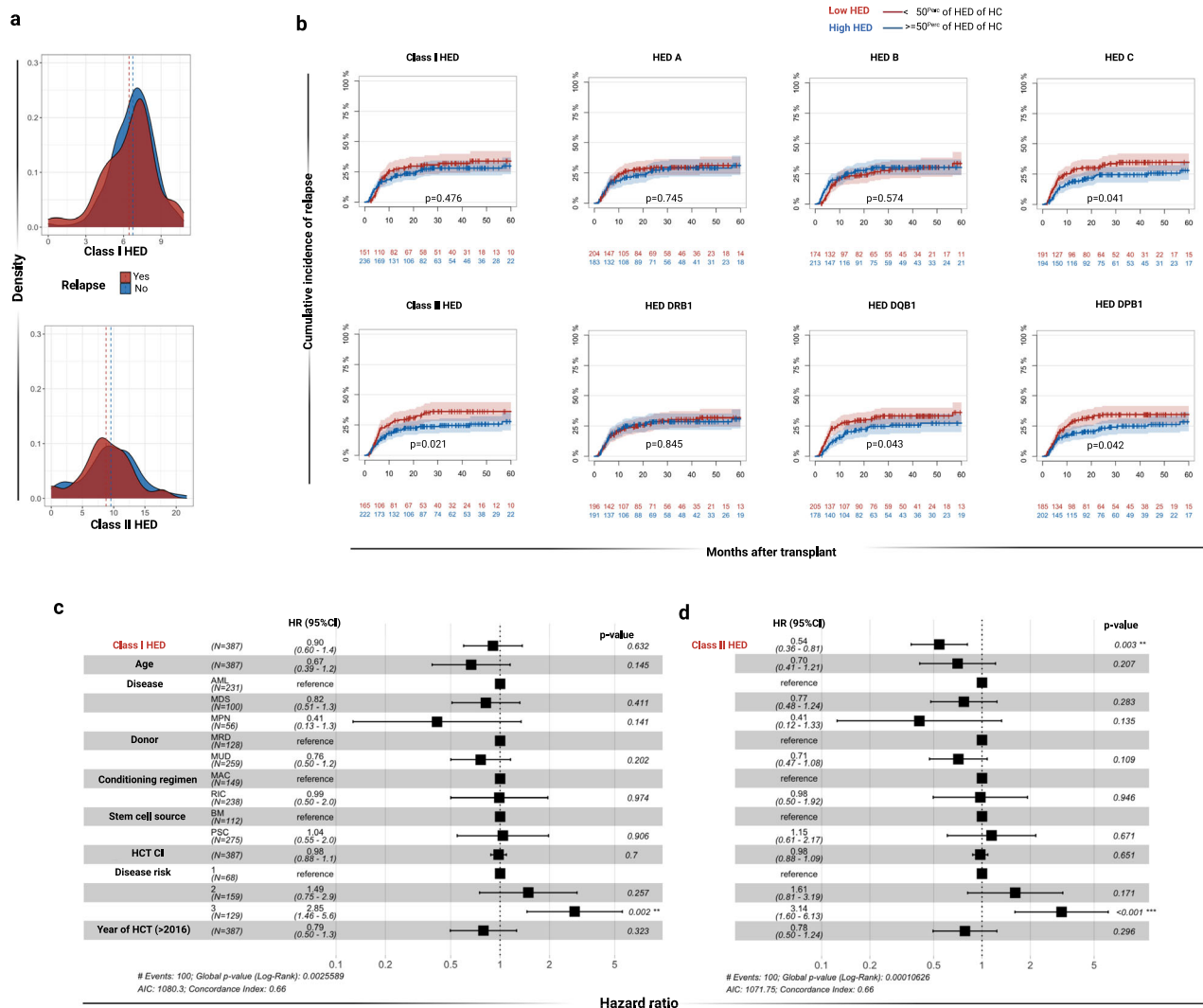
We then assessed the distribution and the frequency of HLA mutations in a subset of longitudinally genotyped patients relapsing after allo-HCT ( $N=48$ ). Specimens at diagnosis ( $N=40$ ), at post-chemotherapy ( $N=9$ ) and post-allo-HCT ( $N=44$ ) relapse were sequenced for both HLA mutations and myeloid genes (Source Data 1, Table 2). We used a newly implemented pipeline designed to recognize somatic mutations and allelic losses within HLA loci<sup>29</sup>. Whereas somatic HLA hits were present in 22% ( $N=9/40$ ) of patients at diagnosis and in 38% ( $N=17/44$ ) of patients at post allo-HCT relapse, in cases with relapse following chemotherapy no HLA hits were found (Fig. 3a, Source Data 2). In the 2 cases harboring HLA aberrations at diagnosis, no hits were found at relapse after chemotherapy ( $N=2$ ; Fig. 3a).

HLA lesions were classified as missense, non-sense, frameshift, splicing or losses and were distributed across class I and II alleles (Source Data 2). Most affected loci were A, C, and DQB1 with allelic losses occurring predominantly in DRB1 and DQB1 (Fig. 3b). Mutations and losses were seen irrespectively of donor type and transplant setting ( $N=10/20$ , 50% in matched related;  $N=2/8$ , 25% in haploidentical;

$N=5/17$ , 29% in unrelated transplants, Fig. 3c). Of note and in line with previous results, HLA hits occurred in the mismatched haplotype in the haploidentical setting. With a median time to post-transplant disease recurrence of 5.8 months (1.67–56.3), HLA aberrations were more frequent in patients with late (>6 months) relapses (42% vs 29% in early relapses; Fig. 3d, e).

When we analyzed the frequency and distribution of most common somatic myeloid driver hits (for panel description see Supplementary Data 3) in relationship to HLA lesions, no difference in terms of myeloid landscape in samples *prior* to transplant *vs.* post-transplant relapses were noticed. Therefore, HLA aberrations accounted for the predominant genomic lesions in post-transplant relapses. However, reverse analysis indicated that relapses with newly acquired HLA lesions were enriched in *RUNX1*, *DNMT3A*, *EZH2*, *EP300* mutations (Fig. 3g) as compared to those with wild type HLA. Of note is that the acquisition of HLA mutations was not related to HED. In fact, no differences were found with regard to the global, class I, II and locus specific divergence between HLA wild type and mutant sub-cohorts, suggesting that HED was not indicative of the immune pressure or propensity for immune escape in this setting. Similarly, mutated HLA loci did not show higher HED as compared to respective loci in healthy controls (Supplementary Fig. 6).

Notably, all patients sequenced at relapse showed a prevalent recipient chimerism, along with the fact that specific HLA events were not found in recipient prior to transplant. These observations indicated that HLA mutations were somatic and occurred at relapse.



**Fig. 2 | Impact of HED on cumulative incidence of relapse. a** Density distribution of class I and class II HED scores in post-HCT relapsed (red) and non-relapsed patients (blue). **b** Impact of per-class and per-locus HED scores on cumulative incidence (CI) of relapse. Low (red) and high (blue) HED are defined according to the 50th percentile of the corresponding locus in healthy controls. Shaded bands represent 95% confident interval. Non-adjusted *p*-values indicate the significance of the log-rank test. **c** Cause specific multivariable cox regression analysis of relapse including class I HED as main effect variable. **d** Cause specific multivariable cox

regression analysis of relapse including class II HED as main effect variable. In (c) and (d) disease risk was intended as: 1 = low; 2 = intermediate; 3 = high. Black squares indicate the odd ratio and error bars the 95% confident intervals. All the *p*-values were two sided. Source data 1. Abbreviations: HCT-CI hematopoietic cell transplant comorbidity index, AML acute myeloid leukemia, MDS myelodysplastic syndromes, MPN myeloproliferative neoplasms, MRD matched related donor, MUD matched unrelated donor, MAC myeloablative conditioning regimen, RIC reduced intensity conditioning regimen, HR Hazard ratio 95% CI: 95% confident interval.

Donor lymphocyte infusion (DLI) for the treatment of relapse, was performed in half of the patients of this subgroup (18 AML and 7 MDS) at a median time of 7.8 (range 2.9–61.08) months after transplant. DLI were administered mostly in matched related context (14, 56%). A complete response was seen in 5 patients (20%). Nine patients of this subgroup harbored HLA aberrations both in class I and II alleles spanning across exonic, intronic or untranslated (UTR) regions. Of them 6 with exonic mutations or allelic losses did not achieve response to DLI, while 3 patients harboring UTR mutations in HLA-DPA1, -DPB1 and -A were still alive at >1 year post-relapse with controlled disease after DLI and other immunochemotherapy-based strategies (including 5-azacytidine and sorafenib).

### HLA and non-HLA immune transcriptional dysregulation in relapses following allo-HCT

In addition to somatic HLA hits, we also investigated the transcriptional changes of HLA and immune genes occurring from diagnosis to post allo-HCT relapses in a subgroup of 13 relapsed patients using deep

RNA sequencing of longitudinal (pre and post-transplant) samples (Supplementary Fig. 7; Supplementary Data 4). In post-transplant specimens, HLA genes were often down-regulated, specifically class II loci HLA-DRB1, -DQB1 and -DPB1 (Fig. 4a, Source Data 3). In those with low HLA expression, somatic HLA aberrations were virtually absent, except for 1 case with a mutation in HLA-DRB1 (Fig. 4b). Overall, expression of classical immune checkpoint genes was not significantly altered at relapse, whereas mRNA levels of interferon-gamma response pathway and antigen presentation machinery genes were predominantly decreased (Fig. 4c). In addition, we also observed a down-modulation of Th1 and Th2 cytokine-induced signaling pathways, toll like receptor signaling, neutrophil degranulation, TP53-mediated apoptosis and caspase activation in post-transplant relapses (Supplementary Fig. 8; Supplementary Data 5).

As an illustrative example, we performed single-cell RNA sequencing on pre-sorted CD34+ cells obtained at diagnosis and at post-transplant relapse of an AML patient with *IDH2*, *DNMT3A* and *RUNX1* mutations (Supplementary Fig. 9). CD34 enrichment was

**Table 2 | Patient and transplant characteristics of the genotyped cohort (N = 48)**

	N (% or IQR)
Median Age at HCT	49 (20–73)
Disease	
AML	35 (73%)
MDS	13 (27%)
Disease status at transplant	
CR1	32 (67%)
CR2	16 (33%)
Donor type	
MRD	20 (42%)
MUD	20 (42%)
Haplo	8 (16%)
Graft type	
BM	26 (54%)
PBSC	22 (46%)
Conditioning regimen	
RIC	25 (52%)
MAC	23 (48%)
Median FUP (months)	12.33 (2.4–109.5)
Acute GvHD (II–IV)	16 (33%)
Chronic GvHD (moderate-severe)	10 (20%)
Median time of relapse after HCT (months)	5.8 (1.67–56.3)
DLI recipients	25 (52%)
Time to DLI (months)	7.8 (2.9–61.08)
Extramedullary disease at relapse	4 (8%)

IQR interquartile range, HCT hematopoietic cell transplantation, AML acute myeloid leukemia, MDS myelodysplastic syndrome, MPN myeloproliferative neoplasms, MRD matched related donor, MUD matched unrelated donor, MMUD mismatched unrelated donor, Haplo haploidentical donor, BM bone marrow, PBSC peripheral blood stem cells, MAC myeloablative, RIC reduced intensity, GvHD graft versus host disease, DLI donor lymphocyte infusion.

performed to mitigate donor cell contamination in this specimen, collected when donor chimerism was <30%. Although no changes in HLA expression and immune checkpoint regulators were found, the top represented down-regulated signatures included an enrichment of IL-1, IL-2, IL-5, TNF and TGF response pathways, indicating an impaired adaptive immune response.

### Low contribution of non-HLA immunogenomic aberrations in post-transplant leukemia relapses

We then asked whether, as in other malignancies, AML/MDS relapses could be constellated by immunogenomic alterations in genes outside HLA loci. To answer this question, we performed whole exome sequencing (WES) on a subset of patients at diagnosis (N = 9), post-chemotherapy relapse (N = 2) and post-HCT relapse (N = 8). These samples were investigated for the presence of mutations affecting antigen presentation and processing machinery and immune checkpoint regulators on target cells (Supplementary Data 5). Applying stringent criteria for variant selection (see methods) and excluding all possibly germline variants, as well as variants with unknown significance (reported in Supplementary Data 6), we did not find a specific enrichment in pathogenic or likely pathogenic somatic aberrations in post-transplant specimens, supporting the concept that direct immune evasion from HLA-related mechanisms represents the major contributor of GvL escape. Of note is that two of these patients (one at diagnosis and the other at post-transplant relapse) were carriers of HLA aberrations, confirmed in the respective WES specimens.

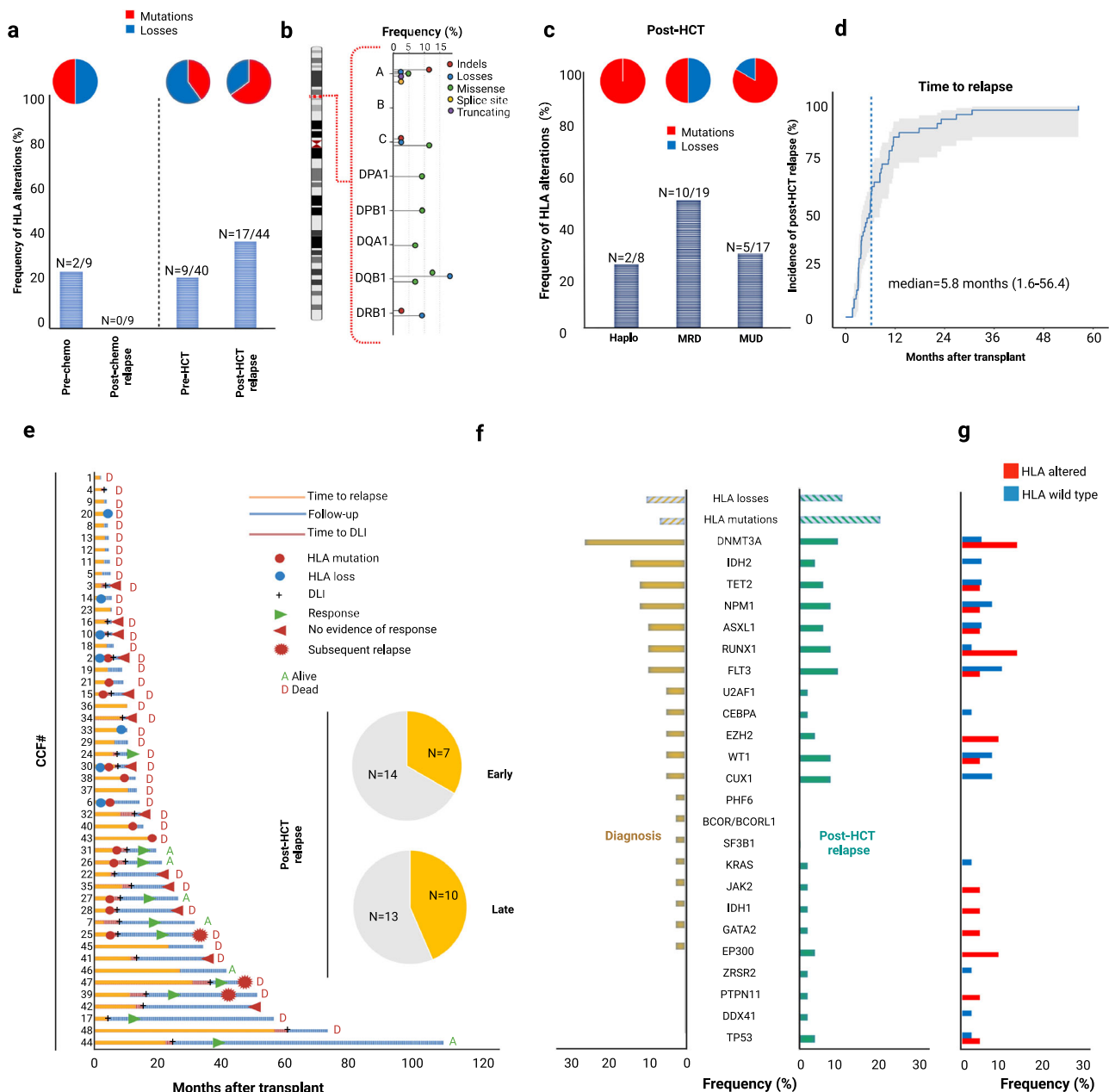
**Evaluation of contribution of HED and KIR ligands on relapse**  
HLA-C alleles serve as KIR-ligand and can be classified in C1 (Asp80) and C2 groups (Lys80) according to their KIR specificity<sup>30–32</sup>. HLA-C2 homozygosity with or without donor activating KIR2DS2 genotype (or other KIR genotypes) has been shown to increase the risk of relapse in HLA-matched allo-HCT<sup>30</sup>. Thus, the effect of HED in C locus on propensity for relapse could depend on the presence of homozygous C1/C1 or C2/C2 vs heterozygous C1/C2 genotypes. When we classified all the recipient/donor HLA-C alleles of the matched cohort according to the genotypic groups and investigated the clinical outcomes, we did not observe any impact on OS, acute and chronic GvHD and relapse (Supplementary Fig. 10A–D) in univariable models. Similarly, when C haplotype was tested in a multivariate setting (including also class I or C HED, age, donor and conditioning type), we did not observe any contributing effect on relapse, while HED-C continued to affect the risk of recurrence (Supplementary Fig. 10E, F). To complete these observations, KIR genotypes were characterized for 13 matched related donor (MRD)/recipient pairs and described in Supplementary Data 7.

## Discussion

The structural variability of HLA loci along with the presentation of a diversified spectrum of antigens cooperate with a highly heterogeneous TCR repertoire to maintain immune competence<sup>33</sup>. During tumor immunotherapy with immune checkpoint inhibitors, various modes of immune escape have been described including adaptive and genetically-encoded clonal escape<sup>20,34–36</sup>. Similarly, in immune-mediated AA, somatic LOH, deletions and mutations of HLA loci have been described<sup>12–14</sup>. Such pathophysiologic mechanisms may also be operative as acquired resistance factors in allogeneic anti-leukemia surveillance following allo-HCT, wherein either down-modulation of HLA expression or clonal deletion of disparate alleles have been described at relapse<sup>6–8</sup>. Here, through a combination of NGS-based HLA and myeloid genotyping and immune-transcriptomic analyses, we defined more granular molecular mechanisms culminating in immune-resistance phenotypes.

With a previously validated bioanalytic framework<sup>29,37</sup>, we were able to identify somatic single nucleotide mutations and small indels in HLA genes along with specific allelic losses. We show that somatic HLA mutations can constitute a potential escape mechanism in patients with MN relapsing after allo-HCT in both matched related and unrelated settings. Indeed, both mutations and deletions of one of the HLA alleles could be detected in fully matched allo-HCT, whereas it is remarkable that in the mismatched setting HLA mutations affected both matched and mismatched alleles. We therefore conclude that the immune selection process leading to clonal escape targets the allele presenting the most immunodominant antigenic peptide, likely not limited to the mismatched ones<sup>6</sup>. Indeed, HLA LOH events often involve large genomic portions of the short arm of chromosome 6, thereby affecting a whole haplotype and not only the mismatched allele and resulting in a global impairment of antigen presentation<sup>4,38–40</sup>. The finding of HLA mutations in relapsed allo-HCT is further supported by the low response rate to DLI. The analysis of the available cohort showed that patients with HLA mutations (including non-UTR alterations) had decreased likelihood of responsiveness to this salvage strategy.

In contrast to tumor control post-immunotherapy, class II HLA genes more than class I were the major drivers of immunologic escape. This result underscores the role of tumor-surveillance of CD4 T-cell specificities<sup>41–44</sup> and antigen-presenting cells, including leukemia-derived dendritic cells<sup>45–48</sup> in GvL effect. As to the function of HLA-C alleles, whose HED scores also influenced the risk of relapse, one can speculate that their allelic heterogeneity determines the degree of natural killer (NK) alloreactivity in the genetic context of inhibitory and stimulatory KIR genes<sup>49–52</sup>. Indeed, our multivariate analysis showed



**Fig. 3 | HLA and myeloid landscape of AML and MDS relapsing after allo-HCT.** **a** Barplot indicating the frequency of HLA mutations in each studied group. Numbers of patients harboring HLA aberrations are reported on the top of each bar. Pie charts capture the frequency of mutations and losses in each group. **b** Lollipop chart showing the frequency of each HLA aberration. The height of the line denotes the frequency, while the color of each dot indicates the type of aberration. **c** Frequency of HLA alterations according to donor group (Haplo: haploidentical donor); MRD (matched related donor); MUD (matched unrelated donor). Pie charts represent the distribution of mutations or losses in each group.

**d** Time dependent curve indicating the incidence of relapse of the sequenced group (median time to relapse was 5.8 months). **e** Swimming plot capturing main clinico-biological information for each patient of the sequenced cohort. The yellow barplot indicates the frequency of HLA alterations in late (after 6 months) vs early (before 6 months) post-HCT relapses. **f** Barplots showing the frequency and distribution of HLA mutations, losses and mutations in main myeloid driver genes in AML/MDS at diagnosis (yellow) and at post-HCT relapses (green). **g** Frequency of mutations in myeloid driver genes in HLA-altered vs HLA wild type post-HCT specimens. Source Data 1 and 2.

that locus C HED could influence, over the KIR ligand stratification, the probability of disease recurrence.

A reduced immunopeptidome diversity due to low amino acid divergence at the antigen-binding site level between HLA alleles may impair the GvL effect, configuring a pathophysiological condition similar to the loss of heterozygosity. Indeed, the simple divergence of recipient germline HLA polymorphisms impacts on the risk of relapse. However, high divergence did not translate into selection pressure for the emergence of HLA escape mutants, thus generating an immunogenetic environment able to prevent relapse without favoring HLA

aberrations. To that end, the impact of high HED is also demonstrated by its relationship to the TCR specificity spectrum. For instance, our study also showed that high global mean HED is indeed associated with a broad spectrum of anti-tumor and pathogen-directed TCR specificities. Nevertheless, the locus-specific frequencies of HLA mutations were in line with the findings of the analysis on divergence. Class II loci constantly appeared more mutated than class I genes, with DQB1 and DRB1 alleles strongly compromised by losses and mutations.

Another important aspect of immune-escape relapse after allo-HCT is highlighted by the results of the transcriptional analysis of





somatic dysfunction of HLA heterogeneity<sup>33</sup>. In particular, we demonstrate that germline-determined class II HLA divergence and somatic class II HLA mutations, indels or losses can enable an environment of GvL resistance, immune evasion and unfavorable outcomes. Future studies exploring larger cohorts of patients are warranted to better establish how germline and somatic HLA dysfunction may inform daily clinical practice.

## Methods

### Study design and IRB approval

This is a retrospective study aiming at dissecting the genetic determinants of immune escape in AML and MDS patients undergoing allogeneic HCT; three centers contributed to this effort providing clinical and biological data and material: Cleveland Clinic (CCF, Ohio, USA), Nancy Hospital (France), Washington University in St. Louis (WashU, Missouri, USA), (Supplementary appendix, Supplementary Fig. 1; Table 1). We integrated immunogenomic and transcriptomic data from 494 patients (Table 1; Data Source 1) who received allo-HCT for AML ( $N = 294$ ), MDS ( $N = 125$ ) and MPN ( $N = 75$ ) and followed in all three institutions: CCF ( $N = 344$ ), Nancy Hospital ( $N = 143$ ) and WashU ( $N = 7$ ). Pre-transplant clinical HLA genotyping was available in 487 patients and was used as benchmark for HED computation. An ad-hoc targeted high throughput sequencing for classical HLA loci and myeloid genes was performed on selected samples at disease onset ( $N = 40$ ), post-chemotherapy relapse ( $N = 9$ ) and post-transplant relapse ( $N = 44$ ). In addition, 13 patients presenting with disease recurrence were profiled by means of RNA sequencing on sequential diagnosis/post-transplant relapse samples (median 179 days post-HCT, IQR 129–297), whereas single cell transcriptomics was performed on paired specimens for a patient with early relapse (day +92). This research was conducted under the Institutional Review Board (IRB) and local ethics committees of the three centers. All the procedures involving human subjects were carried out under the legacy and the ethical principles of the Declaration of Helsinki.

### Sample collection

Blood or bone marrow (BM) specimens at diagnosis, post-chemotherapy or post-HCT relapses were collected in ethylenediaminetetraacetic acid (EDTA) tubes, and cryopreserved until further use after Ficoll-Paque isolation and suspension in Dimethyl sulfoxide (DMSO) and Fetal bovine serum (FBS) containing media. For each sample we prioritized DNA extraction for HLA genotyping, and targeted myeloid high throughput sequencing. T-cell receptor immunosequencing was performed on a subset of selected samples, taking into account recipients' HED scores. In case of residual cells, RNA extraction and sequencing were performed. BM specimens collected at the moment of diagnosis and post-transplant relapse for a patient experiencing early relapse were profiled for single-cell RNA sequencing, after flow-based CD34+ cell separation.

### Genotyping studies

HLA sequencing and myeloid targeted sequencing were performed on patient samples included in the biorepository of the Translational Hematology and Oncology Research Department at CCF.

Genomic DNA was isolated directly from cryopreserved unfractionated peripheral or BM blood mononuclear cells with the Nuclei Lysis Solution (Promega) according to manufacturer's instructions.

HLA targeted sequencing was performed with TruSight HLA v2 (Illumina, San Diego California) as indicated previously<sup>29,57</sup>. In brief, 11 HLA loci (Class I HLA-A, B, and C; Class II HLA-DRB1/3/4/5, HLA-DQA1, HLA-DQB1, HLA-DPA1, and HLA-DPB1) were amplified with a long-range polymerase chain reaction (PCR). After amplification, a transposon-based DNA tagmentation was applied to generate DNA amplicons, via DNA fragmentation and addition of adapter sequences. Additional PCR steps provided sequence adapters and indexing

primers to generate sequencing-ready DNA libraries. Prepared libraries were then loaded directly onto a MiSeq System for sequencing. SSO-PCR methods were used to generate HLA genotypes for patients transplanted before 2013. HLA genotyping data used for HED computation were issued from the histocompatibility laboratories of the institutions involved and were coded based on a 2 or 4-field (4–6 digits) nomenclature<sup>58</sup>.

Myeloid targeted sequencing was performed on 94 specimens paralleled sequenced for HLA, using a custom panel for detection of myeloid somatic variants from three sequencing platforms: TruSight, TruSeq, and Nextera (Illumina, San Diego, CA, USA). Forty-one most frequent leukemia associated genes common to the three panels were considered for this study (Supplementary Data 3). Sequencing libraries were generated according to an Illumina paired-end library protocol. The enriched targets were sequenced using a HiSeq 2000 or MiSeq (Illumina, San Diego, CA, USA). Paired-end sequenced reads were aligned with the Burrows-Wheeler Aligner (BWA)<sup>59</sup> to GRCh37 reference and post-alignment processing included sorting, marking of duplicates, indexing, base recalibration, according to Genome Analysis Tool Kit v.4 best practices<sup>60</sup>. Variant calling was performed with HaplotypeCaller and VarScan2<sup>60,61</sup>. Variants were annotated using Annovar. Variants with minimum coverage less than 20 or number of high-quality reads less than 5 were filtered out. An in-house developed bio-analytic pipeline identified somatic/germline mutations using sequences derived from controls and mutational databases such as dbSNP138, 1000 Genomes or ESP 6500 database, and Exome Aggregation Consortium (ExAC) as previously published<sup>62–64</sup>. Median coverage of the myeloid panel was 826x. For our variant calling algorithm, the threshold of 20 reads' depth was set-up in previous studies from our group<sup>62,65,66</sup>. Our somatic mutation detection algorithm was previously validated and demonstrated an overall accuracy of 98.7% through comparison with PCR sanger sequencing<sup>65,67</sup>.

Whole exome sequencing was applied to a subset of samples studied for the presence of molecular events in other immune non-HLA genes, related to antigen presentation machinery and T cell activation and potentially involved in immune escape (Supplementary appendix).

### HLA mutational analysis

The bioinformatic approach to investigate somatic HLA mutational status is reported in the supplementary appendix and has been previously described<sup>29,57</sup>.

### KIR genotyping and HLA-C status assignment

KIR genotyping was performed as previously described, using a sequence-specific oligonucleotide probe platform (One Lambda, West Hills, CA) and sequence-specific primers (Thermo Fisher Scientific, Waltham, MA)<sup>68,69</sup>. Recipient HLA-C status was assigned according to the allelic C group configurations: I) C1 homozygous: if patients carried only alleles belonging to the supertypes: C\*01, C\*03, C\*07, C\*08, C\*09, C\*10, C\*12, C\*14, C\*16, C\*17; C2 homozygous, in case of presence of supertypes C\*02, C\*04, C\*05, C\*06, C\*15 and C1/C2 heterozygous in presence of a alleles of both groups<sup>30,31</sup>.

### TCR $\beta$ chain sequencing and analysis

Immunosequencing of the CDR3 regions of human TCR $\beta$  chains was performed using the ImmunoSEQ Assay (Adaptive Biotechnologies, Seattle, WA), as previously described<sup>70–72</sup>. In brief, extracted genomic DNA was amplified in a bias-controlled multiplex PCR, with (i) a first PCR step consisting in forward and reverse amplification primers specific for every V and J gene segment, to allow the amplification of the hypervariable CDR3 region, and (ii) a second PCR adding a proprietary barcode sequence and Illumina adapter sequences. CDR3 libraries were sequenced on an Illumina MiSeq system according to the manufacturer's instructions. ImmunoSeq Analyzer 3.0 suite was used

for sample export and preliminary statistics and quality control steps while R Bioconductor<sup>73</sup> environment and Immunarch R<sup>74</sup> suite were used for all the downstream analyses as previously described (Supplementary Data 8)<sup>21</sup>.

### HED computation

High-quality, protein level (2nd field) HLA data in patient and control cohorts were used as input for HED computation. HED scores were generated for all the subjects and genotypes in the study using the algorithm published by Pierini and Lenz (<https://sourceforge.net/projects/granthamdist/>) for the calculation of the amino acid sequence divergence at the antigen binding sites of HLA molecules<sup>75</sup>. Briefly, starting from a dictionary including all the protein sequences of exons 2 and 3 for class I alleles and exon 2 for class II alleles, assembled from the IPD-IMGT/HLA database v.3.41 we calculated HED for 6 class I (A, B, C) and II HLA loci (DRB1, DQB1, DPB1). The scores so computed were used for all the downstream analyses.

### Bulk RNA sequencing and analysis

Paired diagnosis/post-transplant relapsed samples obtained from 13 patients (7 previously reported<sup>53</sup>) served for transcriptomic studies. For this last group of samples, total RNA was purified with the NucleoSpin RNA kit (Takara Bio USA, Inc.; Mountain View, CA, USA) according to the manufacturer's instruction. RNA quality check was performed with Agilent 2100 Bioanalyzer. mRNA was enriched using oligo(dT) beads and then was fragmented randomly by adding fragmentation buffer. The cDNA was synthesized using mRNA template and random hexamers primer, after which a custom second-strand synthesis buffer (Illumina, dNTPs, RNase H, and DNA polymerase I) was added to initiate the second-strand synthesis. After a sequential process of terminal repair, a ligation, and sequencing adapter ligation, the double-stranded cDNA library was completed through size selection and PCR enrichment. Illumina technology (NovaSeq 6000) was used for sequencing, after pooling, and >30 million reads were acquired for each sample. FastQC was used to check sequenced reads quality. After trimming and adapter removal, raw reads were mapped to the human hg19 reference genome using RNA STAR<sup>76</sup>. For the transcriptomic study, cohort assembly was performed starting from read counts for all the samples in study. Batch effect was removed through a two-way ANOVA algorithm<sup>77</sup> via limma package prior to further analyses, taking into account the design matrix used to describe comparisons between the samples (diagnosis vs post-HCT relapse), to mitigate the effects derived from different sequencing batches.

Genes that were lowly expressed across >80% of the samples were filtered out. For each sample a normalization factor was calculated through the trimmed mean of M values (TMM) method and final logarithmic counts per million were calculated (log2CPM). Differential expression and gene set enrichment analysis were assessed using edgeR 3.32.1 and limma 3.46 with R computational environment (v. 4)<sup>78</sup>. Benjamini-Hochberg procedure was used for multiple testing correction.

### Single-cell RNA sequencing and analysis

Pre-sorted CD34+ cells BM cells from patient CCF#8 were processed for single-cell library preparation as previously described<sup>79</sup>. Library preparation was performed as per manufacturer instructions with the Illumina Nextera XT DNA sample preparation kit (Illumina, San Diego, California). Briefly a first tagmentation reaction was carried out on 1 ng of cDNA at 55 C for 10 min. The amplification of adapter-ligated fragments was performed using Nextera PCR master mix, index primers and barcodes were added to identify each cell for 15 cycles. The PCR products from each cell were then pooled together and purified using Sera-mag SpeedBeads (0.6:1 ratio). A final quality check of the cDNA library was performed using an Agilent high-sensitivity DNA chip, obtaining a broad peak between 300–800 bp. Single cells were

sequenced using 100 bp paired end sequencing on the HiSeq2500. Raw reads were assessed using FastQC v0.11.5 and aligned to the hg19 human reference genome using the STAR aligner v2.5.3a, after trimming and adapter removal. Rsubread v1.32.4 was used to assemble read matrices from aligned bam files. Cells with fewer than 150,000 reads, along with <65% unique mapping and <35% exonic region coverage, were excluded from further analysis. Genes expressed in less than 10% of cells per patient were removed from downstream analysis. Only those cells that passed quality control were included in the downstream analysis. Seurat R package was used on feature-barcoded matrices for clustering and differential analyses and for visualization purposes.

### Statistical analyses

Median, interquartile ranges (IQR), mean and 95% CI intervals were used where appropriate. Frequency and distribution of categorical variables were expressed as percentage. For all relevant comparisons, after testing for normal distribution, comparative analyses between two groups were performed by Wilcoxon matched-pair signed rank test at 95% CI. Fisher's exact test or Chi-square were applied for independent group comparisons.

Each HED variable was categorized in high and low HED based on the 50th percentile cutoff of the distribution in healthy controls.

Probabilities of survival for OS, defined as the time from transplantation to death for any cause or last follow-up, was calculated using Kaplan-Meier estimates<sup>80</sup>, with differences between the curves based on log-rank tests for univariate comparisons. Cumulative incidence of acute and chronic GVHD and relapse were calculated in a competing risk setting, where death was considered the competing event<sup>81</sup>.

Multivariate cox regression models were built on cumulative incidence of relapse retaining class I and II HED as main effect variables.

All statistical tests were two-sided, and a *P*-value < 0.05 was considered statistically significant.

All of the analyses and data visualization were performed using the statistical computing environment R (4.0.0 R Core Team, R Foundation for Statistical Computing, Vienna, Austria) and excel Microsoft 365. BioRender.com was used for figures' assembly.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All the data that support the findings of this study are available within the Article and Supplementary Files.

Genomic data is available through the dbGAP-controlled access database, accession number dbGaP: [phs003235.v1](https://dbgap.ncbi.nlm.nih.gov/ah/equery.fcgi?db=phs003235.v1). Access can be granted through dbGAP, and contact can be made to Jaroslaw P. Maciejewski ([maciejj@ccf.org](mailto:maciejj@ccf.org)). There are no restrictions on who will be granted access. Source data are provided with this paper.

### Code availability

The pipeline used for the HLA mutational study has been deposited in the following repository: <https://github.com/SMNPAG/HLA-mutations> under the <https://doi.org/10.5281/zenodo.7651570>.

### References

1. Passweg, J. R. et al. Hematopoietic cell transplantation and cellular therapy survey of the EBMT: monitoring of activities and trends over 30 years. *Bone Marrow Transpl.* **56**, 1651–1664 (2021).
2. Döhner, H. et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447 (2017).

3. O'Neill, A. T. & Chakraverty, R. Graft versus leukemia: current status and future. *Perspect. JCO* **39**, 361–372 (2021).
4. Vago, L. & Gojo, I. Immune escape and immunotherapy of acute myeloid leukemia. *J. Clin. Invest.* **130**, 1552–1564 (2020).
5. Hong, S. et al. Pattern of somatic mutation changes after allogeneic hematopoietic cell transplantation for acute myeloid leukemia and myelodysplastic syndromes. *Bone Marrow Transpl.* **57**, 1615–1619 (2022).
6. Vago, L. et al. Loss of mismatched HLA in leukemia after stem-cell transplantation. *N. Engl. J. Med.* **361**, 478–488 (2009).
7. Toffalori, C. et al. Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation. *Nat. Med.* **25**, 603–611 (2019).
8. Stölzel, F. et al. Clonal evolution including partial loss of human leukocyte antigen genes favoring extramedullary acute myeloid leukemia relapse after matched related allogeneic hematopoietic stem cell transplantation. *Transplantation* **93**, 744–749 (2012).
9. Vago, L., Toffalori, C., Ciceri, F. & Fleischhauer, K. Genomic loss of mismatched human leukocyte antigen and leukemia immune escape from haploidentical graft-versus-leukemia. *Semin. Oncol.* **39**, 707–715 (2012).
10. Dermime, S. et al. Immune escape from a graft-versus-leukemia effect may play a role in the relapse of myeloid leukemias following allogeneic bone marrow transplantation. *Bone Marrow Transpl.* **19**, 989–999 (1997).
11. Shyr, D. C. et al. HLA-haplotype loss after TCR $\alpha\beta$ /CD19-depleted haploidentical HSCT. *Bone Marrow Transpl.* **56**, 733–737 (2021).
12. Ahmad, M., Rees, R. C. & Ali, S. A. Escape from immunotherapy: possible mechanisms that influence tumor regression/progression. *Cancer Immunol. Immunother.* **53**, 844–854 (2004).
13. Ritter, C. et al. Epigenetic priming restores the HLA class-I antigen processing machinery expression in Merkel cell carcinoma. *Sci. Rep.* **7**, 2290 (2017).
14. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).
15. Fuchs, E. J. et al. HLA informs risk predictions after haploidentical stem cell transplantation with posttransplantation cyclophosphamide. *Blood* **139**, 1452–1468 (2022).
16. Sajulga, R., Bolon, Y.-T., Maiers, M. J. & Petersdorf, E. W. Assessment of HLA-B genetic variation with an HLA-B leader tool and implications in clinical transplantation. *Blood Adv.* **6**, 270–280 (2022).
17. Petersdorf, E. W. et al. Role of HLA-DP expression in graft-versus-host disease after unrelated donor transplantation. *J. Clin. Oncol.* **38**, 2712–2718 (2020).
18. Arrieta-Bolaños, E. et al. A core group of structurally similar HLA-DPB1 alleles drives permissiveness after hematopoietic cell transplantation. *Blood* 2022015708. <https://doi.org/10.1182/blood.2022015708> (2022).
19. Ruggeri, A. et al. Integrating biological HLA-DPB1 mismatch models to predict survival after unrelated hematopoietic cell transplantation. *Haematologica*. <https://doi.org/10.3324/haematol.2021.280055> (2022).
20. Chowell, D. et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat. Med.* **25**, 1715–1720 (2019).
21. Pagliuca, S. et al. Clinical and basic implications of dynamic T cell receptor clonotyping in hematopoietic cell transplantation. *JCI Insight* **6**, e149080 (2021).
22. Daull, A.-M. et al. Class I/Class II HLA evolutionary divergence ratio is an independent marker associated with disease-free and overall survival after allogeneic hematopoietic stem cell transplantation for acute myeloid leukemia. *Front. Immunol.* **13**, 841470 (2022).
23. Roerden, M. et al. HLA evolutionary divergence as a prognostic marker for AML patients undergoing allogeneic stem cell transplantation. *Cancers (Basel)* **12**, 1835 (2020).
24. Merli, P. et al. Human leukocyte antigen evolutionary divergence influences outcomes of paediatric patients and young adults affected by malignant disorders given allogeneic haematopoietic stem cell transplantation from unrelated donors. *Br. J. Haematol.* <https://doi.org/10.1111/bjh.18561> (2022).
25. Lin, W.-Y. et al. Genome-wide association study identifies susceptibility loci for acute myeloid leukemia. *Nat. Commun.* **12**, 6233 (2021).
26. Chen, D.-P., Chang, S.-W., Wang, P.-N., Hus, F.-P. & Tseng, C.-P. Association between single nucleotide polymorphisms within HLA region and disease relapse for patients with hematopoietic stem cell transplantation. *Sci. Rep.* **9**, 13731 (2019).
27. Boudreau, J. E. et al. KIR3DL1/HLA-B subtypes govern acute myelogenous leukemia relapse after hematopoietic cell transplantation. *J. Clin. Oncol.* **35**, 2268–2278 (2017).
28. Pagliuca, S. et al. The similarity of class II HLA genotypes defines patterns of autoreactivity in idiopathic bone marrow failure disorders. *Blood* 2021012900. <https://doi.org/10.1182/blood.2021012900> (2021).
29. Pagliuca, S. et al. Molecular landscape of immune pressure and escape in aplastic anemia. *Leukemia* <https://doi.org/10.1038/s41375-022-01723-w> (2022).
30. Venstrom, J. M. et al. HLA-C –dependent prevention of leukemia relapse by donor activating KIR2DS1. *N. Engl. J. Med.* **367**, 805–816 (2012).
31. Stringaris, K. et al. Donor KIR genes 2DL5A, 2DS1 and 3DS1 are associated with a reduced rate of leukemia relapse after HLA-identical sibling stem cell transplantation for acute myeloid leukemia but not other hematologic malignancies. *Biol. Blood Marrow Transplant.* **16**, 1257–1264 (2010).
32. Sobecks, R. M. et al. Influence of killer immunoglobulin-like receptor/HLA ligand matching on achievement of T-cell complete donor chimerism in related donor nonmyeloablative allogeneic hematopoietic stem cell transplantation. *Bone Marrow Transpl.* **41**, 709–714 (2008).
33. Pagliuca, S., Gurnari, C., Rubio, M. T., Visconte, V. & Lenz, T. L. Individual HLA heterogeneity and its implications for cellular immune evasion in cancer and beyond. *Front. Immunol.* **13**, 944872 (2022).
34. O'Donnell, J. S., Teng, M. W. L. & Smyth, M. J. Cancer immunoeediting and resistance to T cell-based immunotherapy. *Nat. Rev. Clin. Oncol.* **16**, 151–167 (2019).
35. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
36. Mizuno, S. et al. Immunogenomic pan-cancer landscape reveals immune escape mechanisms and immunoeediting histories. *Sci. Rep.* **11**, 15713 (2021).
37. Gurnari, C. et al. Clinical and Molecular Determinants of Clonal Evolution in Aplastic Anemia and Paroxysmal Nocturnal Hemoglobinuria. *JCO* 22.00710. <https://doi.org/10.1200/JCO.22.00710> (2022).
38. Afaible, M. G. et al. SNP array-based karyotyping: differences and similarities between aplastic anemia and hypocellular myelodysplastic syndromes. *Blood* **117**, 6876–6884 (2011).
39. Savage, S. A. et al. Genome-wide association study identifies HLA-DPB1 as a significant risk factor for severe aplastic anemia. *Am. J. Hum. Genet.* **106**, 264–271 (2020).
40. Montesin, M. et al. Somatic HLA class I loss is a widespread mechanism of immune evasion which refines the use of tumor mutational burden as a biomarker of checkpoint inhibitor response. *Cancer Disco.* **11**, 282–292 (2021).

41. Ferraro, F. et al. Immunosuppression and outcomes in adult patients with de novo acute myeloid leukemia with normal karyotypes. *Proc. Natl Acad. Sci. USA* **118**, e2116427118 (2021).
42. Stevanović, S., van Schie, M. L. J., Griffioen, M. & Falkenburg, J. H. HLA-class II disparity is necessary for effective T cell mediated Graft-versus-Leukemia effects in NOD/scid mice engrafted with human acute lymphoblastic leukemia. *Leukemia* **27**, 985–987 (2013).
43. Matte-Martone, C. et al. Differential requirements for myeloid leukemia IFN- $\gamma$  conditioning determine graft-versus-leukemia resistance and sensitivity. *J. Clin. Investig.* **127**, 2765–2776 (2017).
44. Jandin, A. et al. Age-related immune cell dynamics influence outcomes after allogeneic haematopoietic cell transplantation. *Br J Haematol.* <https://doi.org/10.1111/bjh.18822> (2023).
45. Woiciechowsky, A., Regn, S., Kolb, H.-J. & Roskrow, M. Leukemic dendritic cells generated in the presence of FLT3 ligand have the capacity to stimulate an autologous leukemia-specific cytotoxic T cell response from patients with acute myeloid leukemia. *Leukemia* **15**, 246–255 (2001).
46. Dickinson, A. M. et al. Graft-versus-leukemia effect following hematopoietic stem cell transplantation for leukemia. *Front Immunol.* **8**, 496 (2017).
47. Schroeder, M. A., Choi, J., Staser, K. & DiPersio, J. F. The role of Janus kinase signaling in graft-versus-host disease and graft versus leukemia. *Biol. Blood Marrow Transpl.* **24**, 1125–1134 (2018).
48. Li, N. et al. Anti-CD3 preconditioning separates GVL from GVHD via modulating host dendritic cell and donor T-cell migration in recipients conditioned with TBI. *Blood* **113**, 953–962 (2009).
49. Joncker, N. T., Fernandez, N. C., Treiner, E., Vivier, E. & Raulat, D. H. NK cell responsiveness is tuned commensurate with the number of inhibitory receptors for self-MHC class I: the rheostat model. *J. Immunol.* **182**, 4572–4580 (2009).
50. Boudreau, J. E. & Hsu, K. C. Natural killer cell education in human health and disease. *Curr. Opin. Immunol.* **50**, 102–111 (2018).
51. Boudreau, J. E. et al. Cell-extrinsic MHC class I molecule engagement augments human NK cell education programmed by cell-intrinsic MHC class I. *Immunity* **45**, 280–291 (2016).
52. Vivier, E. et al. Innate or adaptive immunity? The example of natural killer cells. *Science* **331**, 44–49 (2011).
53. Christopher, M. J. et al. Immune escape of relapsed AML cells after allogeneic transplantation. *N. Engl. J. Med.* **379**, 2330–2341 (2018).
54. Dufva, O. et al. Immunogenomic landscape of hematological malignancies. *Cancer Cell* **38**, 380–399.e13 (2020).
55. Pagliuca, S. et al. Comprehensive transcriptomic analysis of VISTA in acute myeloid leukemia: insights into its prognostic value. *Int. J. Mol. Sci.* **23**, 14885 (2022).
56. Gambacorta, V. et al. Integrated multiomic profiling identifies the epigenetic regulator PRC2 as a therapeutic target to counteract leukemia immune escape and relapse. *Cancer Discov.* **12**, 1449–1461 (2022).
57. Gurnari, C. et al. Is nature truly healing itself? Spontaneous remissions in paroxysmal nocturnal hemoglobinuria. *Blood Cancer J.* **11**, 187 (2021).
58. Robinson, J. et al. IPD-IMGT/HLA database. *Nucleic Acids Res.* gkz950. <https://doi.org/10.1093/nar/gkz950> (2019).
59. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
60. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1–11.10.33 (2013).
61. Koboldt, D. C. et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
62. Nagata, Y. et al. Invariant patterns of clonal succession determine specific clinical features of myelodysplastic syndromes. *Nat. Commun.* **10**, 5386 (2019).
63. Hirsch, C. M. et al. Consequences of mutant TET2 on clonality and subclonal hierarchy. *Leukemia* **32**, 1751–1761 (2018).
64. Awada, H. et al. Invariant phenotype and molecular association of biallelic TET2 mutant myeloid neoplasia. *Blood Adv.* **3**, 339–349 (2019).
65. Hirsch, C. M. et al. Molecular features of early onset adult myelodysplastic syndrome. *Haematologica* **102**, 1028–1034 (2017).
66. Nagata, Y. et al. Machine learning demonstrates that somatic mutations imprint invariant morphologic features in myelodysplastic syndromes. *Blood* **136**, 2249–2262 (2020).
67. Makishima, H. et al. Dynamics of clonal evolution in myelodysplastic syndromes. *Nat. Genet.* **49**, 204–212 (2017).
68. Sobecks, R. M. et al. Survival of AML patients receiving HLA-matched sibling donor allogeneic bone marrow transplantation correlates with HLA-Cw ligand groups for killer immunoglobulin-like receptors. *Bone Marrow Transpl.* **39**, 417–424 (2007).
69. Hong, S. et al. Influence of killer immunoglobulin-like receptors and somatic mutations on transplant outcomes in acute myeloid leukemia. *Transplant. Cell. Ther.* **27**, 917.e1–917.e9 (2021).
70. Robins, H. S. et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* **114**, 4099–4107 (2009).
71. Robins, H. et al. Ultra-sensitive detection of rare T cell clones. *J. Immunol. Methods* **375**, 14–19 (2012).
72. Carlson, C. S. et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat. Commun.* **4**, 2680 (2013).
73. Gentleman, R. *Bioinformatics and computational biology solutions using R and Bioconductor.* (Springer Science+Business Media, 2005).
74. Nazarov, V., Immunarch.Bot & Rumynskiy, E. immunomind/immunarch: 0.6.5: Basic single-cell support. <https://doi.org/10.5281/ZENODO.3367200> (2020).
75. Pierini, F. & Lenz, T. L. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. *Mol. Biol. Evol.* **35**, 2145–2158 (2018).
76. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
77. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).
78. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
79. Stetson, L. C. et al. Single cell RNA sequencing of AML initiating cells reveals RNA-based evolution during disease progression. *Leukemia* **35**, 2799–2812 (2021).
80. Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457 (1958).
81. Fine, J. P. & Gray, R. J. A proportional hazards model for the sub-distribution of a competing risk. *J. Am. Stat. Assoc.* **94**, 496–509 (1999).

## Acknowledgements

This work was supported by US National Institute of Health (NIH) grants R35 HL135795, R01HL123904, R01 380HL118281, R01 HL128425, R01 HL132071, Edward P. Evans Foundation and The Leukemia & Lymphoma Society TRP Award 6645-22 (all to J.P.M.), Aplastic Anemia and MDS International Foundation, Italian Society of Hematology, Fondation ARC pour la Recherche sur le Cancer, MDS Foundation Tito Bastianello Award (to S.P.); VeloSano Pilot Award and Vera and Joseph Dresner Foundation–MDS (to V.V.). C.G. was supported by a grant from the Edward P. Evans Foundation and the American-Italian Cancer Foundation. We greatly thank Lucia D’Aprano, for her IT assistance in the

development of the HLA mutational pipeline. We thank all the reviewers who contributed to improving this work.

### Author contributions

S.P. designed the study, collected, analyzed and interpreted the data, developed the pipeline for the variant calling in HLA region, performed the bioinformatic and statistical analyses, and wrote the manuscript. C.G. performed NGS experiments, clinical and molecular data collection and participated in the analysis interpretation and critical manuscript revision. C.H. developed NovoHLA and the methodology for the calculation of copy number variation. S.He. helped in optimizing the bioanalytical HLA workflow and developing the associated bioinformatic pipeline. S.Ho. collected clinical and genotyping data. A.A., A.D. performed HLA genotyping in Nancy Hospital. M.D.A., M.T., H.E.C., B.K.H., N.S.M., F.F. participated in patient recruitment and management and helped with clinical data collection and interpretation, gave important intellectual inputs and edited the manuscript. V.V. helped in data collection and interpretation, figure conceptualization, gave important intellectual inputs and edited the manuscript. J.P.M. designed and conceptualized the study, supervised genomic experiments, provided funding and resources, interpreted the data analysis and edited the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-38113-4>.

**Correspondence** and requests for materials should be addressed to Jaroslaw P. Maciejewski.

**Peer review information** *Nature Communications* thanks Sawa Ito, Naser Ansari-Pour and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023