Recurrent transcriptional responses in AML and MDS patients treated with decitabine

Pawan Upadhyay[a,1], **Jeremy Beales**[a,2], **Nakul M. Shah**[b], **Agata Gruszczynska**[a], **Christopher A. Miller**[a], **Allegra A. Petti**[c], **Sai Mukund Ramakrishnan**[a], **Daniel C. Link**[a], **Timothy J. Ley**[a], **John S. Welch**[a]

[a]Division of Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, MO;

[b]Department of Medicine, Washington University School of Medicine, St. Louis, MO;

[c]Department of Neuro-logical Surgery, Washington University School of Medicine, St. Louis, MO

Abstract

The molecular events responsible for decitabine responses in myelodysplastic syndrome and acute myeloid leukemia patients are poorly understood. Decitabine has a short serum half-life and limited stability in tissue culture. Therefore, theoretical pharmacologic differences may exist between patient molecular changes in vitro and the consequences of in vivo treatment. To systematically identify the global genomic and transcriptomic alterations induced by decitabine in vivo, we evaluated primary bone marrow samples that were collected during patient treatment and applied whole-genome bisulfite sequencing, RNA-sequencing, and single-cell RNA sequencing. Decitabine induced global, reversible hypomethylation after 10 days of therapy in all patients, which was associated with induction of interferon-induced pathways, the expression of endogenous retroviral elements, and inhibition of erythroid-related transcripts, recapitulating many effects seen previously in in vitro studies. However, at relapse after decitabine treatment, interferon-induced transcripts remained elevated relative to day 0, but erythroid-related transcripts now were more highly expressed than at day 0. Clinical responses were not correlated with epigenetic or transcriptional signatures, although sample size and interpatient variance restricted the statistical power required for capturing smaller effects. Collectively, these data define global hypomethylation by decitabine and find that erythroid-related pathways may be relevant because they are inhibited by therapy and reverse at relapse.

Graphical Abstract

Among acute myeloid leukemia (AML) therapies, decitabine has several unique features. It is well-tolerated, which enables administration to older and frail patients. It can be administered in the outpatient setting. And, until recently, it was commonly used as a single agent. However, the mechanisms of decitabine activity and responses *in vivo* remain incompletely understood.

Decitabine is a cytosine analog that is incorporated into newly synthesized DNA. When incorporated into DNA, this substituted base interacts covalently with DNMT1, the maintenance DNA methyl- transferase, inhibiting DNA methylation across the genome, and the hypomethylation induced by decitabine is reversible after the drug is withdrawn [1–7].

We previously used Illumina BeadChip technology to assess in vivo methylation changes induced after 10 days of decitabine [8,9]. This approach provides broad but incomplete CpG analysis (485,000 CpGs, representing ~1.5% of CpGs in the genome, with a bias toward CpGs near promoter/enhancer regions). We observed statistically overlapping but incomplete hypomethylation on day 10 in responding and non-responding patients, suggesting that decitabine "hit the target" in both groups of patients and that equivalent hypomethylation occurred across diverse genomic features (e.g., promoters, gene bodies, islands, and non-island shores). Whole-genome bisulfite sequencing provides a more complete analysis of DNA methylation changes associated with decitabine therapy in vivo, but has not previously been reported for the study of treated patients.

Initial in vitro approaches suggested that decitabine-induced perturbation of the methylome results in transcriptional reprogramming and derepression of potential tumor suppressors, and cellular differentiation effects [10]. Further analysis identified re-expression of multiple

embryonic transcripts, including Melanoma Antigen Genes (MAGE) transcripts [11], cancer testes antigens [12,13], *Nanog* [14], and *Oct4* [15]. Decitabine also has been shown to induce expression of a wide range of transposable elements [16]. Activation of these epigenetically silenced transcripts could lead to anti-leukemic responses via the expression of recognizable neo-antigens, through cellular stressors, or through induced interferon responses [11,17,18].

To more comprehensively define the epigenetic and transcriptional changes induced by decitabine therapy, we evaluated primary bone marrow samples from a series of patients treated with single-agent 10-day decitabine [8]. Samples were assessed using whole-genome bisulfite sequencing (WGBS), RNA sequencing (RNA-seq), and single-cell RNA sequencing (scRNA-seq). We identified patterns of global hypomethylation, but were unable to identify regions of the genome that were differentially susceptible to decitabine-induced hypomethylation on day 10. In addition, we also were not able to find regions differentially susceptible to re-methylation by day 28. These two findings suggest decitabine has non-specific, global effects on the genome. During treatment, decitabine induced signatures of maturation, interferon activation, and inhibited erythroid signatures (which included a set of porphyrin synthesis genes, globin genes, and erythroid maturation genes). At relapse, the interferon activation signatures persisted, whereas inhibition of erythroid signatures was reversed, with expression levels increased relative to day 0. This suggests that interferon signaling may not be relevant for decitabine responses, whereas erythroid pathways may be relevant for response.

METHODS

Patients and sample collection

All patients evaluated in this study had been enrolled in a prospective clinical trial using decitabine 20 mg/m^2/day on days 1–10 of 28-day cycles (Clinicaltrials.gov NCT01687400) [8]. The trial was approved by the institutional review board at Washington University in St. Louis and was conducted in accordance with the provisions of the Declaration of Helsinki. All the patients provided written informed consent that explicitly included genome sequencing and data sharing with qualified investigators.

The molecular mechanisms of decitabine responses remain incompletely understood. To capture diverse possible response mechanisms, NCT01687400 enrolled a breadth of patients typically exposed to decitabine, including transfusion-dependent myelodysplastic syndromes (MDS), older AML, and relapsed/refractory AML. In this manuscript we evaluate samples collected in NCT01687400. For technical reasons, we biased analysis to patients with high tumor burden, defined by exome sequencing or cytogenetics, regardless of blast counts, and included both MDS and AML subjects. For all studies, cryopreserved viable cells were thawed and processed immediately with minimal manipulation. Total DNA or RNA was extracted for WGBS or RNA Seq or viable cells were subject to standard single cell analysis preparation.

Quantification and statistical analysis

Statistical analysis was performed using either the software described in the supplemental methods, in R (Version 3.6, R Foundation, Vienna, Austria), or GraphPad Prism (Version 9, GraphPad, San Diego, CA). Hypothesis testing in R was performed using x^2 tests for categorical variables, t tests for continuous variables, Wilcoxon signed-rank tests for comparing means of gene expression quartiles, or Fisher's exact test for differences between proportional data.

Deposition of sequence data

All sequencing data for all studies were deposited to the dbGaP study ID phs000159.

RESULTS

Whole-genome methylation analysis

We used whole-genome bisulfite sequencing (WGBS) to evaluate trios of samples from 29 patients collected from bone marrow aspirates on days 0, 10 and 28 during decitabine therapy (treatment consisted of 20 mg/m^2/day on days 1–10 of 28-day cycles) (Supplementary Table E1, online only, available at www.exphem. org). Sequencing resulted in coverage of ~89% of the CpGs in the human reference genome, with a mean of $7.6 \times$ coverage for each sample ($n = 87$, 29 samples each for day 0, day 10, and day 28) (Supplementary Table E2). As expected, we observed more complete coverage of CpG sites in WGBS data as compared with our prior array data (27 million CpGs vs. 450,000 interrogated; Supplementary Figure E1A). Day 10 samples were associated with global hypomethylation in patient bone marrow samples compared with day 0 samples, and we observed incomplete restoration of methylation levels by day 28 (Figure 1; Supplementary Figure E1B). Neither the extent of hypomethylation on day 10, nor the extent of remethylation on day 28 correlated with response (Supplementary Figure E2), consistent with prior results [8].

We examined context-specific global methylation patterns observed at annotated genes, transcription factor binding sites (TFBSs), DNase-resistant sites (BP-DNase), and CpG islands and across 25 different chromatin annotated states (ChromHMM-25). These annotated regions were each associated with canonical methylation patterns at day 0 (e.g., hypermethylation upstream of a transcriptional start site, hypomethylation at the transcriptional start site and hypermethylation along the gene body). At day 10, these patterns were globally retained across all evaluated annotated regions but were associated with global, nonspecific hypomethylation. Likewise, these patterns were retained during remethylation at day 28, suggestive of global, random hypomethylation and remethylation processes (Figure 1C; Supplementary Figure E1B) [9]. Using differentially methylated region (DMR) analysis, we could not identify recurrent, high-confidence DMRs within the data set that were recurrent across patients or clinically defined subgroups (all identified recurrent DMRs were associated with outlier cases, and none passed manual review).

Interpatient heterogeneity dominated global transcriptional variance

To examine the transcriptional changes induced by 10 days of decitabine treatment, we evaluated RNA-seq data from the bone marrow aspirate cells of 30 MDS/AML patients collected on days 0 and 10 of decitabine treatment (Supplementary Table E1). Total bone marrow aspirate cells were used, and samples with evidence of tumor burden involving more than 70% of the marrow were prioritized for analysis (heterozygous founding clone variant allele frequency [VAF] >35%). Within an unsupervised hierarchical analysis of the top 1,000 most variable transcripts across all samples, individual samples clustered together by patient rather than by treatment day (Figure 2A). Differentially expressed transcripts were identified between day 0 and day 10 samples, noting only modest fold changes for most and more transcripts with increased expression ($N = 427$) than decreased expression ($N = 228$) (Figure 2B; Supplementary Table E3). Among these transcripts, many have been previously identified as regulated by decitabine exposure (e.g., *ANPEP* (alanyl aminopeptidase, membrane), *HIRA* (histone cell cycle regulator), *SEL1L* (SEL1L adaptor subunit of ERAD E3 ubiquitin ligase), *IRF9* (interferon regulatory factor 9), *SERPINA1* (serpin family A member 1) (Up); *HBB* (hemoglobin subunit β), *HBA1* (hemoglobin subunit α1) and *HBA2* (hemoglobin subunit a2) (down). Within upregulated transcripts, we observed multiple lineage-determining transcription factors (e.g., *CEBPE* [CCAAT enhancer-binding protein e], *GATA1* [GATA binding protein 1]), as well as interferon-regulated transcripts (e.g., *IFI44* [interferon induced protein 44]). A separate gene set enrichment analysis (GSEA) analysis examining expressed transcripts between days 0 and 10 revealed positive enrichment of interferon α and γ response gene sets and negative enrichment of erythroid-related and MYC target gene sets (Figure 2C), consistent with previous reports [9,19–21].

Derepression of endogenous retroviruses

As activation of endogenous retroviruses (ERVs) by decitabine has been documented in various cancer types, including MDS and AML, we investigated differential expression of ERV transcripts in bone marrow samples after decitabine treatment at day 10. We observed increases in total ERV expression at day 10; however, it was modest in absolute terms (i.e., less than twofold) (Figure 3A,B). In total, 17 ERV families were significantly overexpressed (adjusted p value < 0.05), which included HERVH, HERV9, and HERV3 (Figure 3C,E; Supplementary Table E4), which overlap with previously reported decitabine-induced ERV families [22]. Induced ERV families were heterogeneous between patients, and we could not identify specific families with consistent regulation by decitabine across any clinically or molecularly defined subgroups of patients (Figure 3D).

Interferons have been associated with the activation of ERVs [16]. Therefore, we assessed the correlation of the increase in ERV family expression and a list of annotated interferon-activated transcripts [23]. We did observe a trend toward a positive correlation between ERV families and expression of some interferon-induced genes (e.g., *IFITM1, OAS1*), but none were significant (p value < 0.05) (Figure 3F).

scRNA-seq-based interrogation of MDS/AML patients' bone marrow

To improve the resolution of detection and to define decitabine effects within malignant cells, we performed scRNA-seq on total bone marrow aspirate cells from 10 patients

collected on days 0 and 10 of decitabine treatment (Supplementary Table E5). Given restrictions based on the cost of these studies, our comparisons of responder and nonresponder subsets could not be adequately powered. Therefore, we selected cases that were associated with clinical responses (4 cases with CR, 4 with mCR, 6 with CRi, 5 with CCR). Cases were prioritized with multiple mutations that could be evaluated for transcriptional effects within subsets of cells, and these results were integrated with prior data from two normal bone marrow donor samples [24].

Data were processed and cells were included based on the number of detected transcripts (700–5,000), mitochondrial counts (<10%), and ribosomal counts (<50%) (Supplementary Figure E3 and Supplementary Table E6). We used a *k*-nearest-neighbor algorithm trained on the Data Management and Access Plan (DMAP) database to estimate the lineage of each cell after performing typical data filtration methods with the Partek Flow software (Figure 4; Supplementary Figure E4A). Again, we observed strong interpatient heterogeneity, with bone marrow cells clustering based on patient origin rather than treatment status (Figure 4A,B; Supplementary Figure E4A,B). We observed that myeloid cells and lymphoid cells (predominantly T cells) clustered separately, and that myeloid cells formed clusters with monocytic/common myeloid precursor (CMP) or hematopoietic stem cell (HSC))/ erythroid signatures (Figure 4C; Supplementary Figure E4C). The bone marrow cells from normal donors clustered with both myeloid and lymphoid cell patient subsets, suggesting that much of the transcriptional hierarchy of normal bone marrow lineage populations was still reflected in these MDS/AML patients (Supplementary Figure E4). Surprisingly, transcriptional differences induced by decitabine treatment at day 10 did not result in separate Uniform Manifold Approximation and Projection (UMAP) clusters (Figure 4B; Supplementary Figure E4B). We also did not observe recurrent enrichment of any lineage-defined cell type or the proportion of *MIKI67*-positive cells at day 10 (Supplementary Figure E4D).

Differential gene expression analysis was performed by first filtering out lymphoid cells based on CD3E expression and lymphoid cell DMAP signatures to enable a more enriched analysis of the residual "myeloid" cells (Figure 4). Within the remaining "myeloid" cells from the combined 10 patients, we identified 120 genes that were significantly differentially expressed (fold change ≥ 2.0, false discovery rate [FDR] ≤ 0.05) (Figure 4D; Supplementary Table E7). Again, the majority were upregulated (Figure 4E, *n* = 101). We noted upregulation of a set of previously reported decitabine-regulated genes, includ-ing *CEBPE* [25], *COL14A1* (collagen type XIV *a*1 chain), *IFI27* (interferon a-inducible protein 27) [26–28], *PNMA5* (PNMA family member 5) [9,11], *TRPM4* (transient receptor potential cation channel subfamily M member) [29], *PRG2* (proteoglycan 2, pro-eosinophil major basic protein) [30], and *IRF7* (interferon regulatory factor 7) [27,31]. Furthermore, a separate GSEA analysis revealed significant (FDR ≤ 0.05) positive enrichment of interferon *a* and *γ* response gene sets and negative enrichment of TNFa signaling via NFkB and heme metabolism (Supplementary Table E8), consistent with our findings in RNA-Seq and previous reports [9,19–21].

Recurrent decitabine-induced transcriptional changes in scRNA-seq

Because interpatient heterogeneity reduced our ability to detect decitabine-induced changes, we assessed differential expression between days 0 and 10 within the scRNA-seq data from individual patients and then evaluated recurrence effects between patients. Differentially expressed genes (DEGs) (FDR 0.05, fold change [FC] –2, 2) were noted in all patients (median = 316 per case, range: 94–699; Supplementary Figures E5–E14). The absolute fold change noted for DEGs within individuals tended to be larger than in the "combined" analysis, but day 10 transcripts were rarely greater than 10-fold different from day 0. In myeloid cells across MDS/AML patients, we observed some heterogeneity in response within individual patients, but transcriptional shifts tended to occur globally; we observed only two cases with "blocks" of cells that exhibited discordant responses that might represent subclones or subpopulations with differential responses to decitabine compared with other cellular subsets within an individual patient (patients 1014 and 1062; Supplementary Figures E6F and E9F).

Individual cases were independently assessed for pathway-level regulation using GSEA. Recurrently altered GSEA pathways (FDR 0.05 and recurrent in at least 5 cases) overlapped with the GSEA results from RNA-seq and again included heme metabolism (down in 6 cases); interferon α, interferon γ, and inflammatory pathways (up in 4, 3, and 4 cases, but also downregulated in some cases); and coagulation (up in 5 cases) (Figure 5A).

DEG analysis between days 0 and 10 within individual patients identified a total of 3,870 genes that were significantly regulated on day 10 (FC 2.0, FDR 0.05) across 10 MDS/AML patients (Supplementary Table E9). Within this set, we observed 104 genes that were recurrently altered in at least 4 of 10 cases. These included myeloid maturation-related transcripts (*CEBPE, ELANE* [elastase, neutrophil expressed], *MPO* [myeloperoxidase], *CSTG* [cathepsin G], *AUZ1* [azurocidin 1]); interferon-regulated transcripts: (*IFI27, IFI30, IFI44L* [interferon-induced protein 44 like], *IFITM1* [interferon-induced transmembrane protein 1]), *EGR1* [early growth response 1], *E2F1* [E2F transcription factor 1], *FOXM1* [forkhead box M], *ID1* [inhibitor of DNA binding 1], *CCR1* [C–C motif chemokine receptor 1], *CD14*); and globin gene transcripts: *HBB, HBD, and HBA1/2* (Figure 5B; Supplementary Figure E15).

Because we had observed induction of maturation-related transcripts in both RNA-seq and scRNA-seq, we sought to quantify global maturation programs within the scRNA-seq data. We performed the analysis in the PartekFlow software using the "AUCell" task. The input to AUCell is a gene set, and the output is a gene set "activity" (AUC) score assigned to each cell [32]. We used previously defined gene signatures developed using AML populations, which separate hematopoietic stem cell/progenitor-like signatures (HSC-like), granulocyte/monocyte progenitor-like signatures (GMP-like), and myeloid tumor signatures [33]. Using this approach, we observed a significant increase in GMP-like programs in 7 of 10 cases following decitabine therapy, whereas HSC-like and myeloid tumor signatures were not consistently modified across the 10 cases (Supplementary Figure E16).

As a secondary validation of expression regulation, the most differentially expressed transcripts identified in scRNA-seq were re-examined within the RNA-seq data set. Within

the top three, up- and downregulated transcripts *ATXN3L, HIRA, SEL1L, HBB, HBA1,* and *HBA2* were all differentially expressed with a significance of $p < 0.05$ when examined in isolation (i.e., without multiple testing correction), and exhibited broad trends in the RNA-seq data that reflected scRNA-seq results (Supplementary Figure E17A). Of the top three interferon-regulated transcripts and top three maturation-related transcripts, only *CEBPE, IFI44,* and *GATA1* were associated with at least $p < 0.05$ in the RNA-seq data (Supplementary Figure E17B).

To determine whether these decitabine-regulated transcripts identified via scRNA-seq were recurrently observed beyond this data set, we performed a literature search and identified 11 previously published gene expression studies of decitabine treatment (Supplementary Table E10A). Within these published studies, upregulation of *ANPEP, COL14A1, IFI27, IRF9, PNMA5,* and *PRG2* was noted in 5 of 11 studies, and upregulation of *IFI44L, IFI6, IRF7, MPO, MX1* was reported in 3 of 11 studies. Lower expression levels of *HBB, HBA1, and HBA2* were reported in 2 of 4 studies (most studies did not report lower expression levels with decitabine) (Supplementary Table E10B,C). These studies examined cell line or primary culture samples in vitro, and variable concentrations of decitabine were used for treatment. Despite these caveats, this approach appears to validate recurrent alterations in the abundance of several transcripts in AML and MDS samples that were recurrently identified in scRNA-seq.

Mutation-defined subclonal transcriptional signatures

Using a recently developed algorithm [34], we evaluated the expression of somatic mutations within individual cells in the scRNA-seq data analyzed on day 0, independent of decitabine treatment. Of the 10 patients evaluated, only one displayed a transcriptional shift associated with a subclonal mutation (Figure 6A). Patient 1058 harbored a mutation (A303P) in *CEBPA (CCAAT Enhancer Binding Protein α)* that was associated with transcriptional clusters (C1 and C5) and an HSC cell population signature, whereas the *CEBPA* non-mutant clusters (C2, C8, and C9) were associated with megakaryocyte erythroid progenitor (MEP)/erythroid population signatures (Figure 6A). By comparing the cell populations between the clusters containing *CEBPA* A303P cells (*CEBPA* mutant) and the clusters containing *CEBPA* wild-type cells (*CEBPA* wild type), we identified 1,592 genes that were significantly differently expressed (FC ±2.0, FDR 0.05) (Figure 6C,D; Supplementary Table E11). Among the most highly differentially expressed transcripts were lineage-determining transcription factors and maturation-associated transcripts: increased in the mutant cluster, *CEBPA, CEBPE, CSF3R* (colony-stimulating factor 3 receptor), and *OAS2* (20–50-oligoadenylate synthetase 2); decreased in the mutant cluster, *GATA1, HBB, HBD, SLC40A1* (solute carrier family 40 member 1) (Figure 6E), consistent with the lineage shift identified in DMAP assigned cell types.

T-Cell subpopulation analysis

We evaluated T-cell-restricted populations as part of a separate analysis (Supplementary Figure E18A). Like myeloid cells, decitabine treatment was not associated with separate clusters on UMAP projection (Supplementary Figure E18B). Unexpectedly, we identified cells with tumor-associated mutations in the T-cell identified populations (Supplementary

Figure E18C,D). These included genes associated with founding clone events (e.g., *U2AF1*), but also later progression events (e.g., *KRAS*). Differential expression analysis revealed only 59 genes to be differentially expressed at day 10 (fold change ± 2.0, FDR 0.05) (Supplementary Figure E18E,F; Supplementary Table E12). At day 0, bone marrow T cells were less frequently associated with *MKI67* expression than myeloid cells (1% vs. 10%, x^2 $p < 0.0001$; Supplementary Figure E4D), suggesting that they are less proliferative than AML cells, and this may contribute to reduced transcriptional changes induced by a DNA-integrating drug like decitabine.

Transposable element regulation in scRNA-seq

To interrogate the expression of transposable elements (TEs) within the scRNA-seq data, we used the *scTE* [7] pipeline [35]. This approach evaluates the expression of LTR retroposons, LINEs (long interspersed nuclear elements), SINEs (short interspersed nuclear elements), and DNA transposons. Of the 10 evaluated patients, 4 exhibited upregulation of multiple TE families at day 10 after decitabine treatment (Supplementary Figure E19 vs. Supplementary Figure E20). Again, we observed interpatient heterogeneity, although five TE sub-families (*LTR12C, AluY, MIR, AluSg, and L1PA2*) were upregulated in at least 3 patients. However, even in the most recurrently upregulated transcript (LTR12C), the absolute CPM fold changes observed were also modest across the 10 patients (2- to 4-fold; Supplementary Figure E19C). TE expression did not correlate with decitabine treatment or with lineage-defined subpopulations of cells. Correlation of interferon-induced transcripts with LTR, SINE, LINE, and DNATE expression in scRNA-seq data identified trends, but not strong associations (Supplementary Figure E21).

Transcriptional effects after relapse

To determine whether decitabine-induced transcriptional changes persist at relapse, we evaluated 6 patients by performing scRNA-seq analysis of paired diagnosis and relapse samples (Supplementary Table E5). Data were processed and cells were evaluated based on the number of detected transcripts (700–5000), mitochondrial counts (<10%), and ribosomal counts (<50%) (Supplementary Figure E22A–D; Supplementary Table E13). We again noted that cells clustered more strongly by patient than by relapse status (Figure 7A). T cells clustered separately, and AML cells were again organized into monocytic/CMP or HSC/MEP assigned DMAP populations (Figure 7B). Relapse samples were not associated with recurrent shifts in DMAP-assigned lineage populations or consistent changes in *MIKI67* expression (Figure 7C–E).

We also performed an analysis of paired diagnosis and relapse samples to identify patient-specific changes (Supplementary Figures E23–E28). Patients had overlapping UMAP cell populations and DMAP cell types at relapse versus day 0, but no consistent shifts were noted in DMAP assigned cell populations, nor fractions of *MIKI67*+ cells. In only one case (1079; Supplementary Figure E27), the day 0 and relapse samples were organized into distinct UMAP clusters with morphologic shifts from a cluster that included monocyte, dendritic, erythroid, and pre-B-cell populations to a population with monocyte and NK T-cell assigned phenotypes.

Differentially expressed genes could be identified within the "combined" myeloid cells (excluding CD3E+ and DMAP lymphoid cells), and more upregulated genes were associated with the relapse samples (N = 555 upregulated vs. 156 downregulated; Figure 7F; Supplementary Table E14). GSEA analysis was performed on the "combined" myeloid cells at day 0 versus relapse (excluding CD3E+ and DMAP lymphoid cells). We observed a decrease in the Hallmark heme metabolism program and other erythroid transcripts, which contrasted with day 10 results, whereas an increase in interferon pathways paralleled day 10 results (Figure 8A; Supplementary Table E15).

Differential expression analyses were also performed for each individual patient in the relapse cohort and identified a total of 4,065 genes that were significantly dysregulated (fold change ±2.0, FDR 0.05) (Supplementary Table E16). GSEA analyses identified an increase in Hallmark heme metabolism in 4 of 6 patients at relapse, which contrasted with decreased expression of genes in this pathway in 6 of 10 patients at day 10 (Figure 8B). In contrast, the expression of interferon γ pathway genes was increased in 3 of 6 patients at relapse and also increased in 4 of 10 patients at day 10. We detected the expression of multiple globin genes at relapse, which contrasted with their low expression levels on day 10 (Figure 8C vs. Figure 5B). The expression of *MPO* and *AZU1* was lower at relapse, while it was increased on day 10; in contrast, the interferon-induced transcripts *IFI27, IFI44L,* and *IFITM1* were increased at relapse and day 10 (Figure 8C vs. Figure 5B). GMP-Like signatures were reduced in all 6 cases at relapse (Supplementary Figure E29B), which contrasted with the day 10 increase in GMP-like signatures in 7 of 10 patients (Supplementary Figure E16). Myeloid maturation signatures increased in 5 of 6 patients at relapse, with large shifts among two patients with dominant monocyte cell populations (patients 1061 and 1079) (Supplementary Figure E29C).

Manual curation of transcripts with increased expression at relapse in at least 3 cases suggested that the genes within the Hallmark heme metabolism GSEA pathway encompass both globin gene transcripts and erythroid maturation transcripts (Supplementary Table E17). This included a series of transcripts involved in iron metabolism and porphyrin synthesis (*ABCB10, ABCG2, ALAD, CPOX, EPB42, FECH, GLRX5, HMBS, SLC11A1, SLC25A37, SLC25A37, STEAP3, UROD*), as well as glutathione regulation (*BLVRB, HAGH, PRDX2, GCLM*). It also included the transcription factor *KLF1*, as well as other transcription factors associated with erythroid maturation (*HEMGN, TRIM10, TRIM58, TSPO2, MAFB*), a broad set of hemoglobin genes (*HBA1, HBA2, HBB, HBD, HBM, HBQ1*), and erythroid cell membrane proteins (*GYPA, GYPB, GYPE, SLC4A1* (Band 3), *ANK1* (Band 4.2), *RHAG, RHCE, RHD, SPTA1, SPTB, ADD2, DMTN, EPB42*), suggestive of changes in erythroid maturation at relapse. These effects are especially notable in patient 1061 (Supplementary Figure E26), who presented with predominantly monocytic/CMP populations at day 0 and at relapse, but the relapse nevertheless was associated with increased expression of diverse porphyrin synthesis, hemoglobin, and erythroid maturation genes (Supplementary Table E17). We assessed this annotated list of genes for expression differences at day 10 versus day 0 (Supplementary Table E17). *HBB* expression was lower in 6 of 10 cases, whereas all other genes had reduced expression in 2 or fewer patients (Figure 5B; Supplemental Table E17). We also evaluated the GSEA core transcripts associated with heme metabolism at relapse and at day 10. We observed variable

overlaps of porphyrin and glutathione transcript levels at day 10 and induction at relapse, but complete overlap among hemoglobin and erythroid maturation transcripts, including *KLF1,* suggesting that erythroid maturation may be consistently altered in some patients after relapse.

DISCUSSION

Decitabine is a cytosine analogue capable of incorporating into DNA with different effects depending on the dose and the frequency of exposure [9,36,37]. At lower doses, decitabine causes primarily hypomethylation by inhibiting the function of DNMT1. At higher doses, decitabine activates ATR/ATM DNA damage responses and single-strand DNA break repair [38,39], resulting in DNA stress and cytotoxicity.

The clinically relevant effects of in vivo decitabine are theoretically challenging to reproduce accurately in vitro. The serum half-life of decitabine is extremely short (8–18 min) [40–43], and decitabine is unstable in solution, with a half-life of approximately 12 hours, depending on pH and temperature [44]. Decitabine integration into DNA is cell cycle dependent, and thus, decitabine effects depend on intracellular exposure during appropriate cell cycle phases. Different cell types and tissues may also exhibit different responses to decitabine exposure, due to differences in cell cycle kinetics, drug transport and metabolism, and basal epigenetic states [36,45]. In this study, we assessed the effect of genome-wide methylation and transcription changes induced by decitabine clinical treatment, following 10 days of therapy and then at the end of the first cycle of decitabine (day 28). Assessing human bone marrow samples during therapy provides a more dose-appropriate assessment, compared with in vitro models. However, our results largely recapitulated previously published in vitro results, suggesting justification for future in vitro studies, despite theoretical pharmacologic concerns.

We applied WGBS to identify regions of the genome that were preferentially sensitive to or resistant to decitabine-induced hypomethylation. We could identify a series of recurrent statistically significant DMRs, but none of these passed manual review; these were generally associated with a small number of outlier cases and did not contain recurrent phenotypes. WGBS coverage was limited to 5–25 × (Supplementary Table E2) because of cost and anticipating that broad coverage of CpGs would enable accurate detection of local trends through averaging effects, although we cannot rule out the possibility that deeper coverage might have enabled detection of additional effects. In comparison with prior studies, some have identified recurrent CpGs with decitabine-induced hypomethylation [9]. However, large regions have not been identified, and uniform hypomethylation across broad regions of the genome have been reported elsewhere [37], consistent with our data. We noted incomplete hypomethylation on day 10 of decitabine (Supplementary Figure E2). The heterogeneous cell mixture in the bone marrow may contribute to this phenotype (e.g., different cell types have differences in cell cycle kinetics and decitabine metabolism pathways). However, reduction in methylation was also incomplete in a study that evaluated cells after a month of decitabine exposure [46], suggesting there may be a limit to the maximum feasible or tolerable hypomethylation induced by decitabine, either in vitro or after in vivo clinical exposure.

Similar to prior ex vivo decitabine treatment studies [9], we also observed that decitabine-induced transcriptional changes were modest, that global transcriptional signatures were informed primarily by the patient-specific characteristics of the samples, and interpatient differences surpassed recurrent transcriptional changes induced by decitabine (Figure 2). Similar to others, we observed recurrent effects at a small number of loci (*COL14A1,* globin, etc.) and pathways (interferon, inflammation, TNF, heme metabolism, and myeloid maturation) [9,20,21]. We also observed global induction of endogenous retroviral elements (ERVs, SINEs, LINEs, etc.) [47]. However, the absolute effect sizes of gene regulation and retro-element regulation were modest, retro-element expression did not consistently correlate with interferon expression or interferon-induced transcripts, and the specific families and retro-element transcripts induced were heterogeneous between patients.

Our study included both MDS and AML samples; we chose to focus on cases with a high tumor burden detected by exome sequencing or cytogenetics rather than to select cases based on blast counts. Across both WGBS and RNA-seq studies, we could not identify signatures that correlated with clinical characteristics or clinical responses, including MDS versus AML, blast counts, WBCs, or responses. Evaluation of larger cohorts may have enabled detection of smaller, or less recurrent, transcriptional effects and correlation with clinical features. Our goal had been to identify transcriptional effects with the most clinical prognostic utility. Smaller effects may yet be detectable, but these may not be relevant within the context of a clinical receiver operating characteristic curve and accurate prediction of responses to decitabine.

Although 9 cases evaluated with scRNA-seq were associated with multiple somatic mutations, only 1 case had a mutation that was associated with a clear subclonal expression signature in the day 0 sample. In this patient (1058, Figure 6), cells with *CEPBA* A303P clustered independently from cells without this mutation and were associated with clusters of cells with HSC expression signatures (higher expression of *CEPBA, CEPBE, CSF3R*), whereas the *CEBPA* wild-type cells were associated with erythroid clusters (higher expression of *GATA1, HBB, HBD*). In separate studies [24,34], transcription factor mutations (*CEPBA, GATA2, FOXP1*) also associated with unique subclonal transcriptional signatures. Other classes of AML- and MDS-associated mutations affect chromatin structure and/or RNA processing (e.g., epigenetic, splicing, cohesin, tumor suppressors), and would also be expected to induce transcriptional effects. However, mutations in chromatin and RNA processing genes have not yet organized unique transcriptional signatures within subsets of cells, suggesting transcription factor mutations may induce greater transcriptional effects that are more easily detected.

Within groups of cells identified transcriptionally as T cells, we also observed evidence of somatic mutations. The cohort of patients evaluated by scRNA-seq was older (median age = 75, range: 61–85), and 8 of 10 patients had mutations in clonal hematopoiesis-associated genes (*U2AF1, TP53, DNMT3A, TET2, JAK2, RAD21, SF3B1*), suggesting that some clonal variants associated with AML and MDS may occur initially in HSC compartments, but may still contribute to multi-lineage hematopoiesis. However, we cannot fully exclude the possibility that some of these variants may be associated with misclassified cells, doublets, or another technical artifact.

HSC and leukemia stem cell (LSC) signatures have been implicated in relapse. We used AUCell [48] to evaluate shifts in previously defined leukemic HSC signatures [33]. At relapse, we noted a reduction in GMP-like signatures, increase in myeloid tumor signatures, and increase in erythroid transcripts, but no increase in HSC-like signatures (Supplementary Figure E26). As decitabine therapy rarely (if ever) induced deep molecular remissions [8,49], perhaps the selective pressure exerted by this drug is inadequate to provide the selection pressure necessary to allow for rare HSC-like cells to emerge at relapse.

Immune modulation has been proposed as a possible mechanism of action for decitabine [47]. Although we observed interferon pathways induced by decitabine in both the RNA-seq and scRNA-seq data sets, these pathways persisted at relapse, relative to day 0 (Figure 8). This suggests that either the reprogramming effects of decitabine on interferon pathways remain constitutively active through the time of relapse, or cells with activation are preferentially selected as part of the relapse program. Regardless, interferon programs in AML cells persist after decitabine exposure.

In contrast, the expression of a number of genes associated with erythroid development was reduced at day 10, but expression of these pathways was higher in relapse samples. Others have observed that increases in fetal hemoglobin, detected by protein mass spectrometry after 6 weeks of therapy, correlate with clinical decitabine responses [19]. Our data examined early transcriptional effects (day 10), noting decreases in the fetal hemoglobin genes *HBG1* and *HBG2* on day 10, but more consistent reductions in *HBB* than in the fetal g-globin chains. These larger and more consistent reductions in *HBB* may alter ratios of b-globin to g-globin, enabling subsequent relative increases in fetal hemoglobin assembly. Shifts in oxidative phosphorylation states or in the synthesis of specific hemoglobin forms may be a part of the decitabine effects. Alternatively, these may simply be components of myeloid maturation programs, with a reciprocal reduction in erythroid transcripts at day 10, and relapse escape mechanisms that circumnavigate this myeloid maturation by augmenting erythroid maturation.

In sum, these data provide a genome-wide analysis of epigenetic and transcriptional consequences of decitabine treatment in MDS/AML patients undergoing therapy. Collectively, we observe global hypomethylation on day 10 and recurrent induction of interferon-regulated genes, associated with altered myeloid and erythroid maturation, after treatment and at relapse. Interpatient variability has made recurrent trends difficult to define; larger cohorts of patients and integration of comparisons with cohorts treated with azacitidine may allow for a better understanding of the molecular mechanisms associated with response.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Hagemann S, Heil O, Lyko F, Brueckner B. Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. PLoS One 2011;6:e17388. [PubMed: 21408221]

2. Santi DV, Norment A, Garrett CE. Covalent bond formation between a DNA–cytosine methyltransferase and DNA containing 5-azacytosine. Proc Natl Acad Sci USA. 1984;81:6993–7. [PubMed: 6209710]

3. Schermelleh L, Spada F, Easwaran HP, et al. Trapped in action: direct visualization of DNA methyltransferase activity in living cells. Nat Methods 2005;2:751–6. [PubMed: 16179921]

4. Grant SG, Worton RG. Activation of the hprt gene on the inactive X chromosome in transformed diploid female Chinese hamster cells. J Cell Sci 1989;92(Pt 4):723–32. [PubMed: 2480966]

5. Issa JP, Baylin SB, Herman JG. DNA methylation changes in hematologic malignancies: biologic and clinical implications. Leukemia 1997;11(Suppl 1):S7–S11. [PubMed: 9130685]

6. Ganetzky RD, Jiang Y, Falk J, et al. Differences between normal and leukemic stem cell-specific methylome indicates aberrantly silenced genes involved in the pathogenesis of malignant evolution. Blood 2008;112:599.

7. Christman JK. 5-Azacytidine and 5-aza-2ʹ-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 2002;21:5483–95. [PubMed: 12154409]

8. Welch JS, Petti AA, Miller CA, et al. TP53 and decitabine in acute myeloid leukemia and myelodysplastic syndromes. N Engl J Med 2016;375:2023–36. [PubMed: 27959731]

9. Klco JM, Spencer DH, Lamprecht TL, et al. Genomic impact of transient low-dose decitabine treatment on primary AML cells. Blood 2013;121:1633–43. [PubMed: 23297133]

10. Lubbert M. DNA methylation inhibitors in the treatment of leukemias, myelodysplastic syndromes and hemoglobinopathies: clinical results and possible mechanisms of action. Curr Top Microbiol Immunol 2000;249:135–64. [PubMed: 10802943]

11. Zhang Z, He Q, Tao Y, et al. Decitabine treatment sensitizes tumor cells to T-cell-mediated cytotoxicity in patients with myelodysplastic syndromes. Am J Transl Res 2017;9:454–65. [PubMed: 28337274]

12. Qiu X, Hother C, Ralfkiaer UM, et al. Equitoxic doses of 5-azacytidine and 5-aza-2ʹdeoxycytidine induce diverse immediate and overlapping heritable changes in the transcriptome. PLoS One 2010;5:e12994.

13. Weber J, Salgaller M, Samid D, et al. Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2ʹ-deoxycyti- dine. Cancer Res 1994;54:1766–71. [PubMed: 7511051]

14. Wang X, Wang Y, Zuo Q, et al. The synergistic effect of 5Azadc and TSA on maintenance of pluripotency of chicken ESCs by overexpression of NANOG gene. In Vitro Cell Dev Biol Anim 2016;52:488–96. [PubMed: 26822431]

15. Huan Y, Wu Z, Zhang J, Zhu J, Liu Z, Song X. Epigenetic modification agents improve gene-specific methylation reprogramming in porcine cloned embryos. PLoS One 2015;10:e0129803.

16. Chiappinelli KB, Strissel PL, Desrichard A, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell 2015;162:974–86. [PubMed: 26317466]

17. Yang H, Bueso-Ramos C, DiNardo C, et al. Expression of PD-L1, PD-L2, PD-1 and CTLA4 in myelodysplastic syndromes is enhanced by treat- ment with hypomethylating agents. Leukemia 2014;28:1280–8. [PubMed: 24270737]

18. Zhang Z, Chang CK, He Q, et al. Increased PD-1/STAT1 ratio may account for the survival benefit in decitabine therapy for lower risk myelo-dysplastic syndrome. Leuk Lymphoma 2017;58:969–78. [PubMed: 27686004]

19. Stomper J, Ihorst G, Suciu S, et al. Fetal hemoglobin induction during decitabine treatment of elderly patients with high-risk myelodysplastic syn- drome or acute myeloid leukemia: a potential dynamic biomarker of outcome. Haematologica 2019;104:59–69. [PubMed: 30171030]

20. McLaughlin LJ, Stojanovic L, Kogan AA, et al. Pharmacologic induction of innate immune signaling directly drives homologous recombination deficiency. Proc Natl Acad Sci USA 2020;117:17785–95. [PubMed: 32651270]

21. Gilmartin AG, Groy A, Gore ER, et al. In vitro and in vivo induction of fetal hemoglobin with a reversible and selective DNMT1 inhibitor. Haematologica 2021;106:1979–87. [PubMed: 32586904]

22. Daskalakis M, Brocks D, Sheng YH, et al. Reactivation of endogenous retroviral elements via treatment with DNMT- and HDAC-inhibitors. Cell Cycle 2018;17:811–22. [PubMed: 29633898]

23. Li H, Chiappinelli KB, Guzzetta AA, et al. Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. Oncotarget 2014;5:587–98. [PubMed: 24583822]

24. Petti AA, Khan SM, Xu Z, et al. Genetic and transcriptional contributions to relapse in normal karyotype acute myeloid leukemia. Blood Cancer Discov 2021;3:32–49.

25. Negrotto S, Ng KP, Jankowska AM, et al. CpG methylation patterns and decitabine treatment response in acute myeloid leukemia cells and normal hematopoietic precursors. Leukemia 2012;26:244–54. [PubMed: 21836612]

26. Moreaux J, Reme T, Leonard W, et al. Development of gene expression-based score to predict sensitivity of multiple myeloma cells to DNA methylation inhibitors. Mol Cancer Ther 2012;11:2685–92. [PubMed: 23087257]

27. Ramakrishnan S, Hu Q, Krishnan N, et al. Decitabine, a DNA-demethylating agent, promotes differentiation via NOTCH1 signaling and alters immune-related pathways in muscle-invasive bladder cancer. Cell Death Dis 2017;8:3217. [PubMed: 29242529]

28. Young CS, Clarke KM, Kettyle LM, Thompson A, Mills KI. Decitabine–vorinostat combination treatment in acute myeloid leukemia activates pathways with potential for novel triple therapy. Oncotarget 2017;8:51429–46. [PubMed: 28881658]

29. Leung KK, Nguyen A, Shi T, et al. Multiomics of azacitidine-treated AML cells reveals variable and convergent targets that remodel the cell-surface proteome. Proc Natl Acad Sci USA. 2019;116:695–700. [PubMed: 30584089]

30. Flotho C, Claus R, Batz C, et al. The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. Leukemia 2009;23:1019–28. [PubMed: 19194470]

31. Greve G, Schuler J, Gruning BA, et al. Decitabine induces gene derepression on monosomic chromosomes: in vitro and in vivo effects in adverse-risk cytogenetics AML. Cancer Res 2021;81:834–46. [PubMed: 33203699]

32. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 2017;14:1083–6. [PubMed: 28991892]

33. van Galen P, Hovestadt V, Wadsworth Ii MH, et al. Single-cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. Cell 2019;176:1265–81. e1224. [PubMed: 30827681]

34. Petti AA, Williams SR, Miller CA, et al. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. Nat Commun 2019;10:3660. [PubMed: 31413257]

35. He J, Babarinde IA, Sun L, et al. Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. Nat Commun 2021;12:1456. [PubMed: 33674594]

36. Colwell M, Wanner NM, Drown C, Drown M, Dolinoy DC, Faulk C. Paradoxical whole genome DNA methylation dynamics of 5'aza-deoxycyti-dine in chronic low-dose exposure in mice. Epigenetics 2021;16:209–27. [PubMed: 32619143]

37. Lund K, Cole JJ, VanderKraats ND, et al. DNMT inhibitors reverse a specific signature of aberrant promoter DNA methylation and associated gene silencing in AML. Genome Biol 2014;15:406. [PubMed: 25315154]

38. Diesch J, Zwick A, Garz AK, Palau A, Buschbeck M, Gotze KS. A clinical-molecular update on azanucleoside-based therapy for the treatment of hematologic cancers. Clin Epigenet 2016;8:71.

39. Patel K, Dickson J, Din S, Macleod K, Jodrell D, Ramsahoye B. Targeting of 5-aza-2´-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. Nucleic Acids Res 2010;38:4313–24. [PubMed: 20348135]

40. Cashen AF, Schiller GJ, O'Donnell MR, DiPersio JF. Multicenter, phase II study of decitabine for the first-line treatment of older patients with acute myeloid leukemia. J Clin Oncol 2010;28:556–61. [PubMed: 20026803]

41. Cashen AF, Shah AK, Todt L, Fisher N, DiPersio J. Pharmacokinetics of decitabine administered as a 3-h infusion to patients with acute myeloid leukemia (AML) or myelodysplastic syndrome (MDS). Cancer Chemother Pharmacol 2008;61:759–66. [PubMed: 17564707]

42. Chabot GG, Bouchard J, Momparler RL. Kinetics of deamination of 5-aza-2´-deoxycytidine and cytosine arabinoside by human liver cytidine deaminase and its inhibition by 3-deazauridine, thymidine or uracil arabi- noside. Biochem Pharmacol 1983;32:1327–8. [PubMed: 6189497]

43. Momparler RL, Samson J, Momparler LF, Rivard GE. Cell cycle effects and cellular pharmacology of 5-aza-2'-deoxycytidine. Cancer Chemother Pharmacol 1984;13:191–4. [PubMed: 6207950]

44. Lin KT, Momparler RL, Rivard GE. High-performance liquid chromatographic analysis of chemical stability of 5-aza-2´-deoxycytidine. J Pharm Sci 1981;70:1228–32. [PubMed: 6170748]

45. Aoyama S, Nakano H, Danbara M, Higashihara M, Harigae H, Takahashi S. The differentiating and apoptotic effects of 2-aza-5´-deoxycytidine are dependent on the PU.1 expression level in PU.1-transgenic K562 cells. Biochem Biophys Res Commun 2012;420:775–81. [PubMed: 22459451]

46. Wong YF, Jakt LM, Nishikawa S. Prolonged treatment with DNMT inhibitors induces distinct effects in promoters and gene-bodies. PLoS One 2013;8:e71099.

47. Kordella C, Lamprianidou E, Kotsianidis I. Mechanisms of action of hypomethylating agents: endogenous retroelements at the epicenter. Front Oncol 2021;11:650473.

48. Parenti S, Rontauroli S, Carretta C, et al. Mutated clones driving leukemic transformation are already detectable at the single-cell level in CD34-positive cells in the chronic phase of primary myelofibrosis. NPJ Precis Oncol 2021;5:4. [PubMed: 33542466]

49. Uy GL, Duncavage EJ, Chang GS, et al. Dynamic changes in the clonal structure of MDS and AML in response to epigenetic therapy. Leukemia 2017;31:872–81. [PubMed: 27740633]
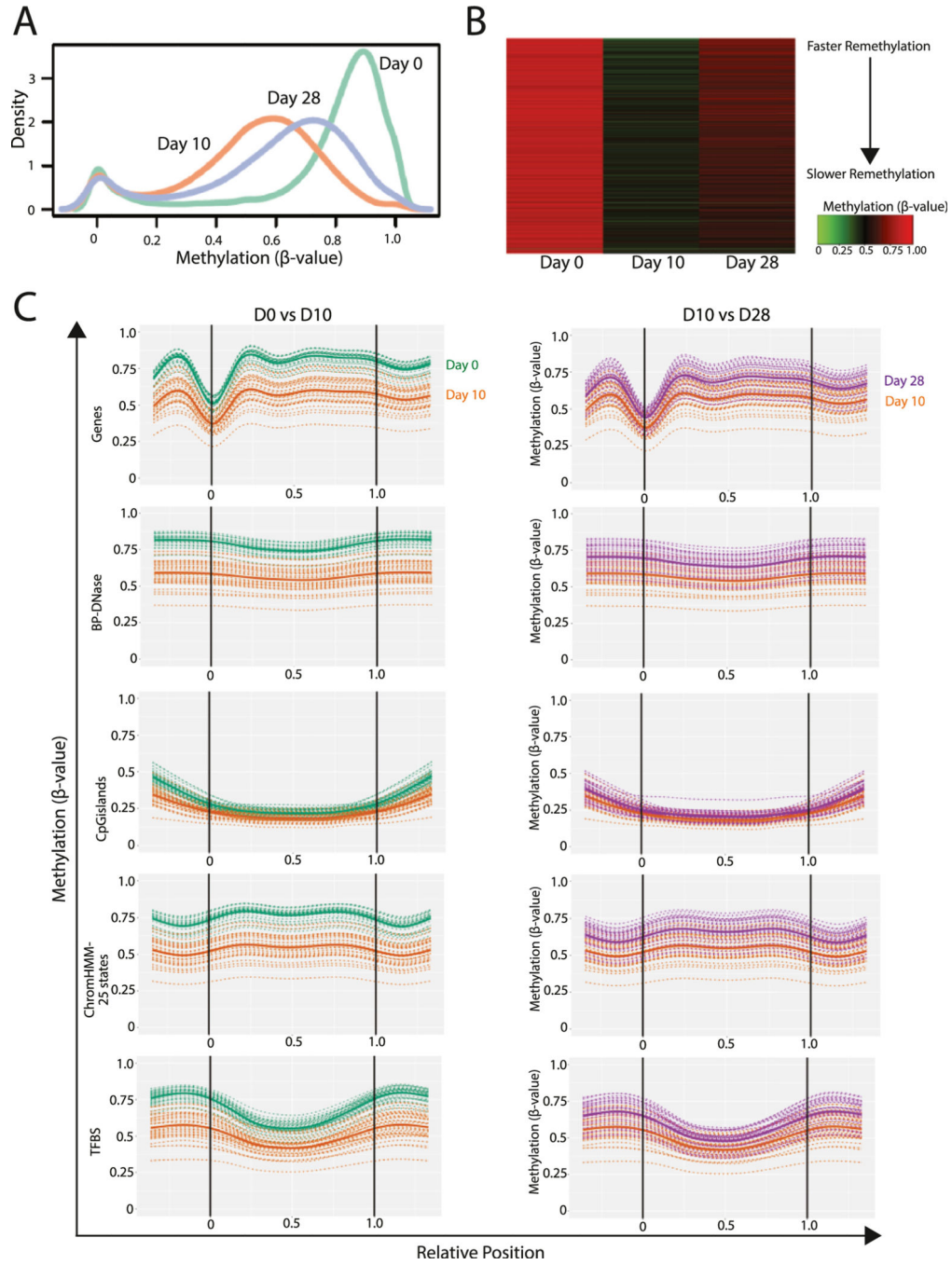
Figure 1.

Alteration in genomewide methylation levels during decitabine treatment in the bone marrow of acute myeloid leukemia and myelo-dysplastic syndrome patients. **(A)** Representative density plots of mean DNA methylation values from whole-genome methylation sequencing across all the sites for bone marrow at days 0, 10, and 28 (patient 1002). **(B)** Heatmap comparison of ~30,000 DMRs defined between the day 0 and day 10 samples of patient 1002, with "passive" plotting of the methylation density of the same regions on day 28. **(C)** Composite plots of DNA methylation levels in various genomic
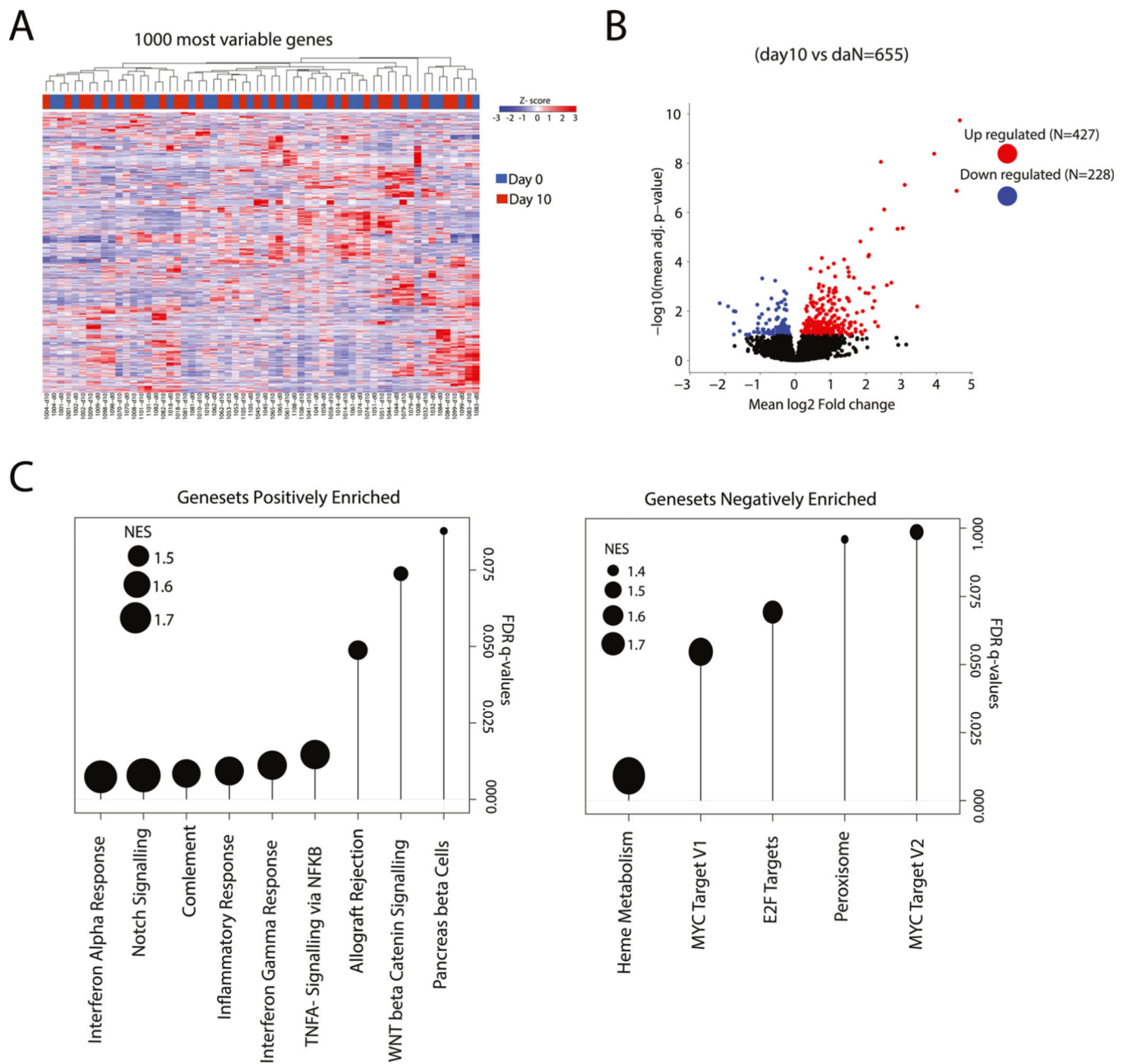
contexts across sites at days 0, 10, and 28. Each region was covered by six equally sized bins and by two flanking regions of the same size. Smoothing was done by cubic splines. Each *dotted line* represents results from a separate patient (n = 29). *DMRs*=differentially methylated regions; *TFBS*=transcription factor binding sites.

Figure 2.
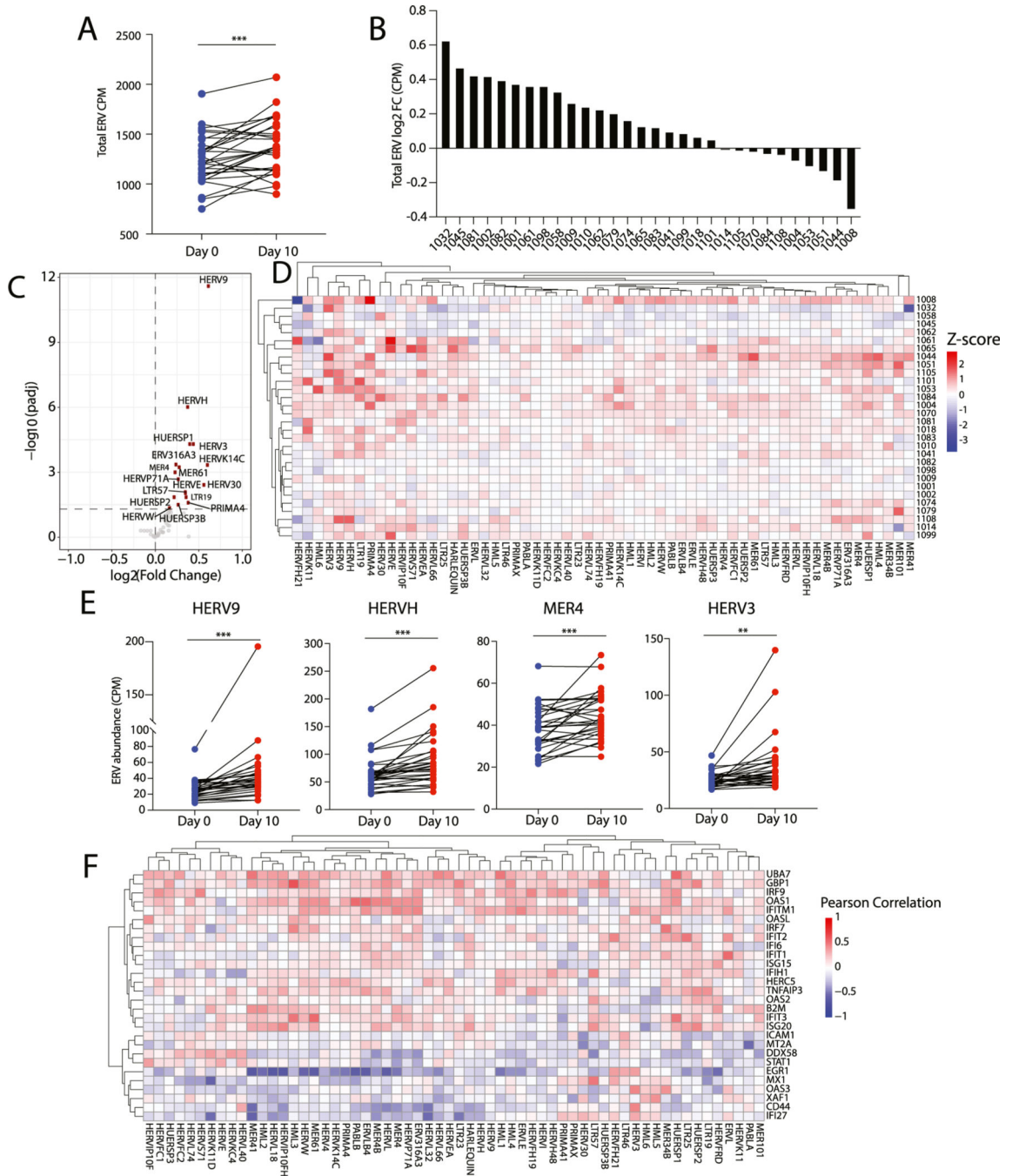Transcriptional changes in the bone marrow of AML and MDS patients during decitabine treatment identified with RNA-seq. (A) Heatmap of the top 1,000 most variably expressed mRNA transcripts across day 0 and 10 bone marrow samples ($n = 30$). (B) Volcano plots with differentially expressed transcripts at day 10 after decitabine treatment across 30 MDS/AML patients (FDR 0.1). (C) Lollipop plot of significantly enriched Hallmark gene sets based on GSEA. The *horizontal axis* represents Hallmark gene sets, the *vertical axis* represents FDR *q* values (0.10), and *circle size* depicts NES for each gene set. *AML*=acute myeloid leukemia; *FDR*=false discovery rate; *GSEA*=gene set enrichment analysis; *MDS*=myelodysplastic syndromes; *NES*=normalized enrichment score.

Figure 3.
Changes in human ERV transcripts during decitabine treatment in the bone marrow of acute myeloid leukemia and myelo-dysplastic syndrome patients identified with bulk RNA sequencing. **(A)** Line plot of total ERV read counts for individual patients on day 0 versus day 10. Significance was assessed with a paired *t* test. **(B)** Bar plot of $\log_2$ fold change in total ERV read counts by the individual patient. **(C)** Volcano plot of differentially expressed ERV families at day 10 of decitabine across 30 patients (significant ERV families in *red*, adjusted *p* value 0.05). **(D)** Heatmap of significantly differentially expressed ERV families
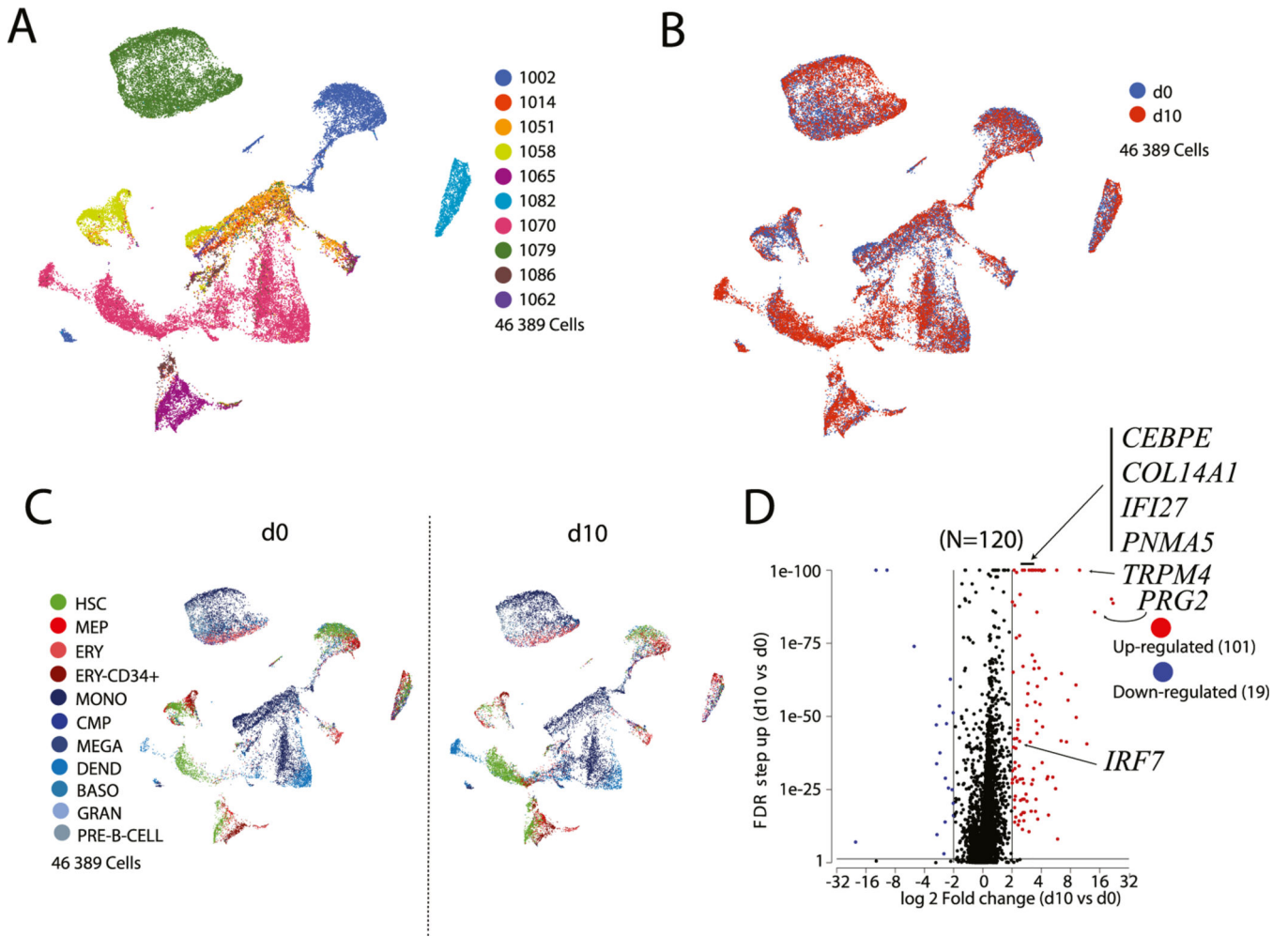
using *Z*-score normalization. **(E)** Line plot of selected significantly differentially expressed ERV families across individual patients. (F) Heatmap comparison of ERV families and transcriptional changes in interferon response genes between day 10 and day 0 bone marrow samples, using Pearson's correlation values. ***$p < 0.001$, **$p < 0.01$. *ERV*=endogenous retrovirus.

Figure 4.
Transcriptional changes in the bone marrow of AML and MDS patients during decitabine treatment identified using single cell RNA sequencing. UMAP projection of scRNAseq data from myeloid cells (CD3E⁻ and DMAP myeloid lineage) in MDS and AML patients based on 10 principal component analysis consolidations and the 1,000 most variable genes in the data set. **(A)** Data colored by the patient. **(B)** Data colored by day of decitabine treatment. **(C)** Data colored by predicted hematopoietic population using DMAP gene expression profiles on days 0 and 10. **(D)** Volcano plot highlighting differentially expressed genes within the combined data from 10 patients (FDR 0.05 and FC 2, 2). *AML*=acute myeloid leukemia; *FC*=fold change; *FDR*=false discovery rate; *MDS*=myelodysplastic syndromes; *UMAP*=Uniform Manifold Approximation and Projection.

A

	1002	1065	1070	1086	1014	1062	1051	1058	1079	1082
Heme Metabolism	-5.5	-3.7	-1.7	-4.7	-4.7	-3.9				
Interferon Gamma Response	2.5	2.8					4.5	3.8		
Inflammatory Response	2.7	2.5					2.6			3.1
Interferon Alpha Response	1.8	3.1					4.5	4.3		3.1
Coagulation	2.9	1.8	1.7				2.3		3.2	

B

	CEBPE	ELANE	MPO	CTSG	AZU1	IFI27	IFI44L	IFITM1	HBB	HBD	HBA1	HBA2	HBM	HBG2	HBG1
1002	81	3	3	13	3	3	3	NS	-77	-13	-15	-30	-72	NS	-3
1014	23	9	32	18	9	6	2	NS	-4	-3	-5	-6	NS	NS	NS
1051	NS	-2	-3	-3	-2	4	4	5	NS	NS	NS	NS	NS	NS	NS
1058	63	3	3	6	3	116	3	2	NS	NS	NS	NS	NS	5	5
1062	26	4	2	5	3	NS	NS	2	-17	NS	NS	NS	NS	NS	NS
1070	10	4	7	3	3	5	NS	0	-2	NS	NS	NS	NS	NS	NS
1086	50	3	4	4	3	5	NS	2	-5	-4	-8	-9	-9	NS	NS
1082	23	NS	NS	NS	NS	NS	NS	4	NS	NS	NS	NS	NS	NS	NS
1079	NS	2	NS	NS	NS	NS	NS	NS	-7	NS	NS	NS	NS	NS	NS
1065	NS	NS	NS	NS	NS	NS	2	NS	NS	NS	NS	NS	NS	6	NS

Figure 5.
Recurrently deregulated genes and gene sets post-decitabine treatment identified using single-cell RNA sequencing in acute myeloid leukemia and myelodysplastic syndrome patients who had responded to decitabine. **(A)** Normalized enrichment score data representation of statistically significant (false discovery rate 0.05) enriched hallmark gene sets based on gene set enrichment analysis in at least 4 of 10 patients (only significant patients are represented by *dots* on the plot). **(B)** Fold change gene expression data representation of selected recurrently regulated differentiation-related genes across 10 patients. Fold change values are indicated for each gene. Green: Upregulated and Red: Downregulated. *NS*=not significant.
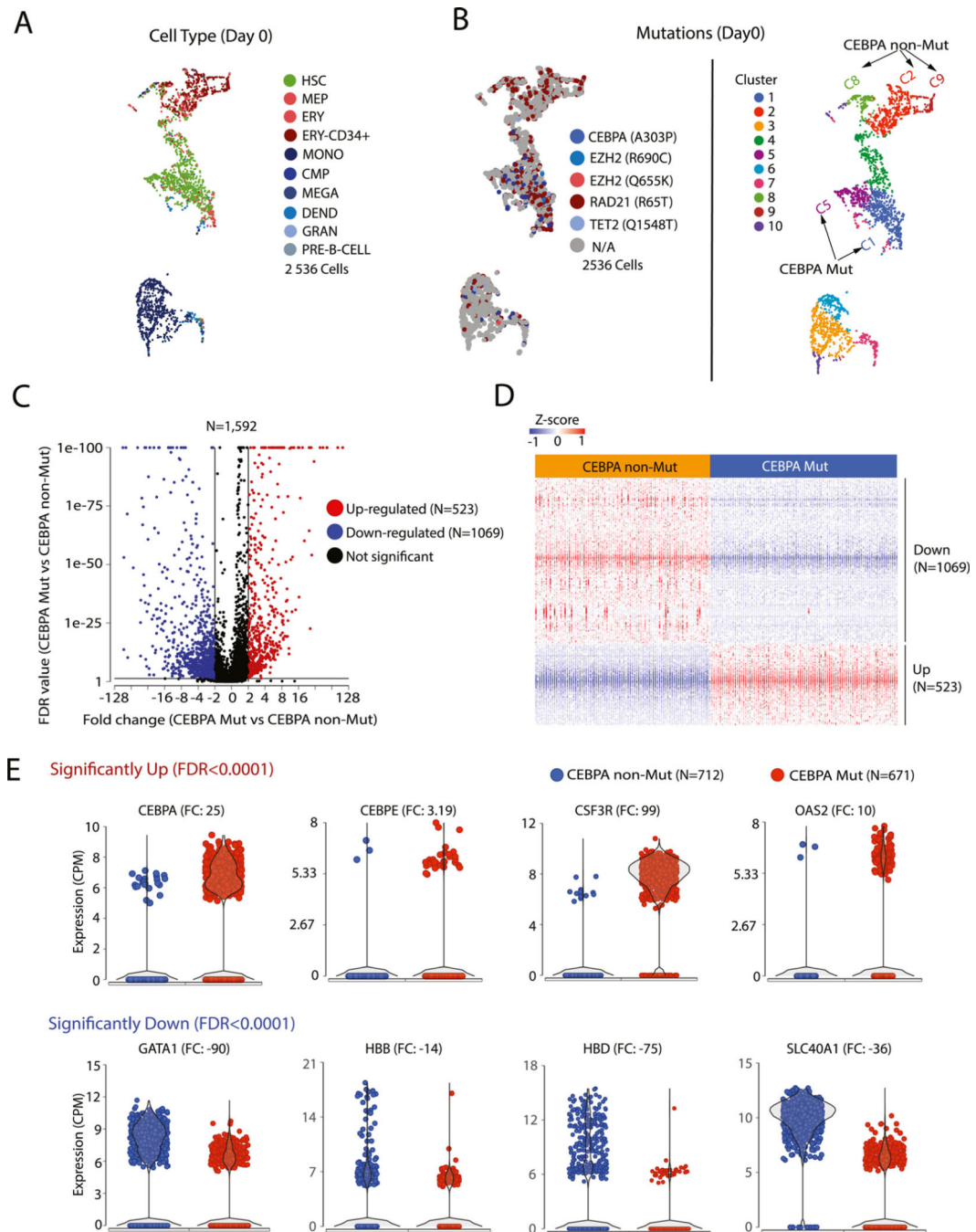
Figure 6.
A mutation-associated transcriptional signature was identified in 1 of 10 patients (CEBPA A303P in acute myeloid leukemia patient 1058). **(A,B)** Uniform Manifold Approximation and Projection of scRNAseq data from bone marrow cells derived from patient 1058, day 0, based on 10 principal component analysis consolidations and the 1,000 most variable genes in the data set. **(A)** Results colored by predicted hematopoietic population using DMAP gene expression profile. **(B)** Results colored by identified mutations and clusters using graph-based clustering. **(C)** Volcano plot revealing significantly differentially expressed

genes between days 0 and 10 (FDR 0.05 and FC −2, 2). **(D)** Heatmap of significantly differentially expressed genes in CEBPA-A303P mutant versus CEBPA wild-type clusters, highlighting the distribution of transcript expression within the clusters. **(E)** Violin plots showing CPM values for selected significantly differentially expressed genes in CEBPA (A303P) mutant versus nonmutant cell clusters. CPM=counts per minute; *FC*=fold change; *FDR*=false discovery rate.
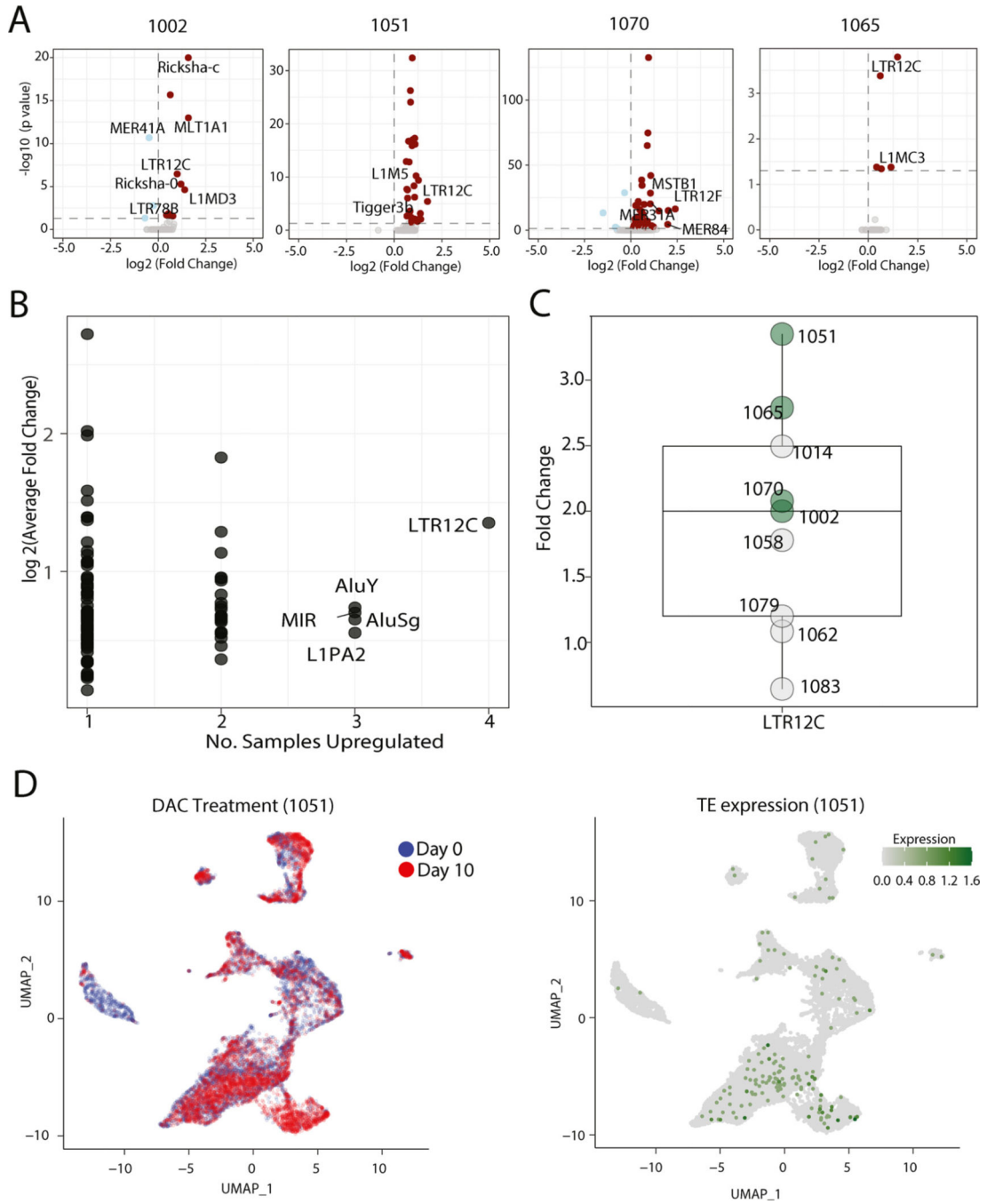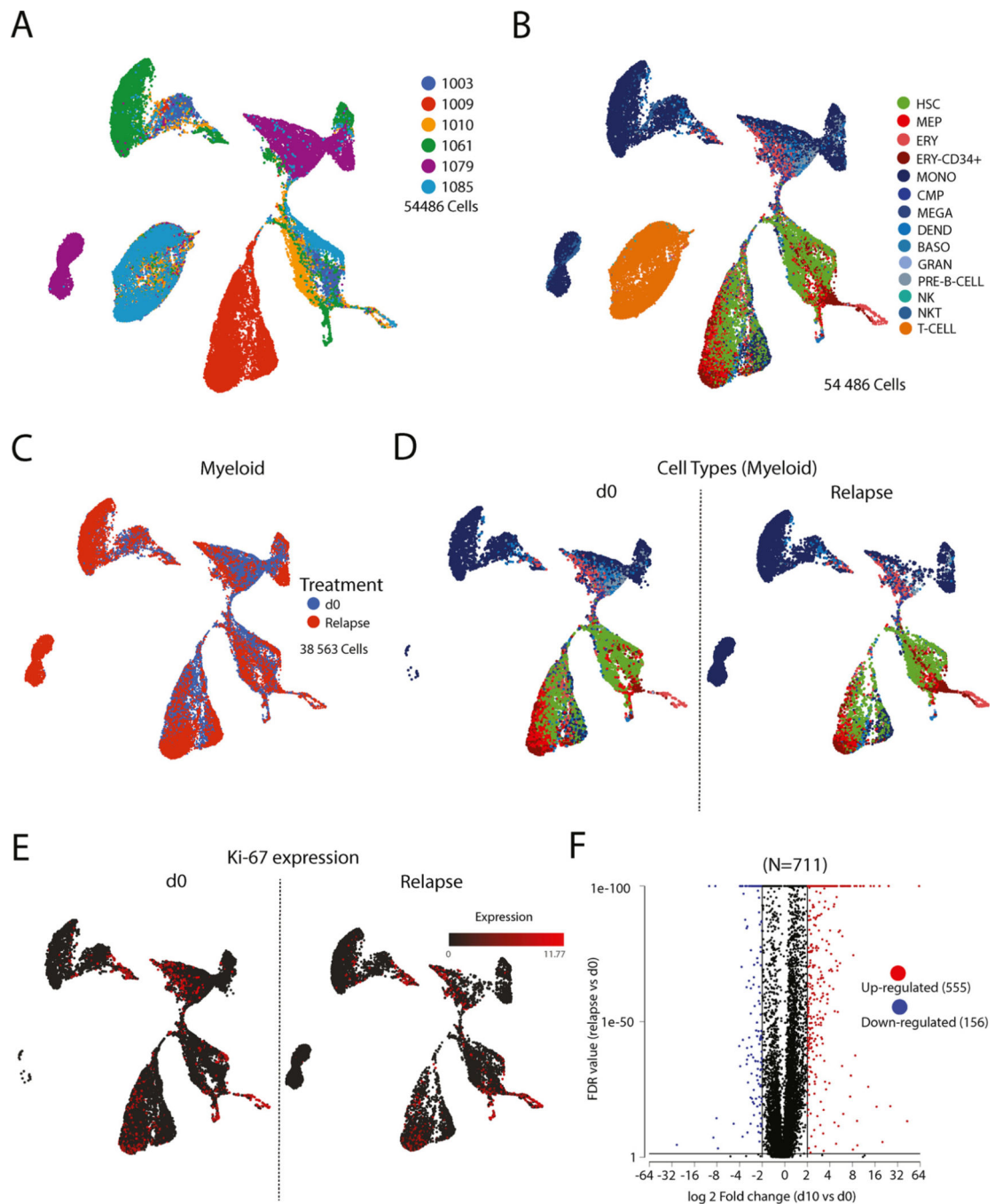
Figure 7.
Changes in human TE transcripts during decitabine treatment identified using single cell RNA sequencing data. **(A)** Volcano plots revealing significantly differentially expressed TEs of four patients who exhibited the largest transcriptional global changes in ERV and TE transcripts between days 0 and day 10 (adjusted *p* value 0.05 and FC -2, 2). **(B)** Plot showing recurrently upregulated TEs at day 10 post decitabine treatment. Selected TE families are highlighted. **(C)** Boxplot showing LTR12C family log$_2$ fold change expression levels at day 10 after decitabine treatment across all 10 MDS and AML patients (*green:*

significant case; *gray:* not significant case). **(D)** t-SNE projection of scRNAseq data from selected patient whole bone marrow samples revealing day 0 and day 10 as well as TE expression levels in individual cells. *AML*=acute myeloid leukemia; *ERV*=endogenous retrovirus; *FC*=fold change; *MDS*=myelodysplastic syndrome; *TE*=transposable element; *UMAP*=Uniform Manifold Approximation and Projection.

Figure 8.

Transcriptional alteration in the bone marrow of AML and MDS patients at relapse versus diagnosis identified using single-cell RNA sequencing. UMAP projection of scRNAseq data from all cells in MDS and AML patients based on 10 PCA consolidations and the 1000 most variable genes in the data set. (A) Data colored by the patient. (B) Data colored by DMAP cell types. (C) Selected myeloid cells (CD3E⁻ and DMAP myeloid lineage) in MDS and AML patients colored based on relapse and day 0. (D) Results were separated between day 0 and relapse colored by DMAP assigned cell types. (E) Data colored by Ki-67 expression

level. (E) Volcano plot highlighting differentially expressed genes within the combined data from six patients (FDR 0.05 and FC −2, 2). *AML*=acute myeloid leukemia; *DMAP*= Data Management and Access Plan; *FC*=fold change; *FDR*=false discovery rate; *MDS*=myelodysplastic syndrome; *UMAP*=Uniform Manifold Approximation and Projection.