Figure S1

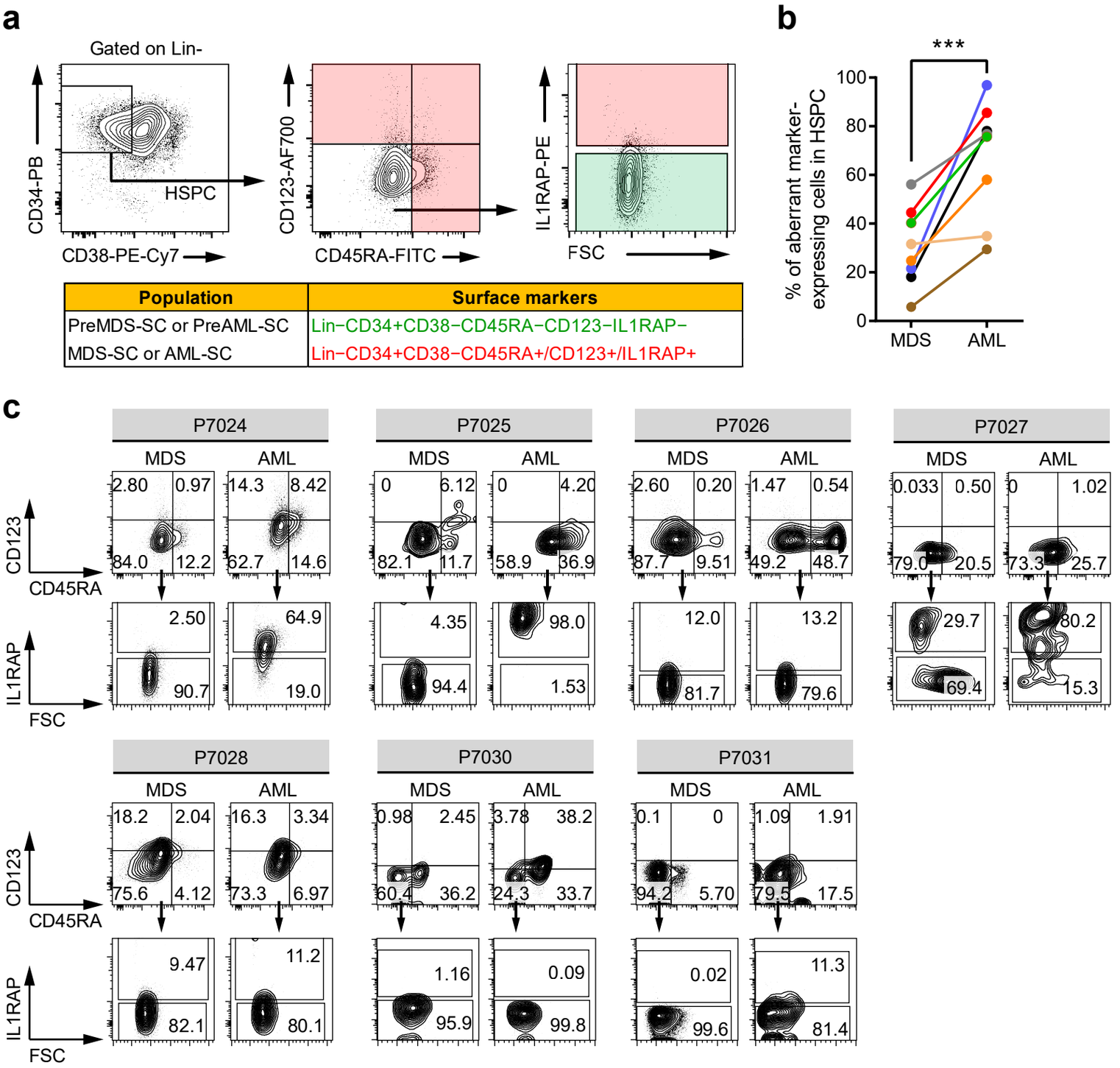

a
Gated on Lin-
CD34-PB
HSPC
CD38-PE-Cy7
CD123-AF700
CD45RA-FITC
IL1RAP-PE
FSC
Population
Surface markers
PreMDS-SC or PreAML-SC
Lin−CD34+CD38−CD45RA−CD123−IL1RAP−
MDS-SC or AML-SC
Lin−CD34+CD38−CD45RA+/CD123+/IL1RAP+
b
***
% of aberrant marker-expressing cells in HSPC
100
80
60
40
20
0
MDS
AML
c
P7024
P7025
P7026
P7027
P7028
P7030
P7031
MDS
AML
CD123
CD45RA
IL1RAP
FSC
2.80
0.97
84.0
12.2
14.3
8.42
62.7
14.6
0
6.12
82.1
11.7
0
4.20
58.9
36.9
2.60
0.20
87.7
9.51
1.47
0.54
49.2
48.7
0.033
0.50
79.0
20.5
0
1.02
73.3
25.7
2.50
90.7
64.9
19.0
4.35
94.4
98.0
1.53
12.0
81.7
13.2
79.6
29.7
69.4
80.2
15.3
18.2
2.04
75.6
4.12
16.3
3.34
73.3
6.97
0.98
2.45
60.4
36.2
3.78
38.2
24.3
33.7
0.1
0
94.2
5.70
1.09
1.91
79.5
17.5
9.47
82.1
11.2
80.1
1.16
95.9
0.09
99.8
0.02
99.6
11.3
81.4

Figure S2

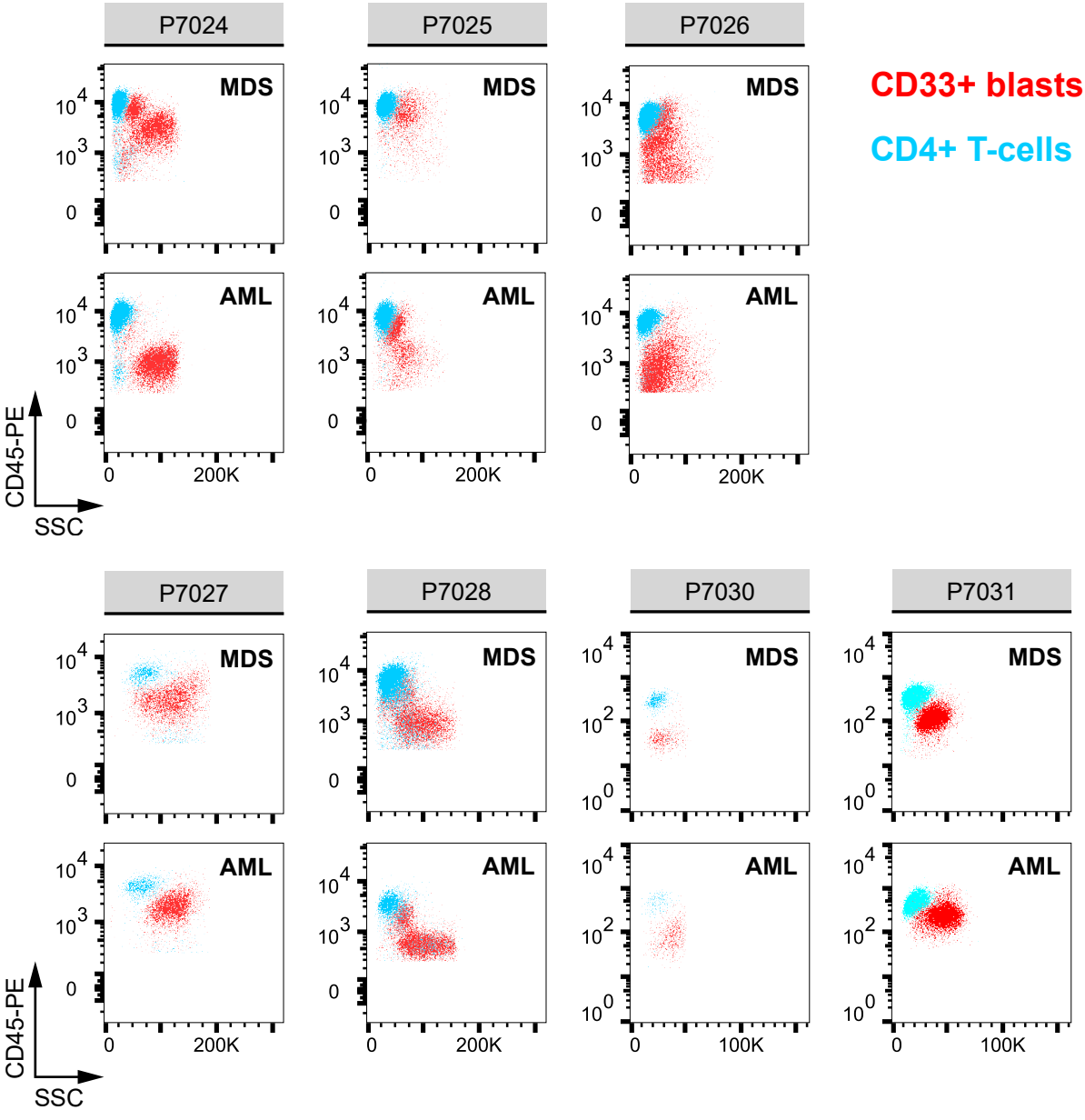
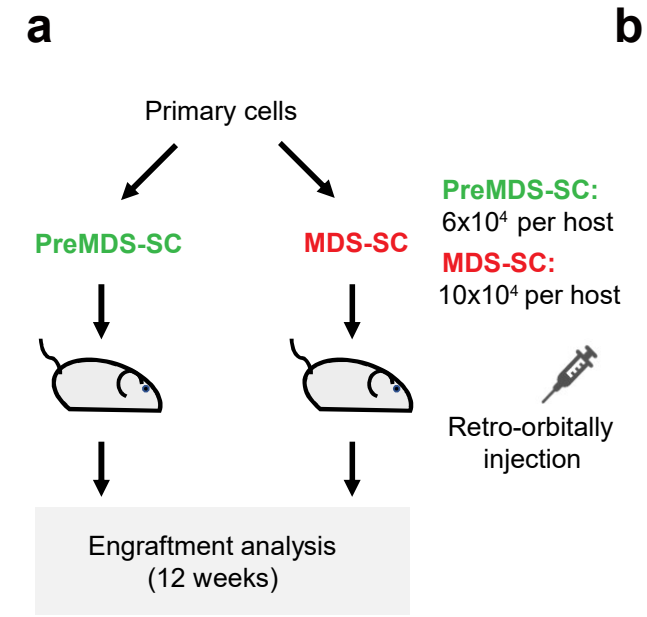
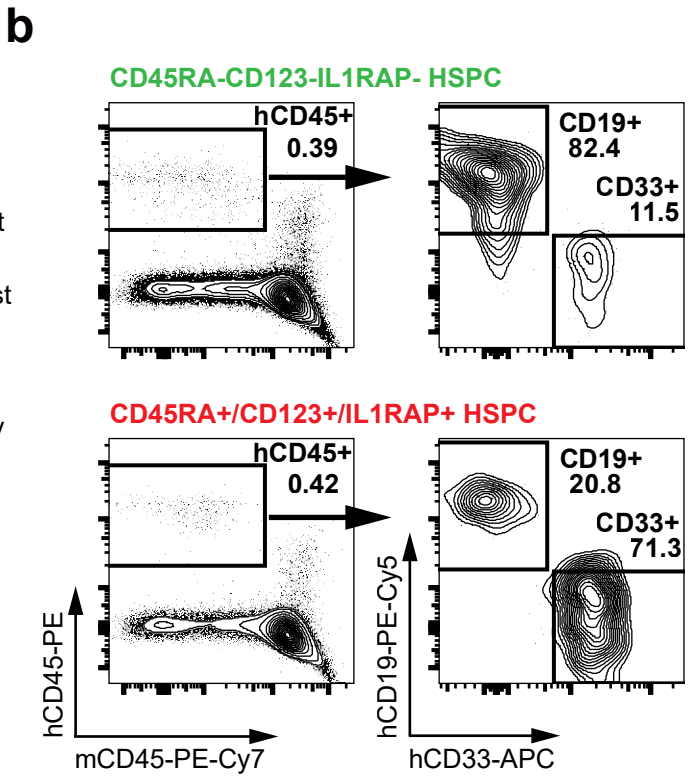
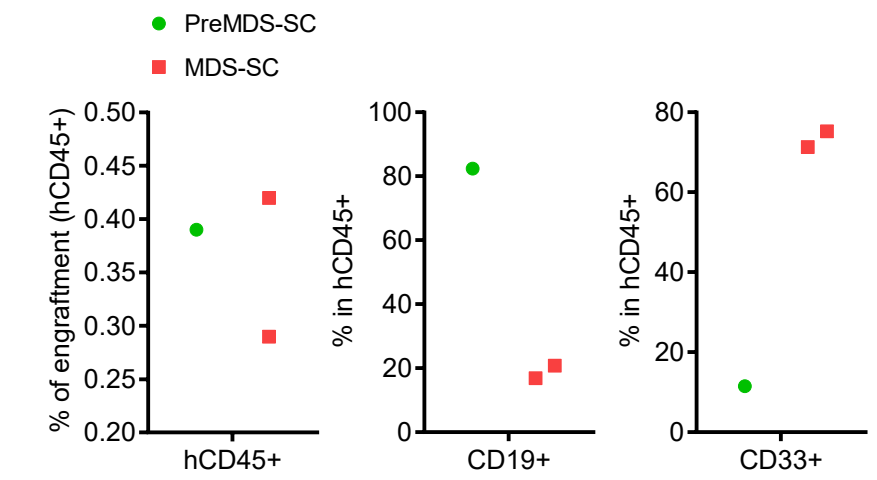
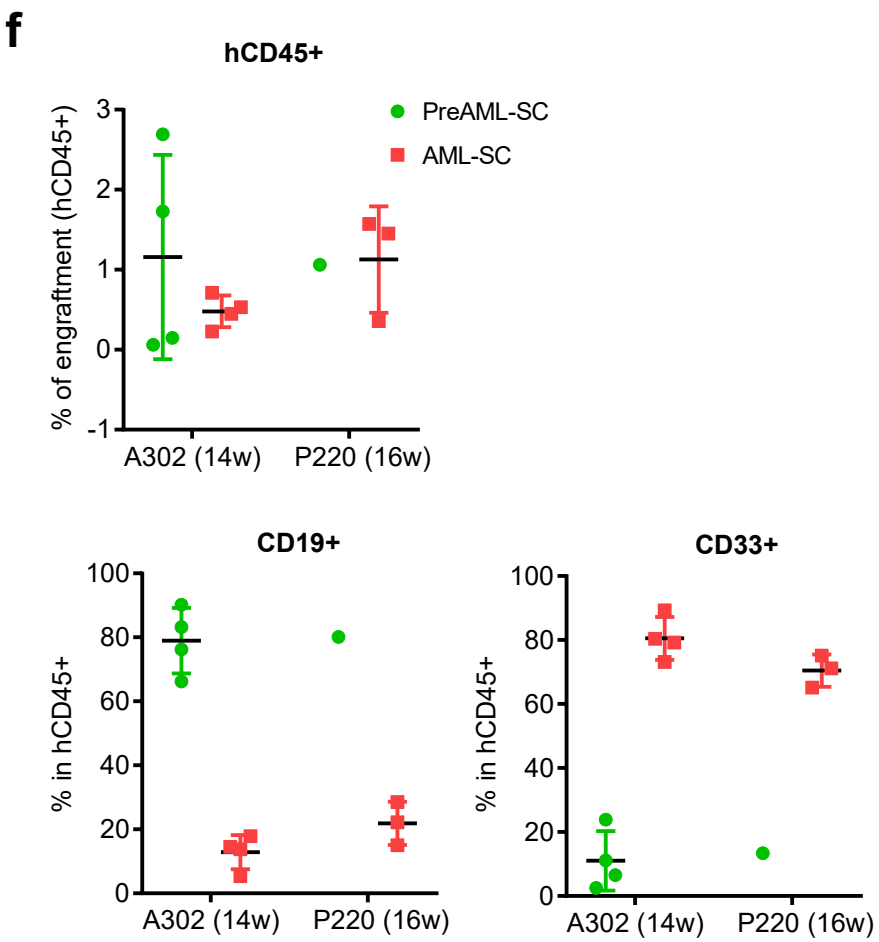

P7024
P7025
P7026
P7027
P7028
P7030
P7031
MDS
AML
CD33+ blasts
CD4+ T-cells
CD45-PE
SSC
0
200K
100K

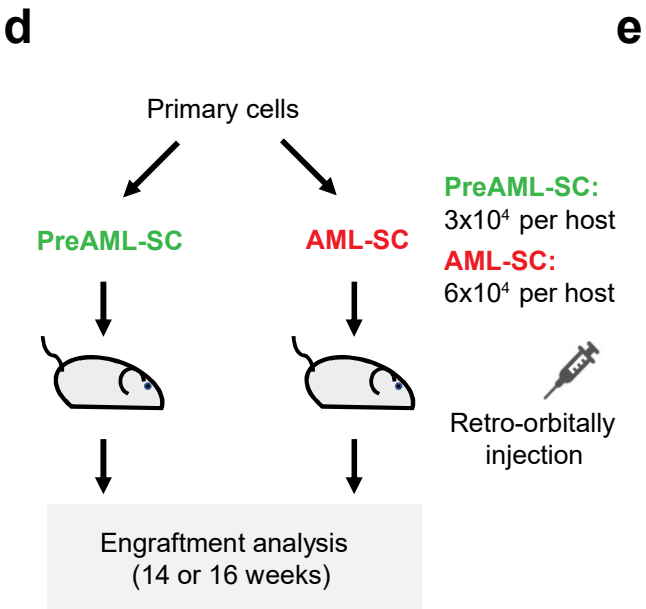
Figure S3


a
Primary cells
PreMDS-SC
MDS-SC
PreMDS-SC:
6x10⁴ per host
MDS-SC:
10x10⁴ per host
Retro-orbitally injection
Engraftment analysis (12 weeks)


b
CD45RA-CD123-IL1RAP- HSPC
hCD45+ 0.39
CD19+ 82.4
CD33+ 11.5
CD45RA+/CD123+/IL1RAP+ HSPC
hCD45+ 0.42
CD19+ 20.8
CD33+ 71.3
hCD45-PE
mCD45-PE-Cy7
hCD19-PE-Cy5
hCD33-APC


c
PreMDS-SC
MDS-SC
% of engraftment (hCD45+)
hCD45+
% in hCD45+
CD19+
CD33+


f
hCD45+
PreAML-SC
AML-SC
% of engraftment (hCD45+)
A302 (14w)
P220 (16w)
CD19+
CD33+
% in hCD45+


d
Primary cells
PreAML-SC
AML-SC
PreAML-SC:
3x10⁴ per host
AML-SC:
6x10⁴ per host
Retro-orbitally injection
Engraftment analysis (14 or 16 weeks)

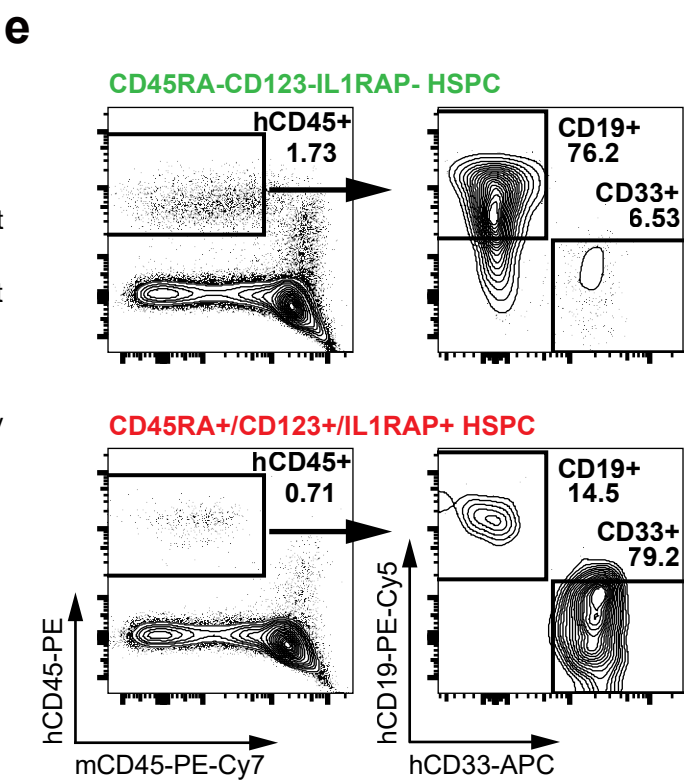

e
CD45RA-CD123-IL1RAP- HSPC
hCD45+ 1.73
CD19+ 76.2
CD33+ 6.53
CD45RA+/CD123+/IL1RAP+ HSPC
hCD45+ 0.71
CD19+ 14.5
CD33+ 79.2
hCD45-PE
mCD45-PE-Cy7
hCD19-PE-Cy5
hCD33-APC

Figure S4

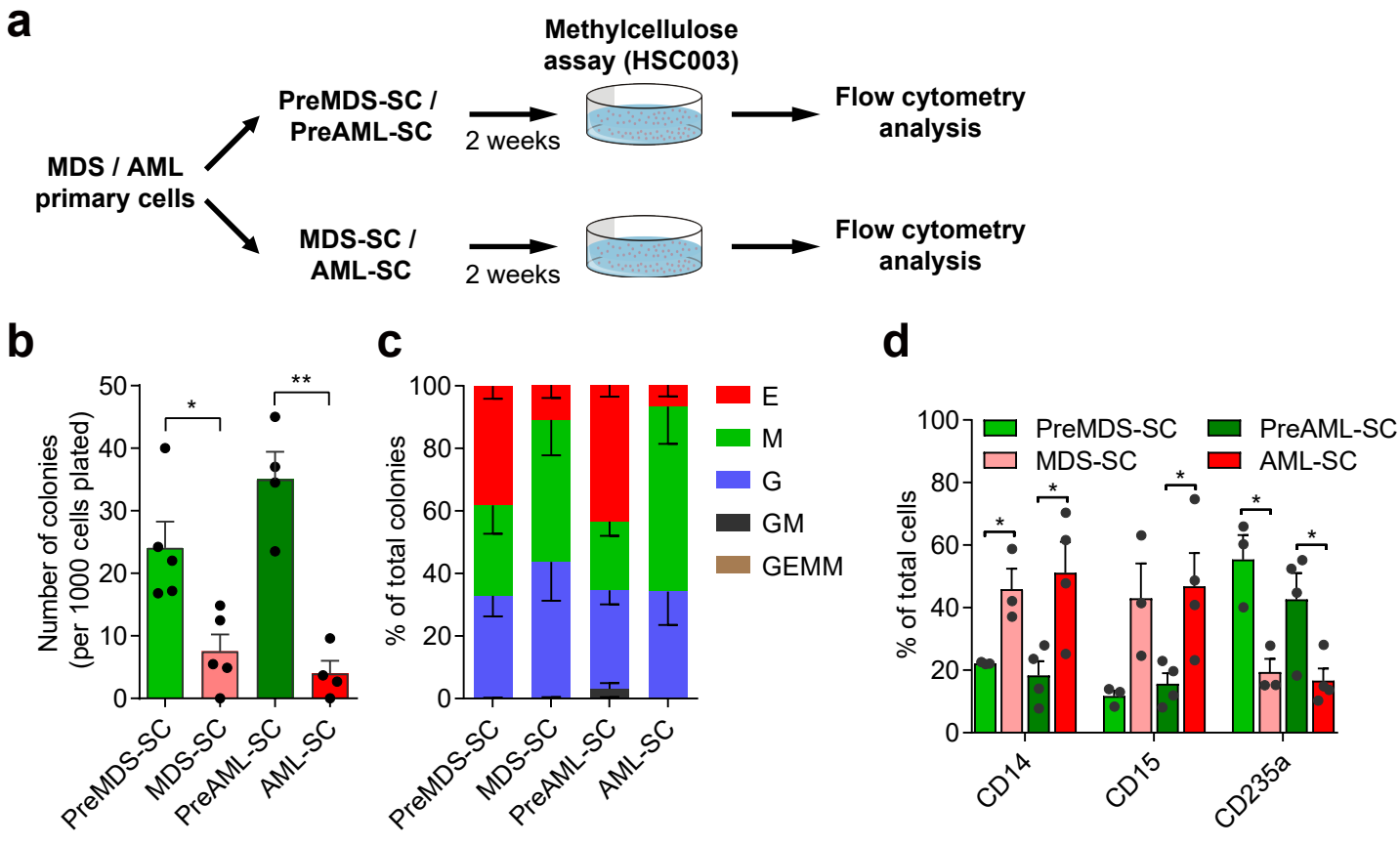

a
Methylcellulose assay (HSC003)
PreMDS-SC / PreAML-SC
2 weeks
Flow cytometry analysis
MDS / AML primary cells
MDS-SC / AML-SC
2 weeks
Flow cytometry analysis
b
Number of colonies (per 1000 cells plated)
PreMDS-SC
MDS-SC
PreAML-SC
AML-SC
c
% of total colonies
E
M
G
GM
GEMM
d
% of total cells
PreMDS-SC
MDS-SC
PreAML-SC
AML-SC
CD14
CD15
CD235a

Figure S5

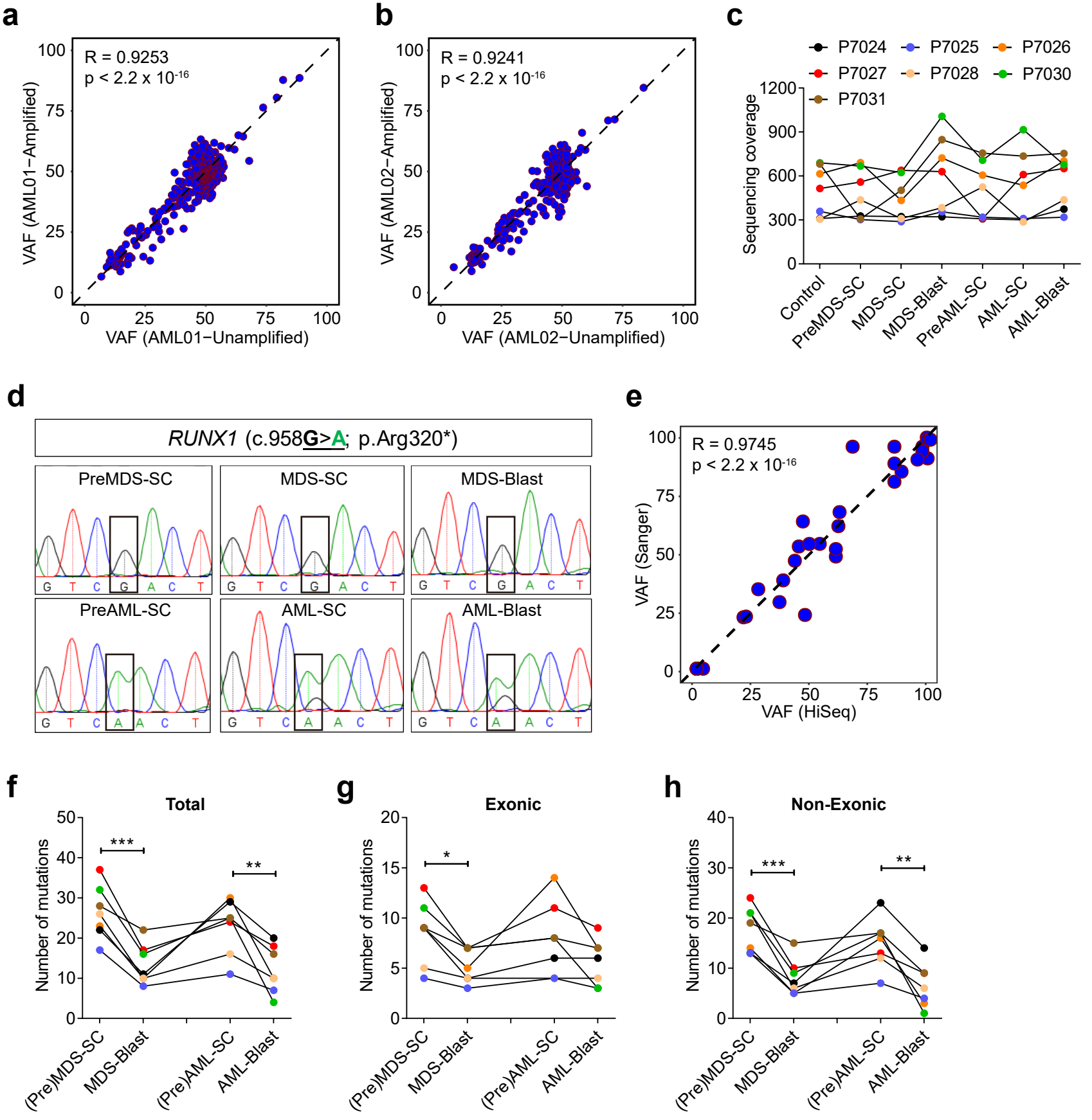
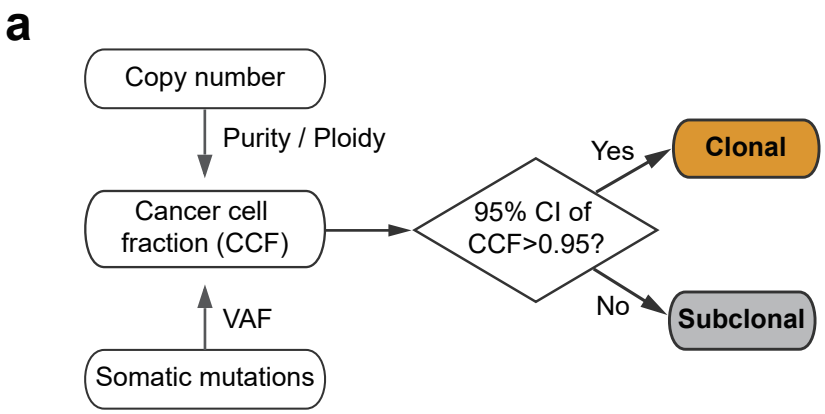

a
R = 0.9253
p < 2.2 x 10-16
VAF (AML01−Amplified)
VAF (AML01−Unamplified)
b
R = 0.9241
p < 2.2 x 10-16
VAF (AML02−Amplified)
VAF (AML02−Unamplified)
c
P7024
P7025
P7026
P7027
P7028
P7030
P7031
Sequencing coverage
Control
PreMDS-SC
MDS-SC
MDS-Blast
PreAML-SC
AML-SC
AML-Blast
d
RUNX1 (c.958G>A; p.Arg320*)
PreMDS-SC
MDS-SC
MDS-Blast
PreAML-SC
AML-SC
AML-Blast
e
R = 0.9745
p < 2.2 x 10-16
VAF (Sanger)
VAF (HiSeq)
f
Total
Number of mutations
***
**
(Pre)MDS-SC
MDS-Blast
(Pre)AML-SC
AML-Blast
g
Exonic
Number of mutations
*
(Pre)MDS-SC
MDS-Blast
(Pre)AML-SC
AML-Blast
h
Non-Exonic
Number of mutations
***
**
(Pre)MDS-SC
MDS-Blast
(Pre)AML-SC
AML-Blast

Figure S6


a
Copy number
Purity / Ploidy
Cancer cell fraction (CCF)
VAF
Somatic mutations
95% CI of CCF>0.95?
Yes
Clonal
No
Subclonal

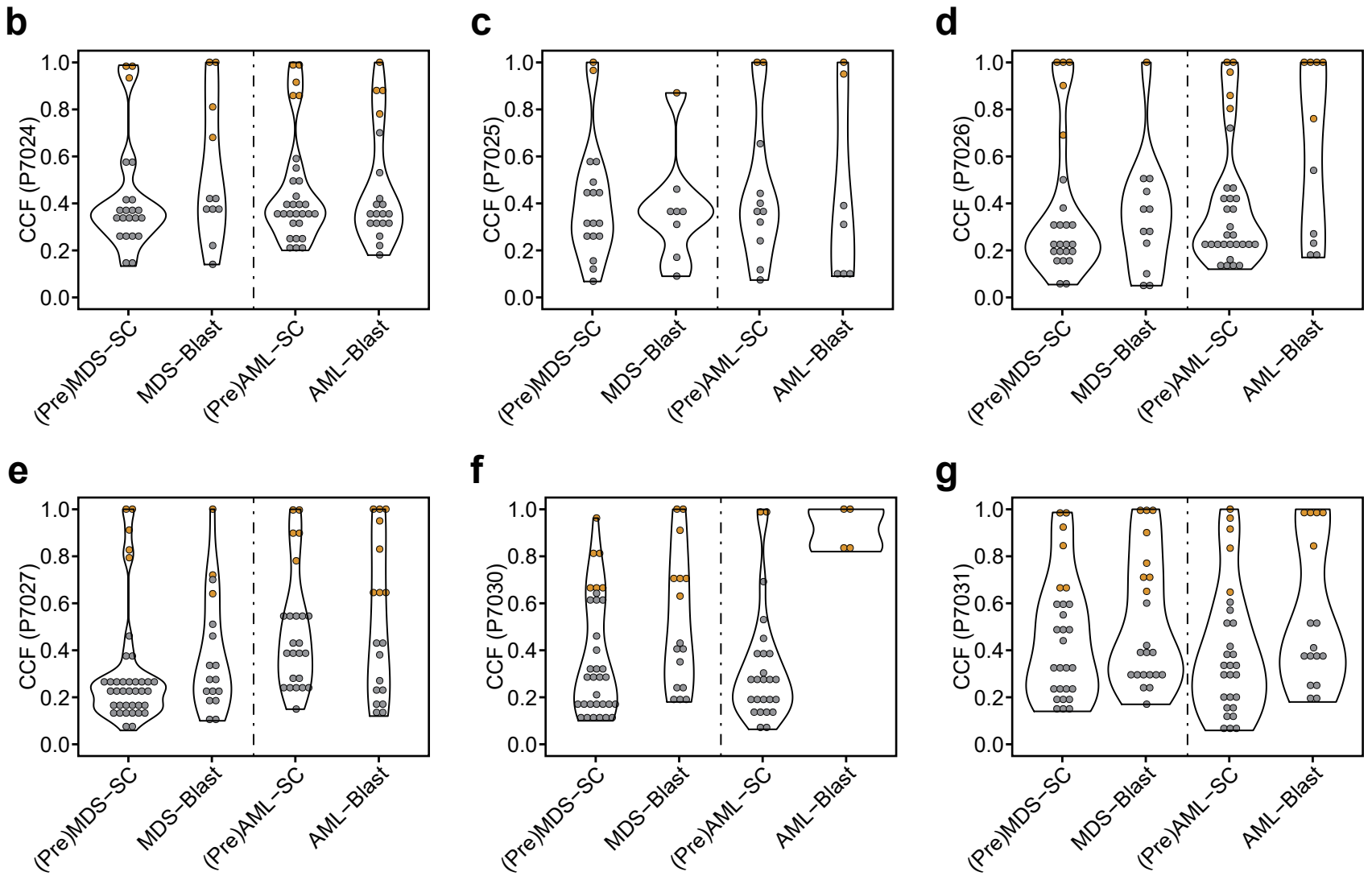

b
c
d
e
f
g
CCF (P7024)
CCF (P7025)
CCF (P7026)
CCF (P7027)
CCF (P7030)
CCF (P7031)
0.0
0.2
0.4
0.6
0.8
1.0
(Pre)MDS−SC
MDS−Blast
(Pre)AML−SC
AML−Blast

Figure S7

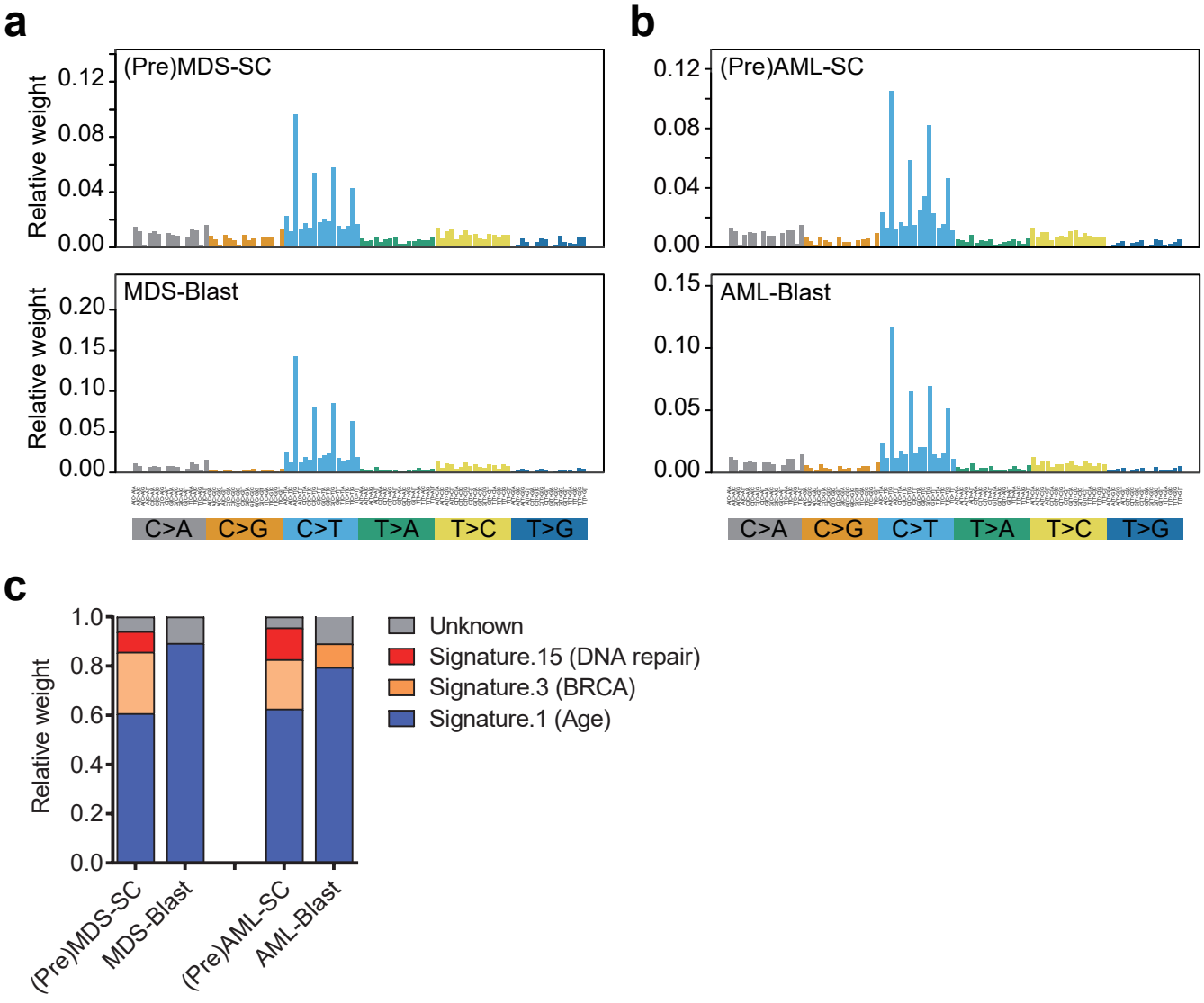

a
(Pre)MDS-SC
Relative weight
0.12
0.08
0.04
0.00
MDS-Blast
Relative weight
0.20
0.15
0.10
0.05
0.00
C>A
C>G
C>T
T>A
T>C
T>G
b
(Pre)AML-SC
0.12
0.08
0.04
0.00
AML-Blast
0.15
0.10
0.05
0.00
C>A
C>G
C>T
T>A
T>C
T>G
c
Relative weight
1.0
0.8
0.6
0.4
0.2
0.0
(Pre)MDS-SC
MDS-Blast
(Pre)AML-SC
AML-Blast
Unknown
Signature.15 (DNA repair)
Signature.3 (BRCA)
Signature.1 (Age)

Figure S8

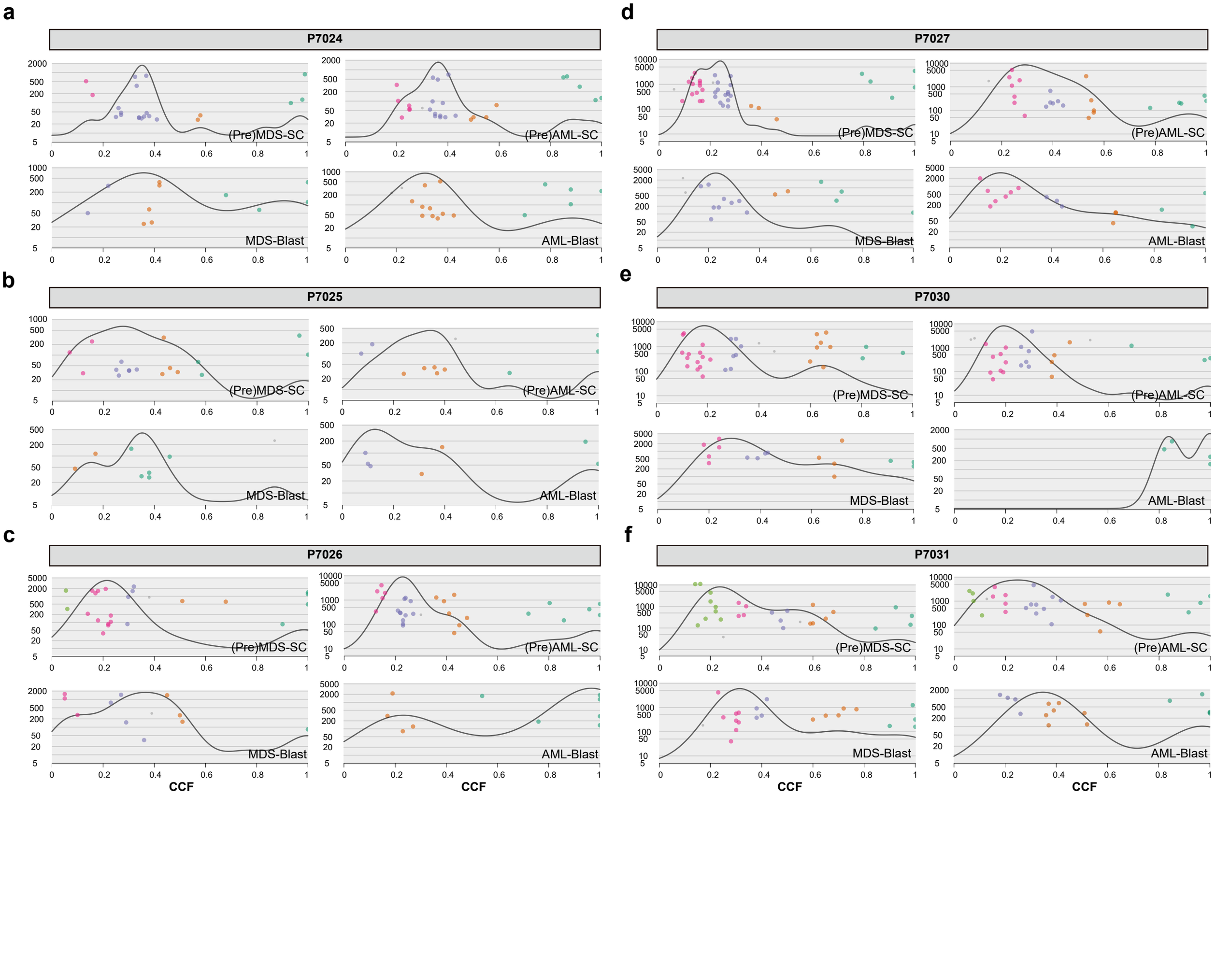

a
P7024
(Pre)MDS-SC
(Pre)AML-SC
MDS-Blast
AML-Blast
b
P7025
(Pre)MDS-SC
(Pre)AML-SC
MDS-Blast
AML-Blast
c
P7026
(Pre)MDS-SC
(Pre)AML-SC
MDS-Blast
AML-Blast
CCF
CCF
d
P7027
(Pre)MDS-SC
(Pre)AML-SC
MDS-Blast
AML-Blast
e
P7030
(Pre)MDS-SC
(Pre)AML-SC
MDS-Blast
AML-Blast
f
P7031
(Pre)MDS-SC
(Pre)AML-SC
MDS-Blast
AML-Blast
CCF
CCF

Figure S9

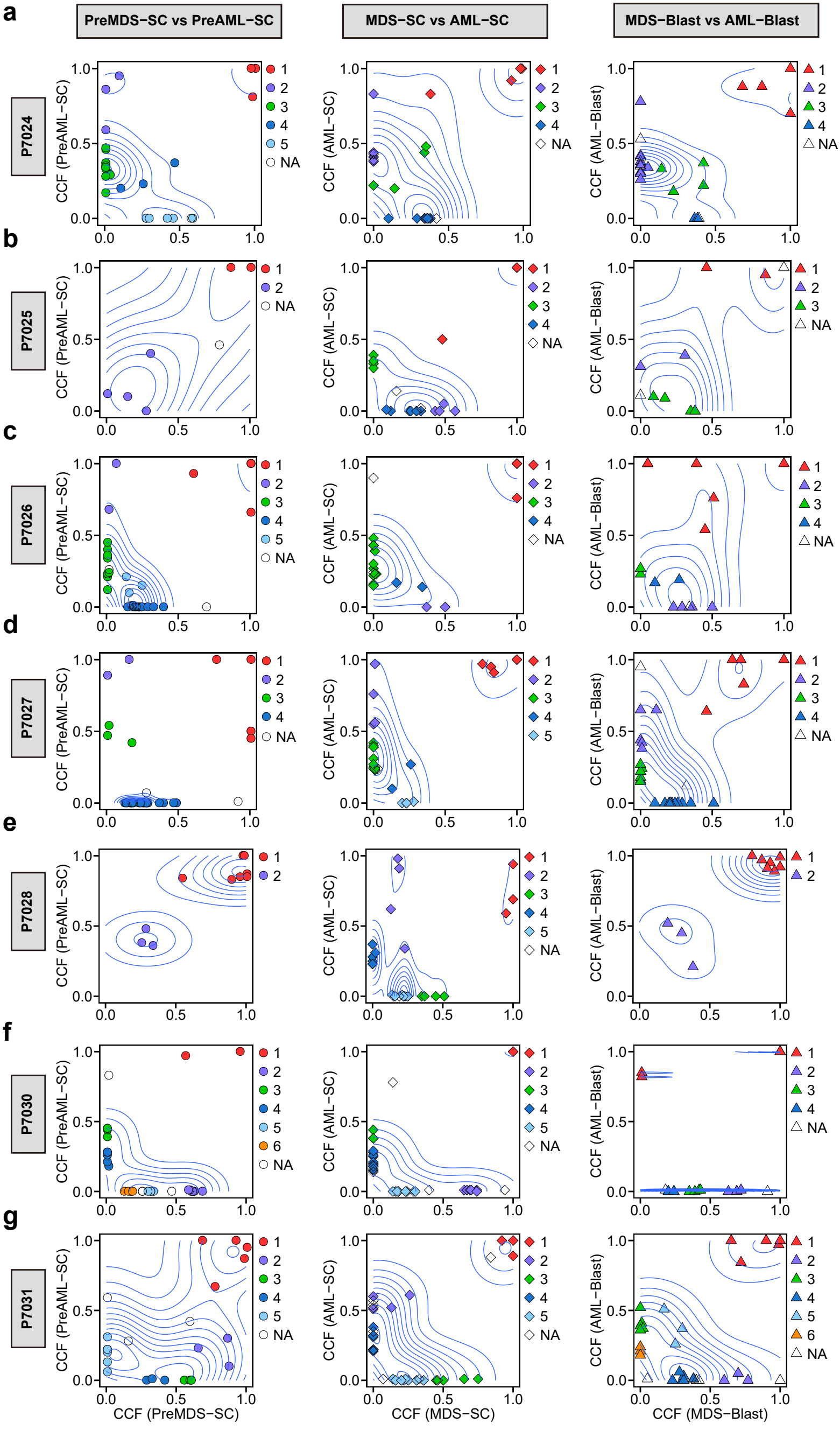

a
PreMDS−SC vs PreAML−SC
MDS−SC vs AML−SC
MDS−Blast vs AML−Blast
P7024
b
P7025
c
P7026
d
P7027
e
P7028
f
P7030
g
P7031
CCF (PreAML−SC)
CCF (AML−SC)
CCF (AML−Blast)
CCF (PreMDS−SC)
CCF (MDS−SC)
CCF (MDS−Blast)
NA

Figure S10

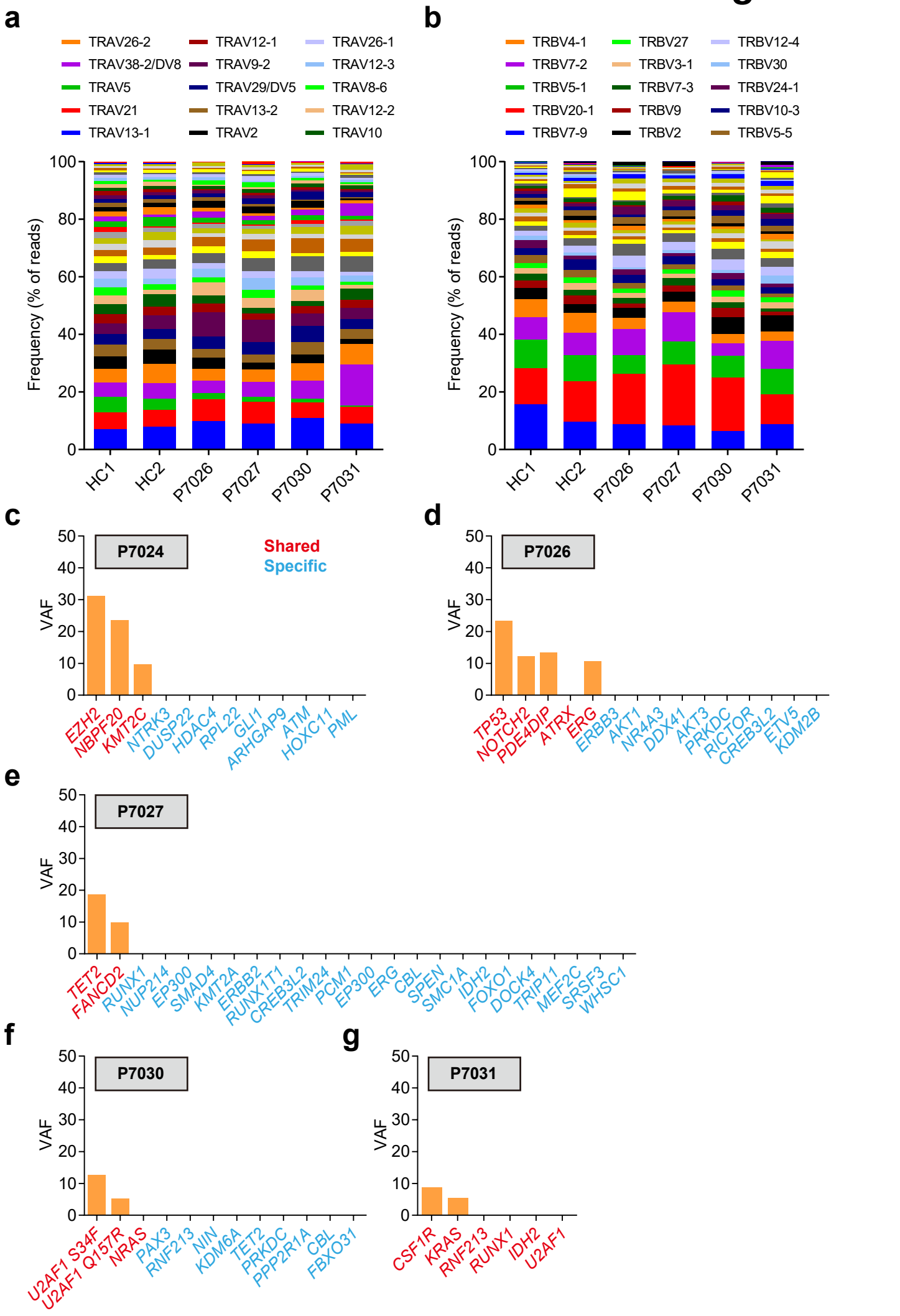

a
TRAV26-2
TRAV12-1
TRAV26-1
TRAV38-2/DV8
TRAV9-2
TRAV12-3
TRAV5
TRAV29/DV5
TRAV8-6
TRAV21
TRAV13-2
TRAV12-2
TRAV13-1
TRAV2
TRAV10
Frequency (% of reads)
HC1
HC2
P7026
P7027
P7030
P7031
b
TRBV4-1
TRBV27
TRBV12-4
TRBV7-2
TRBV3-1
TRBV30
TRBV5-1
TRBV7-3
TRBV24-1
TRBV20-1
TRBV9
TRBV10-3
TRBV7-9
TRBV2
TRBV5-5
c
P7024
Shared
Specific
VAF
EZH2
NBPF20
KMT2C
NTRK3
DUSP22
HDAC4
RPL22
GLI1
ARHGAP9
ATM
HOXC11
PML
d
P7026
TP53
NOTCH2
PDE4DIP
ATRX
ERG
ERBB3
AKT1
NR4A3
DDX41
AKT3
PRKDC
RICTOR
CREB3L2
ETV5
KDM2B
e
P7027
TET2
FANCD2
RUNX1
NUP214
EP300
SMAD4
KMT2A
ERBB2
RUNX1T1
CREB3L2
TRIM24
PCM1
EP300
ERG
CBL
SPEN
SMC1A
IDH2
FOXO1
DOCK4
TRIP11
MEF2C
SRSF3
WHSC1
f
P7030
U2AF1 S34F
U2AF1 Q157R
NRAS
PAX3
RNF213
NIN
KDM6A
TET2
PRKDC
PPP2R1A
CBL
FBXO31
g
P7031
CSF1R
KRAS
RNF213
RUNX1
IDH2
U2AF1

Figure S11

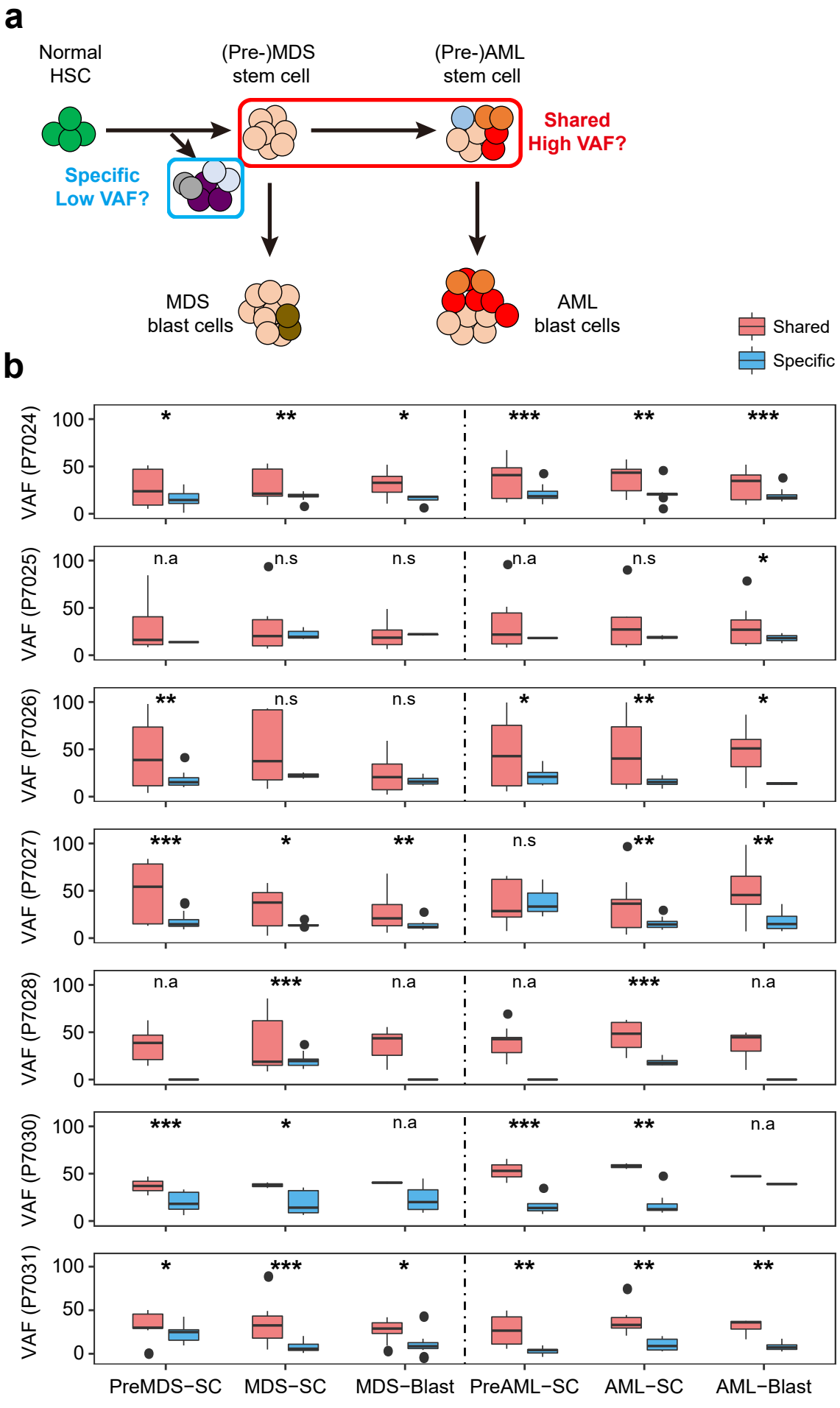
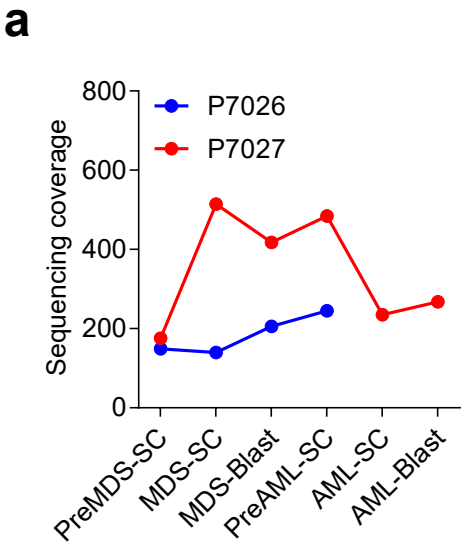
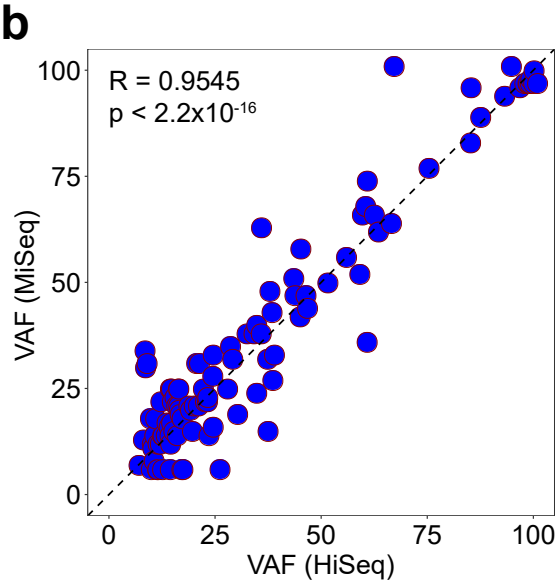

a
Normal HSC
(Pre-)MDS stem cell
(Pre-)AML stem cell
Shared High VAF?
Specific Low VAF?
MDS blast cells
AML blast cells
b
Shared
Specific
VAF (P7024)
VAF (P7025)
VAF (P7026)
VAF (P7027)
VAF (P7028)
VAF (P7030)
VAF (P7031)
100
50
0
n.a
n.s
PreMDS−SC
MDS−SC
MDS−Blast
PreAML−SC
AML−SC
AML−Blast

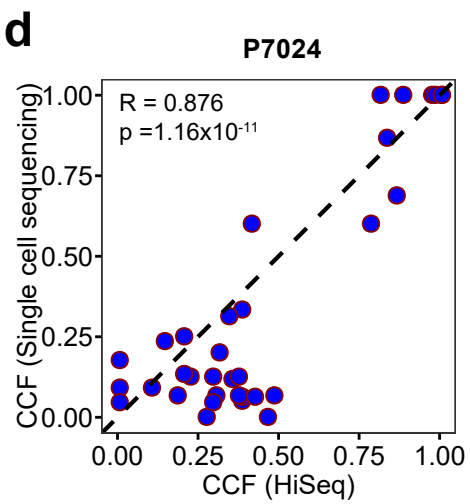
Figure S12

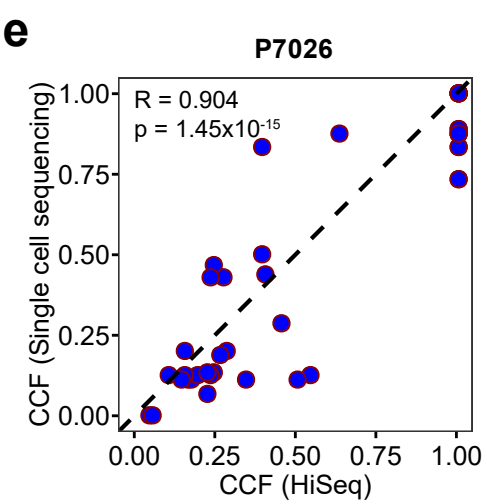


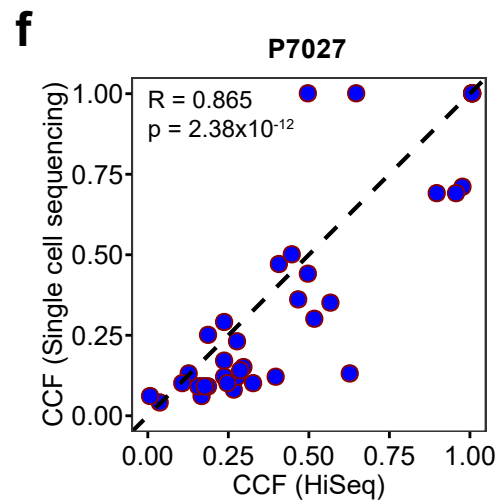


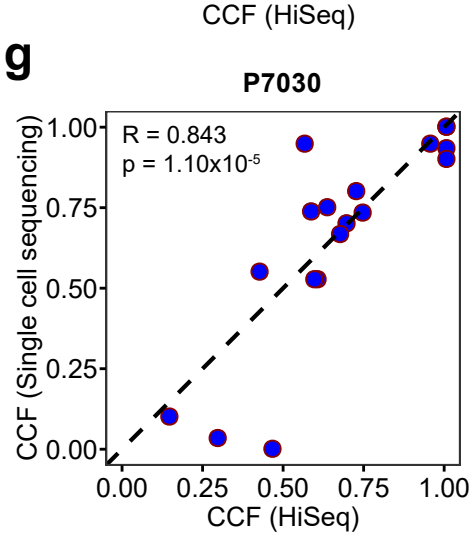
c

	P7024		P7026		P7027		P7030		P7031	
Population	Total	Pass	Total	Pass	Total	Pass	Total	Pass	Total	Pass
PreMDS-SC	28	26	24	19	30	27	26	23	22	22
MDS-SC	28	21	23	21	37	30	36	32	23	22
MDS-Blast	25	20	21	17	29	27	28	22	23	21
PreAML-SC	23	19	23	19	23	21	-	-	-	-
AML-SC	22	18	25	20	25	20	-	-	-	-
AML-Blast	21	18	21	18	23	19	-	-	-	-









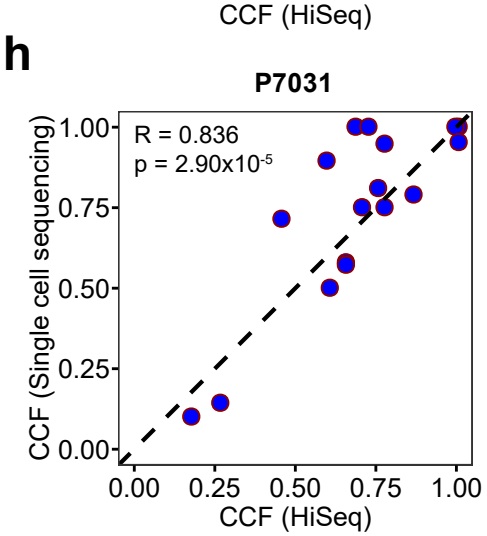


Figure S13

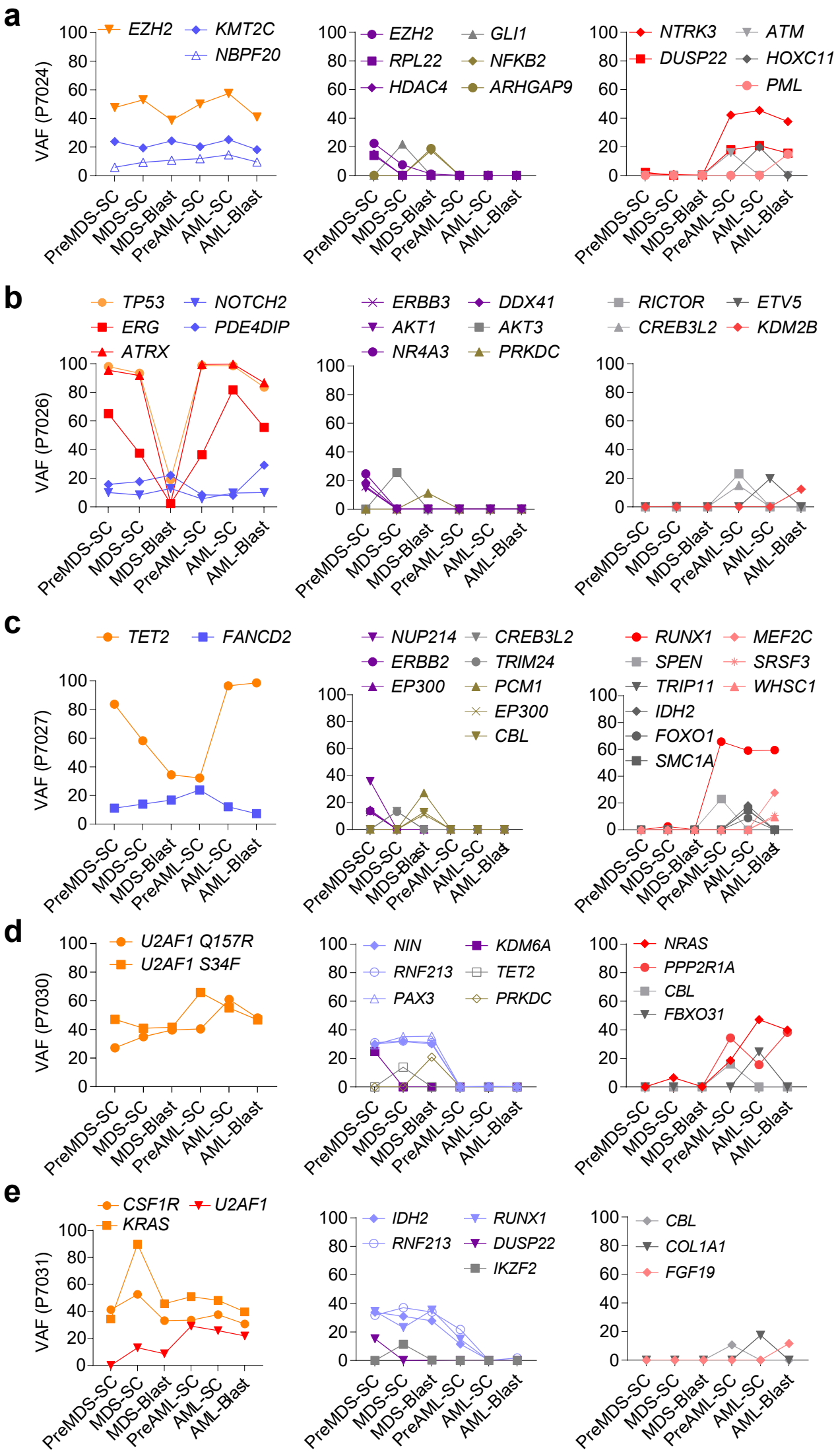

a
VAF (P7024)
EZH2
KMT2C
NBPF20
EZH2
GLI1
RPL22
NFKB2
HDAC4
ARHGAP9
NTRK3
ATM
DUSP22
HOXC11
PML
b
VAF (P7026)
TP53
NOTCH2
ERG
PDE4DIP
ATRX
ERBB3
DDX41
AKT1
AKT3
NR4A3
PRKDC
RICTOR
ETV5
CREB3L2
KDM2B
c
VAF (P7027)
TET2
FANCD2
NUP214
CREB3L2
ERBB2
TRIM24
EP300
PCM1
EP300
CBL
RUNX1
MEF2C
SPEN
SRSF3
TRIP11
WHSC1
IDH2
FOXO1
SMC1A
d
VAF (P7030)
U2AF1 Q157R
U2AF1 S34F
NIN
KDM6A
RNF213
TET2
PAX3
PRKDC
NRAS
PPP2R1A
CBL
FBXO31
e
VAF (P7031)
CSF1R
U2AF1
KRAS
IDH2
RUNX1
RNF213
DUSP22
IKZF2
CBL
COL1A1
FGF19
0
20
40
60
80
100
PreMDS-SC
MDS-SC
MDS-Blast
PreAML-SC
AML-SC
AML-Blast

Figure S14

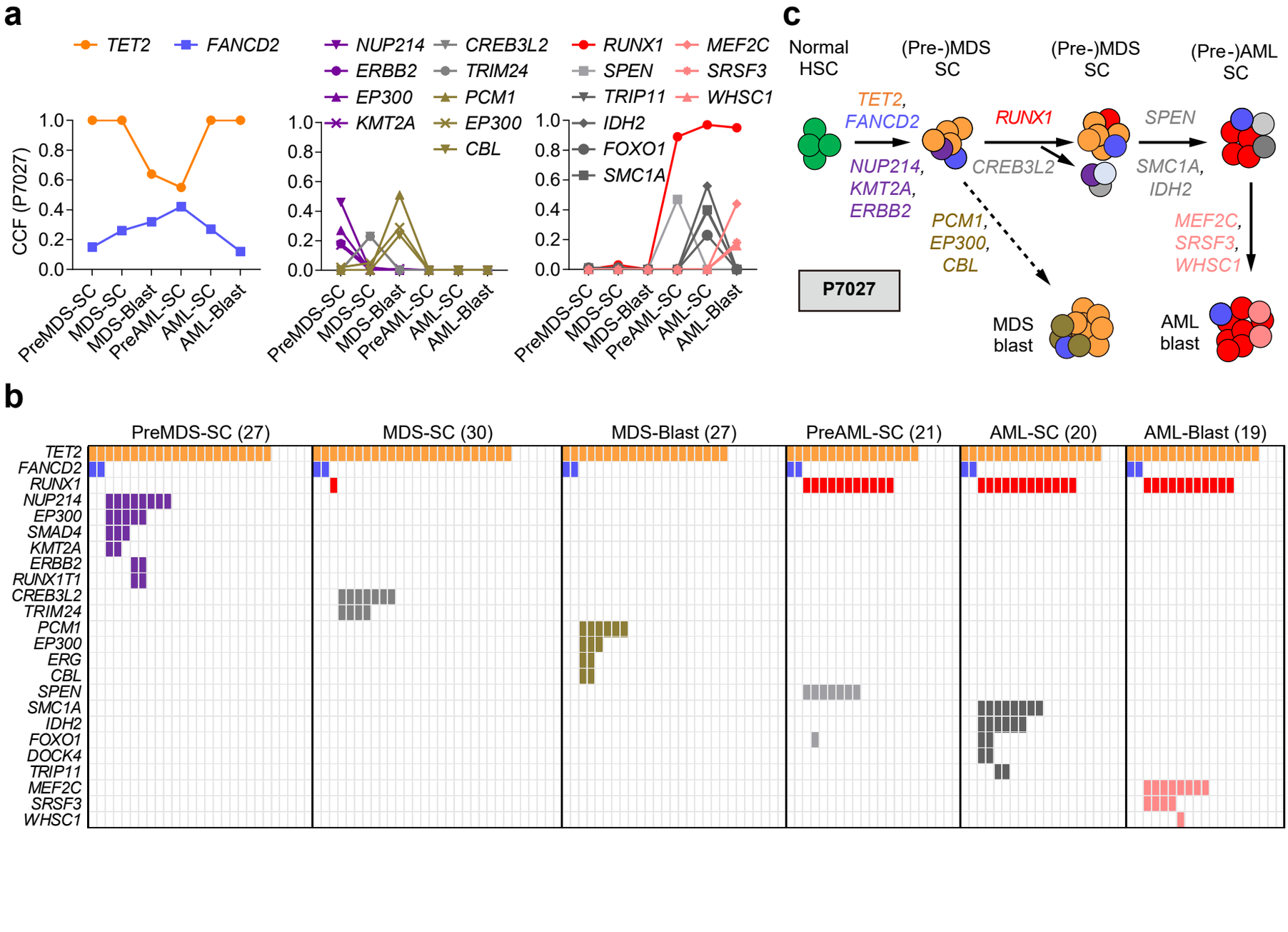
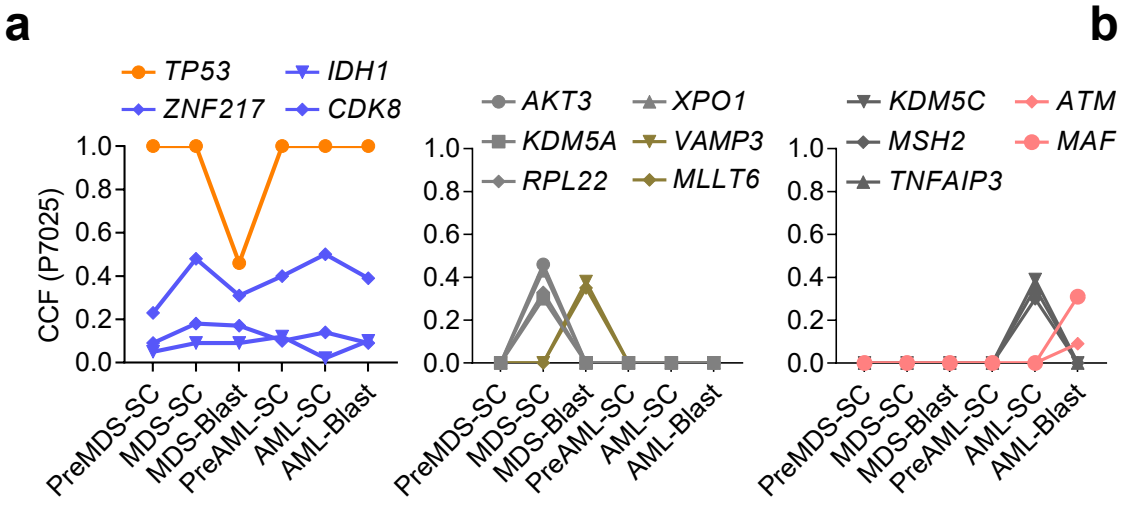
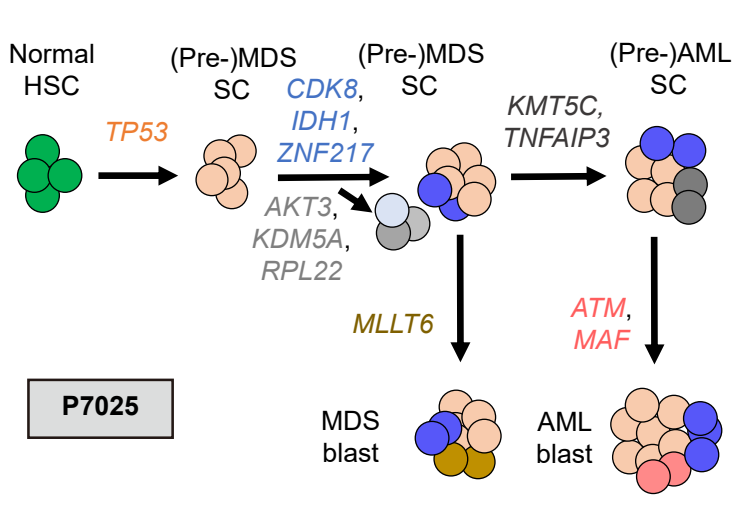
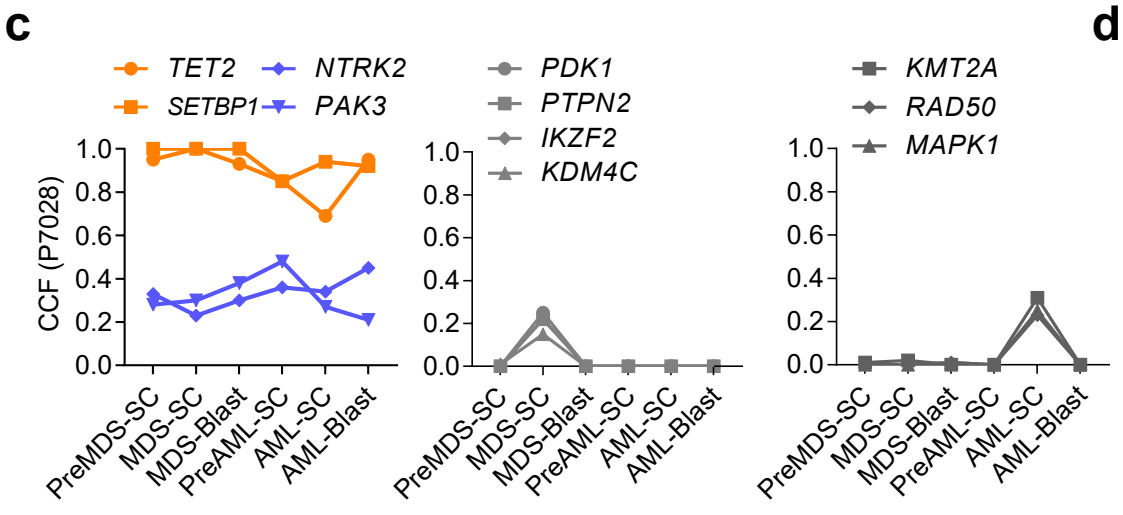
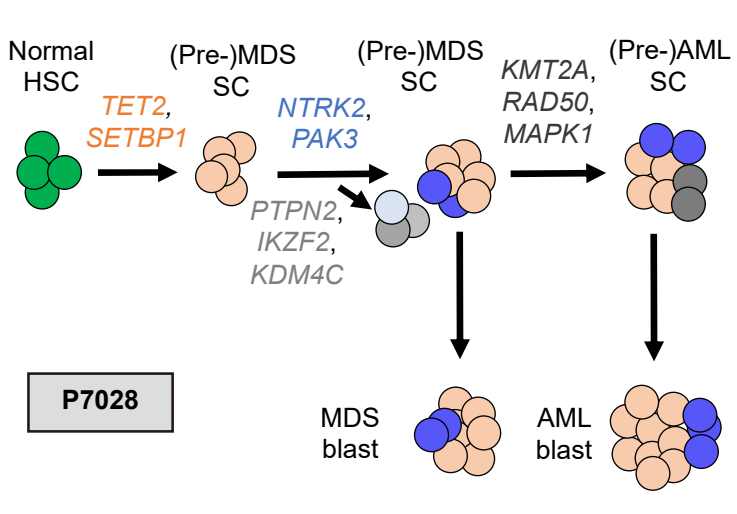

a
TET2
FANCD2
NUP214
CREB3L2
ERBB2
TRIM24
EP300
PCM1
KMT2A
EP300
CBL
RUNX1
MEF2C
SPEN
SRSF3
TRIP11
WHSC1
IDH2
FOXO1
SMC1A
CCF (P7027)
1.0
0.8
0.6
0.4
0.2
0.0
PreMDS-SC
MDS-SC
MDS-Blast
PreAML-SC
AML-SC
AML-Blast
c
Normal HSC
(Pre-)MDS SC
(Pre-)MDS SC
(Pre-)AML SC
TET2, FANCD2
NUP214, KMT2A, ERBB2
RUNX1
CREB3L2
SPEN
SMC1A, IDH2
PCM1, EP300, CBL
MEF2C, SRSF3, WHSC1
P7027
MDS blast
AML blast
b
PreMDS-SC (27)
MDS-SC (30)
MDS-Blast (27)
PreAML-SC (21)
AML-SC (20)
AML-Blast (19)
TET2
FANCD2
RUNX1
NUP214
EP300
SMAD4
KMT2A
ERBB2
RUNX1T1
CREB3L2
TRIM24
PCM1
EP300
ERG
CBL
SPEN
SMC1A
IDH2
FOXO1
DOCK4
TRIP11
MEF2C
SRSF3
WHSC1

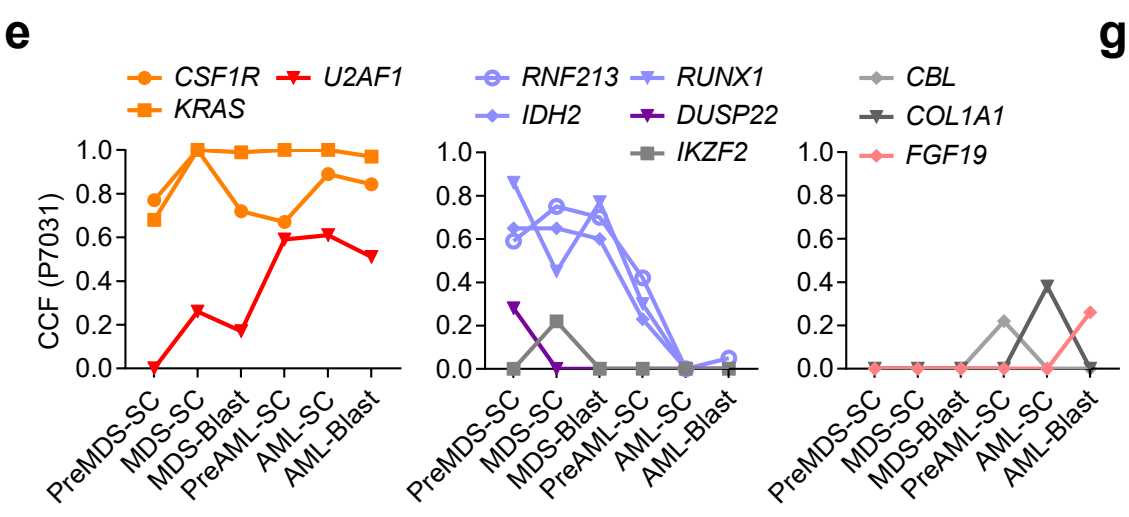
Figure S15

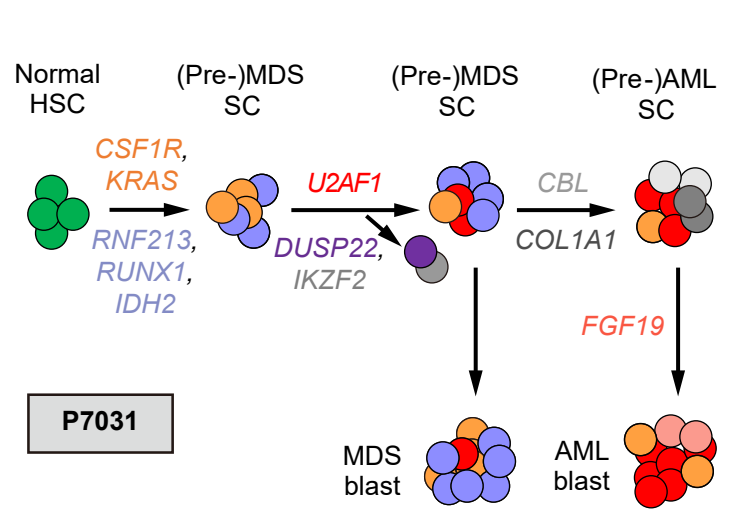

a
TP53
IDH1
ZNF217
CDK8
AKT3
XPO1
KDM5A
VAMP3
RPL22
MLLT6
KDM5C
ATM
MSH2
MAF
TNFAIP3
CCF (P7025)
1.0
0.8
0.6
0.4
0.2
0.0
PreMDS-SC
MDS-SC
MDS-Blast
PreAML-SC
AML-SC
AML-Blast

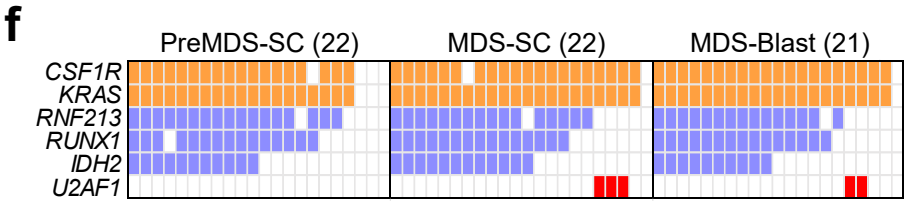

b
Normal HSC
(Pre-)MDS SC
(Pre-)MDS SC
(Pre-)AML SC
TP53
CDK8, IDH1, ZNF217
AKT3, KDM5A, RPL22
KMT5C, TNFAIP3
MLLT6
ATM, MAF
P7025
MDS blast
AML blast


c
TET2
NTRK2
SETBP1
PAK3
PDK1
PTPN2
IKZF2
KDM4C
KMT2A
RAD50
MAPK1
CCF (P7028)
1.0
0.8
0.6
0.4
0.2
0.0
PreMDS-SC
MDS-SC
MDS-Blast
PreAML-SC
AML-SC
AML-Blast


d
Normal HSC
(Pre-)MDS SC
(Pre-)MDS SC
(Pre-)AML SC
TET2, SETBP1
NTRK2, PAK3
PTPN2, IKZF2, KDM4C
KMT2A, RAD50, MAPK1
P7028
MDS blast
AML blast


e
CSF1R
U2AF1
KRAS
RNF213
RUNX1
IDH2
DUSP22
IKZF2
CBL
COL1A1
FGF19
CCF (P7031)
1.0
0.8
0.6
0.4
0.2
0.0
PreMDS-SC
MDS-SC
MDS-Blast
PreAML-SC
AML-SC
AML-Blast


g
Normal HSC
(Pre-)MDS SC
(Pre-)MDS SC
(Pre-)AML SC
CSF1R, KRAS
RNF213, RUNX1, IDH2
U2AF1
DUSP22, IKZF2
CBL
COL1A1
FGF19
P7031
MDS blast
AML blast


f
PreMDS-SC (22)
MDS-SC (22)
MDS-Blast (21)
CSF1R
KRAS
RNF213
RUNX1
IDH2
U2AF1

Supplementary Figure Legends

Fig. S1. Isolation of stem cells from individual human patients with MDS and matched sAML for targeted deep sequencing and single cell sequencing.

a, FACS strategy to isolate preMDS-SC and preAML-SC (green; CD45RA−CD123−IL1RAP− HSPC) and MDS-SC and AML-SC (red; CD45RA+ and/or CD123+ and/or IL1RAP+ HSPC) from longitudinal bone marrow samples of individual patients with MDS who progressed to secondary AML (sAML). **b,** Frequency of stem cells expressing aberrant markers at the MDS and matched sAML stages of individual patients (n=8). ***$p < 0.001$ ($p = 0.0006$; two-tailed paired Student's t test). **c,** FACS plots comparing the expression of CD45RA/CD123/IL1RAP markers in MDS and matched sAML samples from each patient.

Fig. S2. CD45/side scatter gating of blasts and T cells in MDS and matched sAML samples.

Analysis of CD45 versus side scatter (SSC) profiles of blasts and T cells sorted in the studied MDS and matched sAML samples. CD33+ blasts (red) and T cells (blue) were gated in a CD45/SSC bi-dimensional plot.

Fig. S3. Xenotransplantation assays of sorted pre-malignant and malignant stem cells.

a, Schema of xenotransplantation assay. Sorted preMDS-SC and MDS-SC were retro-orbitally transplanted into NBSGW immunocompromised mice (aged 6-8 weeks). PreMDS-SC and MDS-SC were sorted from fresh samples from independent MDS patient based on the same strategy used in the patients sequenced (PreMDS-SC: HSPC (Lin−CD34+CD38−) that were triple-negative for CD45RA, CD123, and IL1RAP; MDS-SC: HSPC expressing at least one (or more) of the markers). Engraftment analysis was performed within the bone marrow of recipient mice 12 weeks after transplantation. **b,** Representative FACS plots of engraftment analysis. The presence of human CD45+ leukocyte engraftment was analyzed by flow cytometry, and the engrafted cells were further examined for the fraction of CD33+ myeloid and

CD19+ lymphoid cells. The number of recipient mice is shown in **c** (each data point represents an individual mouse). **c,** Quantification of engraftment analysis in **b**. **d,** Schema of xenotransplantation assay with sorted preAML-SC and AML-SC. PreAML-SC and AML-SC were sorted from fresh samples from independent AML patients based on the same strategy used in the patients sequenced (PreAML-SC: HSPC (Lin−CD34+CD38−) that were triple-negative for CD45RA, CD123, and IL1RAP; AML-SC: HSPCs expressing at least one (or more) of the markers). Engraftment analysis was performed within the bone marrow of recipient mice 14 or 16 weeks after transplantation. **e,** Representative FACS plots of engraftment analysis. The presence of human CD45+ engraftment was analyzed by flow cytometry, and the engrafted cells were further examined for the fraction of CD33+ myeloid and CD19+ lymphoid cells. The number of recipient mice is shown in **f** (each data point represents an individual mouse). **f,** Quantification of engraftment analysis in **e**. Data are mean ± SEM.

Fig. S4. Methylcellulose differentiation assays of sorted pre-malignant and malignant stem cells.

a, Schema of methylcellulose assay to examine the differentiation of immunophenotypic, sorted pre-malignant and malignant stem cells from independent MDS and AML samples (unpaired). Sorted HSPCs (Lin−CD34+CD38−) that were triple-negative for CD45RA, CD123, and IL1RAP (preMDS-SC and preAML-SC), and HSPCs expressing at least one (or more) of the markers (MDS-SC and AML-SC) were plated in semisolid methylcellulose medium. The colonies of different lineages (E: erythroid; M: macrophage; G: granulocyte; GM and GEMM: mixed lineages) were scored by microscopy 2 weeks after plating. Thereafter, the methylcellulose semisolid medium with colonies was dissolved and resuspend in PBS buffer to make single cell suspensions. Cells were then stained with cell surface markers against CD14, CD15 and CD235a, and analyzed by flow cytometry. **b,** Clonogenicities (per 1000 cells plated) of sorted preMDS-SC and MDS-SC (n=5), as well as preAML-SC and AML-SC (n=4). **c,** Frequency of colonies of different lineages (n=5 for MDS samples; n=4 for AML samples). **d,** Frequency of cells expressing specific lineage markers (n=3 for MDS samples; n=4 for AML samples). If not specified otherwise, data are mean ± SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed paired Student's t test).

Fig. S5. Targeted capture sequencing of sorted stem cells and blasts from MDS and matched sAML samples.

a,b, Comparison of variant allele frequencies between whole genome amplified samples and unamplified samples of 2 different patients with AML. Prior to sequencing, we performed whole genome amplification (WGA) of sorted cell populations using the REPLI-g method, which utilizes the proofreading enzyme Phi 29 polymerase to achieve high-fidelity amplification of genomic DNA. WGA was performed with 10ng DNA of each patient, thereafter 500ng of purified WGA products were subjected to targeted sequencing along with 500ng of unamplified DNA. After pre-processing of the sequencing data, FreeBayes was used for variant detection (see methods). Then potential germline variants reported in dbSNP database with population frequency >1% were removed, before the comparison of VAFs between WGA sample and matched unamplified sample. 425 and 403 variants were obtained for AML01 and AML02, respectively. And we observed that the VAFs in WGA samples were highly consistent with the unamplified samples ($p < 2.2\times10^{-16}$). In addition, we did not observe mutation artifacts that passed the cut-offs for detection of somatic mutations (see methods) when comparing WGA samples to the unamplified samples, indicating that WGA did not distort detection and quantitation of variants. Pearson coefficient R and p-value were calculated by R software. **c,** Mean sequencing coverages in target regions across different cell populations are shown. CD45− non-hematopoietic cells were used as germline control. **d,** Representative results of Sanger sequencing of a *RUNX1* mutation in different cell populations isolated from P7026. The Sanger sequencing was repeated twice. **e,** Correlation of variant frequency estimated by targeted sequencing and Sanger sequencing. A total of 36 pairs of mutations-samples were tested, and Pearson correlation coefficient R and significant p-value calculated by R software are shown. **f-h,** Number of mutations in all genomic regions (**f**), exonic regions (**g**), and non-exonic regions (**h**) across the different cell populations. On average, we observed 8.6 mutations in coding regions (range, 4 to 13) in MDS stem cells compared to 5.3 (range, 3 to 7) in the

MDS blasts (p = 0.012); and 7.8 coding mutations (range, 4 to 14) in AML stem cells compared to 5.6 (range, 3 to 9) in AML blasts (**g**). Interestingly, we also observed high numbers of mutations in non-exonic regions, and these mutations in non-exonic regions were significantly more frequent in stem cell populations compared to blasts for both MDS (on average 17.8 vs. 8.1; p = 0.0008) and AML (15.0 vs. 6.6; p = 0.003) (**h**). *p < 0.05, **p < 0.01, ***p < 0.001 (two-tailed paired Student's t test).

Fig. S6. Distribution of clonal and subclonal mutations in stem cells and blast cells in each patient.

a, Cancer cell fractions (CCFs) based on somatic mutations detected in targeted sequencing were estimated with variant allelic frequency (VAF), ploidy and purity as previous described (33). Mutations were defined as clonal mutation if the upper border of the 95% confidence interval of the CCF was greater than 0.95. **b-g**, Distribution of CCFs of clonal and subclonal mutations in stem cells and blasts of patients P7024 (**b**), P7025 (**c**), P7027 (**d**), P7028 (**e**), P7030 (**f**), and P7031 (**g**). Violin plot is showing frequency distribution (kernel density) of clonal mutations (orange) and subclonal mutations (grey).

Fig. S7. Mutation spectrum in stem cells resembles age-associated and DNA repair-associated mutational signatures.

a,**b**, The 96 trinucleotide mutational spectra of mutations detected in stem cells and blast cells at the stage of MDS (**a**) and sAML (**b**). X-axis is showing the 96 combination of nucleotide changes, and their relative weights inferred by deconstructSigs is shown on the Y-axis. **c**, Relative weight and underlying mechanisms of different mutation signatures inferred from the mutational spectra shown in panel **a, b**.

Fig. S8. Clonal composition of stem cell and blast populations inferred with CCFs of mutations.

Clonal composition of stem cell and blast populations in MDS and sAML, respectively, in patients P7024 (**a**), P7025 (**b**), P7026 (**c**), P7027 (**d**), P7030 (**e**), and P7031 (**f**). Based on the CCFs, mutations are clustered as clones with Sciclone and denoted with same color. Mutation was denoted with grey if the estimated possibility of the mutation to be clustered in the subclone was lower than 0.95.

Fig. S9. Clonal evolution of stem cell and blast populations during progression from MDS to sAML, inferred by CCFs of mutations.

a-e, Comparison of clonal architectures of pre-malignant stem cells (preMDS-SC vs. preAML-SC) (panels on the left), malignant stem cells (MDS-SC vs. AML-SC) (middle panels), and blasts (right panels) in individual patients progressing from MDS to sAML. (**a**) Patient P7024, (**b**) patient P7025, (**c**) patient P7026, (**d**) patient P7027, (**e**) patient P7028, (**f**) patient P7030, (**g**) patient P7031. Each dot represents one somatic mutation (covered by > 20x in both samples shown), and mutations are clustered as clones by SciClone based on the CCFs, indicated by differently colored symbols. Data was visualized with ggplot2 package of R with results from SciClone plus kernel density plot. "NA" indicates unclustered mutations with insufficient sequencing coverage, or mutations with possibility of less than 0.95 in the cluster. For 4 out of 7 patients (P7024, P7025, P7026, P7028) the blasts remained more stable during the MDS-to-sAML transition than the stem cell compartments. In all these patients, the MDS-SC showed massive subclonal differences compared to AML-SC, with subclones that were specific to either the MDS or the sAML stage of the same patient. In two patients (P7024 and P7026) this was additionally accompanied by subclonal differences at the pre-malignant stem cell level (preMDS-SC vs. preAML-SC). Interestingly, in patients P7028 and P7025, preMDS-SC, as well as the MDS blast population showed relatively stable clonal dynamics, whereas MDS-SC underwent extensive evolution during the progression from MDS to sAML. Two patients (P7027 and P7031) showed substantial subclonal evolution in each of the preMDS-SC/preAML-SC, the MDS-SC/AML-SC, and the MDS-blast/AML-blast transitions. Consist patterns of clonal evolution were observed in the analyses with VAFs of mutations (data not shown).

Fig. S10. Somatic mutations shared between samples at the MDS and sAML stages are present in T cells.

a,b, Relative utilization of specific Vα (**a**) and Vβ (**b**) gene segments in T cells isolated from patients' samples, as well as T cells from two healthy controls (HC1 and HC2). The legends show the top 15 gene segments. We hypothesized that the shared mutations were acquired during early stages of disease initiation, thus shared across all cell populations. To test this, we performed targeted re-sequencing of these mutations in matched T cells isolated from MDS bone marrow samples of each patient (**c-g**), as the majority of T cells develop during childhood long before the diagnosis of MDS or AML. To rule out the possibility that T cells were derived from the expansion of a T cell clone arising from mutant HSCs, we performed T cell receptor (TCR) repertoire analysis of the T cells by sequencing the TCR mRNAs (see methods), and found that the TCR repertoire of the patients was highly diverse with polyclonal patterns indistinguishable from healthy controls. (**c-g**), Selected shared mutations, MDS-specific mutations, and AML-specific mutations were examined in sorted CD4+ T cells isolated from the MDS bone marrow sample of the same patient with targeted re-sequencing by Fluidigm and MiSeq. (**c**) Patient P7024, (**d**) patient P7026, (**e**) patient P7027, (**f**) patient P7030, (**g**) patient P7031. Variant allele frequencies (VAFs) of mutations are shown. Data show presence in T cells of most shared mutations, but not of MDS- or AML-specific mutations.

Fig. S11. Shared mutations between MDS and sAML have higher VAF compared to mutations specific to MDS or AML.

a, Schematic non-linear clonal evolution of stem cells in MDS and progression to sAML. Accumulation of mutations gives rise to distinct subclones at the stem cell level, and different subclones contribute to the generation of MDS blasts and sAML progression. Based on this hypothesis, stage-specific (MDS or

AML) mutations would be expected to reside in smaller subset of cells, and thus have lower VAF values. We therefore examined the VAFs of "shared" (present in all sorted cellular subsets) and "specific" (only present in one cellular subset) mutations. **b,** Box plots of VAFs of shared (MDS and AML) or stage-specific (MDS or AML) mutations in different stem and blast cell populations of each individual patient (rows). Results were visualized with the ggplot2 package of R. Lower and upper hinges are showing the 25% and 75% quartiles, respectively. And the segment inside the rectangle shows the median of VAF values. We found that the stage specific mutations consistently had lower VAFs compared to the shared mutations in each patient *p < 0.05, **p < 0.01, ***p < 0.001 (two-tailed unpaired Student's t test). N.s: non-significant; n.a: p-value not available as there were less than 2 data points (mutations) in the either shared or specific group.

Fig. S12. Targeted re-sequencing with Fluidigm Biomark HD followed by Miseq platforms.

a, Somatic mutations detected by HiSeq targeted sequencing were validated by MiSeq following targeted capture with the Fluidigm Biomark platform. We also re-sequenced the original DNA of patients P7026 and P7027 by targeted sequencing. Sequencing coverages of MiSeq data of the tested cell populations from patients P7026 and P7027 are shown. **b,** Comparison of VAFs determined by HiSeq vs MiSeq data. Each dot represents one somatic mutation. X- and Y-axis are showing the variant allele frequency (VAF) of mutations determined from the data of HiSeq and MiSeq, respectively. 82 pairs of mutations-samples were tested, and Pearson correlation coefficient R and significance p-value calculated by R software are shown. We found high concordance between the VAFs determined by HiSeq and MiSeq ($p < 2.2\times10^{-16}$) **c,** Number of single cell sequencing samples tested and passing the quality control in each population. For single cell sequencing, we examined sorted, immunophenotypic preMDS-SC, MDS-SC, MDS blasts, preAML-SC, AML-SC, and AML blasts of the patients. **d-h,** CCFs of mutations estimated by HiSeq targeted sequencing are validated by single cell sequencing. Correlation between mutation CCFs estimated by HiSeq and CCFs determined by single cell sequencing in patient P7024 (**d**), P7026 (**e**), P7027 (**f**), P7030 (**g**), and P7031 (**h**). Examined mutations in each patient are shown in **Fig. 3**, and only

mutations detected by HiSeq and/or single cell sequencing are shown. CCFs in single cell sequencing studies were calculated as the fraction of single cells with the mutation within all single cells carrying early mutations (e.g. *TP53* and *U2AF1*). We found significant correlation between the CCFs estimated by HiSeq of sorted cell populations and CCFs determined by single cell sequencing in all patients, including P7024 ($p = 1.16\times10^{-11}$, n=40 pairs of mutations-samples), P7026 ($p = 1.45\times10^{-15}$, n=40), P7027 ($p = 2.38\times10^{-12}$, n=44), P7030 ($p = 1.10\times10^{-5}$, n=18), and P7031 ($p = 2.90\times10^{-5}$, n=17). Pearson coefficient R and p-value were calculated by *cor.test* function of R.

Fig. S13. Variant allele frequencies of mutations across different stem and blast cell populations.

a-e, Variant allele frequencies (VAFs) in sorted cell populations of the same mutations tested in single cell sequencing shown in **Fig. 3**. (**a**) Patient P7024, (**b**) patient P7026, (**c**) patient P7027, (**d**) patient P7030, (**e**) patient P7031.

Fig. S14. Spatiotemporal subclonal evolution during the progression from MDS to sAML in patient P7027, determined by single cell sequencing of sorted stem and blast cells.

a, CCFs of shared (left), MDS-specific (middle), AML-specific (right) mutations in different stem and blast populations at the MDS and sAML stage of patient P7027. **b,** Single cell targeted sequencing of mutations across different cell populations of patient P7027. **c,** Schematic model of clonal evolution in different stem and blast cell populations in patient P7027. Subclones of MDS stem cells with early founding mutations (i.e. *TET2*) remained present during MDS blast generation as well as AML progression, whereas other mutations, e.g. *PCM1* and *EP300*, only occurred in MDS blasts but not during progression to sAML. Interestingly, we had also identified a *RUNX1* mutation that had a low CCF in MDS-SC, but then expanded at the AML stage. Single cell sequencing confirmed that *TET2* was indeed present in almost all the cells across the different populations, whereas mutation of *FANCD2* resided in a subclone within *TET2*-mutated cells (**b**). Importantly, the mutation of *RUNX1* also resided in a progeny subclone of *TET2* -mutated cells, however, one that was distinct from the subclone with *FANCD2*

mutation (**b**). In addition, within the *TET2*-mutated preMDS-SC, mutations of *NUP214* and *EP300* were acquired successively, however, these never gained dominance within the MDS or AML blast cells (**b**). Within the same parent subclone, *SMAD4* and *KMT2A* mutations were acquired in progeny different from subclones with *ERBB2* and *RUNX1T1* co-mutations (**b**). In this patient, the progression to sAML originated from a subclone of MDS stem cells and was triggered by a *RUNX1* mutation, whereas other mutations, e.g. *PCM1 and EP300*, only occurred in MDS blasts but did not play a role in progression to sAML.

Fig. S15. Spatiotemporal subclonal evolution in sorted stem and blast cells during the progression from MDS to sAML of patients P7025, P7028 and P7031.

a, CCFs of shared (left), MDS-specific (middle), AML-specific (right) mutations in different stem and blast populations at the MDS and sAML stage of patient P7025. **b,** Schematic model of clonal evolution in different stem and blast cell populations in patient P7025. **c,** CCFs of shared (left), MDS-specific (middle), AML specific (right) mutations across all cell populations in patient P7028. **d,** Schematic model of clonal evolution in different stem and blast cell populations in patient P7028. Despite the larger number of stable mutations, results of P7025 and P7028 still indicate an overall model of parallel MDS and sAML evolution at the stem cell level (in these cases, with slightly later branching) rather than linear MDS to sAML progression. **e,** CCFs of shared (left), MDS-specific (middle), AML-specific (right) mutations across all cell populations in patient P7031. **f,** Single cell targeted sequencing of mutations across different MDS cells populations of patient P7031. **g,** Schematic model of clonal evolution in different stem and blast cell populations in patient P7031.