Supplementary Data legends

A-Supplementary Figures

Figure S1.
ETO2-GLIS2 associates with megakaryoblastic and myeloid phenotypes.
Related to Figure 1.
A, Overall survival of EG^+ patients diagnosed for AML (blue, n=10) and AMKL (orange, n=8). Log rank Mantel-Cox test p-value=0.1460.
B, Flow cytometry analyses on two ETO2-$GLIS2^+$ cell lines derived from patients: MO7e (megakaryoblastic phenotype) and CMS (myeloid).
C, Fluorescent in situ hybridization (FISH) analysis with *ETO2* (green) and *GLIS2* (red) probes on MO7e and CMS. The arrows indicate the fusion event on the derivative chromosome 16. Note that MO7e present two derivative chromosome 16 while CMS only shows one.
D, Sequences and schematic representation of the genomic fusion point in MO7e and CMS cells. MO7e cells present an insertion of 87bp at the fusion point that is not present in CMS cells.
E, Quantitative RT-PCR analyses in EG^- HEL cells, and EG^+ MO7e and CMS.
F, RT-PCR detection of the *ETO2-GLIS2* fusion mRNA in purified leukemic populations from the dual phenotype #5 patient.
G, Relative *EG* mRNA expression in EG^+ AMKL (#3, #4, #13) and AML (#5, #6, #15) patient cells measured by quantitative RT-PCR.

Figure S2.
A novel inducible ETO2-GLIS2 expression transgenic murine model.
Related to Figure 2.
A, Schematic representation of the inducible ETO2-GLIS2 (iEG) mouse model. The human ETO2-GLIS2 cDNA is introduced at the HPRT locus (Chromosome X) under the control of a Tet-responsive element (TRE). The reverse transactivator (rtTA) cDNA is inserted at the *ROSA26* locus (Chromosome 6). Upper right panel: *EG* mRNA expression in BM cells grown *in vitro* with the indicated dose of Dox is normalized to *HPRT* mRNA. Mean +/- SEM (n=3) is shown. Lower right panel: EG and HSP70 (loading control) protein levels in the same cells.
B, Relative *EG* mRNA expression in human AMKL/AML and mouse iEG^+ leukemia blasts normalized to *RPLP0*. Orange: human AMKL samples, blue: human AML samples.
C, Western blot on human AMKL and mouse leukemia blasts to detect EG protein (using an anti-ETO2 antibody) expression. RPLP0 served as loading control. HEL: EG-negative control human cell line, M07e and CMS: EG^+ patient-derived cell lines fusion, #13 & #19: EG^+ patient samples. BA/F3: EG-negative control murine cell line, #207-210-291: iEG^+ murine leukemia samples. Black arrow represents the most common isoform of the predicted EG protein (*ETO2* exon 11 fused to *GLIS2* exon 3), red star indicates an unusual isoform observed in patient #19 and resulting from a fusion between *ETO2* exon 11 and *GLIS2* exon 2.
D, Representative flow cytometry analysis and expression of $CD150^+$ and $CD41^+$ gated on LIN^- Kit^+ BM cells from heterozygous iEG females 6-weeks post-induction (CTRL n=9, iEG n=11), p=0.0007.
E, BM $LIN^-CD150^-Kit^+CD41^+$ cells from (D) were flow purified and plated in methylcellulose cultures. After 7 days of culture, individual colonies were picked and analysed by flow cytometry for the expression of megakaryoblastic (CD41) and myeloid (Gr1) markers. Representative flow cytometry analysis of the three types of colonies is shown: $CD41^+$ only colony, $Gr1^+$ only colony and immature colony with a predominance of $CD41^-$ and $Gr1^-$ cells. Right panel: histogram representing the quantification of each type of colonies in 3 independent experiments from 3 independent animals.

F, Kaplan-Meier survival plot of heterozygous iEG females and controls continuously exposed to Dox (CTRL n=11, iEG n=32). Global median survival for iEG mice is 153 days, p<0.0001.
G, Flow cytometry analyses in BM ($CD41^+$ and Kit^+) or spleen ($CD11b^+Gr1^+$) of diseased mice showing 2 different phenotypes at time of analysis (CTRL n=5, Group 1: $CD41^+$ n=4 and Group 2: $Kit^+/CD11b^+Gr1^+$ n=15). $CD41^+$ (CTRL vs. Group 1: p=0.0035, CTRL vs. Group 2: p=0.9705); Kit^+ (CTRL vs. Group 1: p=0.0002, CTRL vs. Group 2: p=0.0590); $CD11b^+Gr1^+$ (CTRL vs. Group 1: p=0.4323, CTRL vs. Group 2: p=0.0025).
H, Histopathological analyses of BM sections of primary mice (CTRL and 2 groups of iEG) and secondary recipients.
Statistical significance is indicated as p values (Student's t test with the exception of the survival curve where a log rank Mantel-Cox test is used). *: p<0.05, **: p<0.01, ***: p<0.001.

Figure S3.
Expression of EG in various hematopoietic progenitors.
Related to Figure 3 and Figure 4.
A, Wright-Giemsa staining of cytospots from FL or ABM derived methylcellulose cultures at passage P3.
B, Wild-type FL (E12.5), new-born BM and ABM progenitors were transduced with ETO2-GLIS2-encoding retroviruses (rEG), flow sorted and replated for 3 rounds in methylcellulose cultures. Flow cytometry analysis at P3 for the *in vitro* analysis wild-type hematopoietic progenitors retrovirally transduced with ETO2-GLIS2. Compare to the phenotype obtained in Fig. 3C with the iEG model using the same conditions.
C, Experimental flow cytometry sorting strategy.
D, Indicated populations purified according to the parameters in (C), were plated into methylcellulose cultures with Dox and analysed 7 days later. Histogram of the percentages of $CD41^+$ and $Gr1^+$ cells from CTRL and iEG methylcellulose cultures at P1 are shown.
E, Number of colonies obtained at P1.
F, iEG FL HSC and MPP were plated in methylcellulose for 4 passages and replated at the 5^{th} passage in presence or absence of Dox. The morphology of cells obtained at P5 +/- Dox is shown. Wright-Giemsa staining. Magnification: x100.
G, Schematic representation of megakaryocytic and myeloid output in cultures from normal FL or ABM progenitors. Orange filling represents a significant (>15%) megakaryoblastic potential and progenitors with myeloid potential are in blue.
H, ABM cells from iEG animal were replated for 8 passages and the $CD41^+$ or $Gr1^+$ populations were flow purified to perform cytospots preparations that were then stained with Wright-Giemsa.
I, iEG transformed cell lines presenting a $CD41^+$ or $Gr1^+$ phenotype were cultured with varying doses of Dox for 7 days and their phenotype was analyzed by flow cytometry. Left panel: histogram representing the quantification of myeloid ($Gr1^+$ cells), megakaryoblastic ($CD41^+$ cells) or both expression ($Gr1^+CD41^+$) populations for each dose. Right panel: Representative flow cytometry analyses of cell lines with the indicated Dox dose.
J, Representative images of single cell-derived colonies presenting an exclusive myeloid ($Gr1^+$), an exclusive megakaryoblastic ($CD41^+$) or a mixed myeloid and megakaryoblastic potential.

Figure S4.
Single cell transcriptome in CTRL and iEG fetal and adult LT-HSC and adult MPP4.
Related to Figure 5.
A, Violin plots of *EG* expression in iEG or CTRL cells (from scRNAseq expression data).
B, Heatmaps of the 500 most differentially expressed genes between CTRL and iEG FL LT-HSC 24 hours after Dox induction. Log2 values are represented. The order of the genes from top to

bottom is conserved for all three heatmaps and is based on the clustering performed for FL HSC (left panel).
C, Heatmaps of the 50 most differentially expressed genes between CTRL and iEG ABM LT-HSC 24 hours after Dox induction. Log2 values are represented. The order of the genes from top to bottom is conserved for all three heatmaps and is based on the clustering performed for ABM HSC (central panel).
D, GSEA using published EG^+ patient-derived gene expression signatures (40, 41) on CTRL vs. iEG FL LT-HSC, ABM LT-HSC and MPP4 scRNAseq data.
E, GSEA using published HSC and MPP4 signatures (38) on CTRL vs. iEG FL and ABM LT-HSC.
F, Violin plot of *Hlf* expression in LT-HSC. The boxplots indicate median, first and third quartiles.

Figure S5.
Transcription factor expression and activity in iEG or CTRL cells.
Related to Figure 5.
A, Experimental design to infer transcription factor activity from scRNAseq expression data.
B, Violin plots of EG expression and activity in FL and ABM LT-HSC. Log2 values of the normalized read counts are represented.
C, Scatter plot of differential expression (log2FC) against activity (Normalized Enrichment Score: NES) obtained by comparing for each gene iEG vs. CTRL LT-HSC. Horizontal lines represent the activity threshold of NES (FDR < 0.05 corresponding to a NES of +/- 2.5) and vertical lines represent a threshold two-fold change in expression in iEG vs. CTRL LT-HSC. Tinted squares indicate factors for which a significant differential activity is inferred without a significant differential expression in iEG vs. CTRL LT-HSC.
D, 2D representation of the cells using t-SNE analysis of expression data or activity data inferred with the EG network. Density plots for each condition was overlaid (Red: LT-HSC iEG [triangles], Blue: LT-HSC CTRL [squares], Grey: MPP4 iEG [diamonds], Green: MPP4 CTRL [dots]).
E, violin plot representation of predicted activity of the indicated factor using targets inferred in the normal network.

Figure S6.
iEG targets LT-HSC to induce aggressive leukemia.
Related to Figure 6 and Figure 7.
A, Quantitative RT-PCR analyses in blasts from patients with AMKL (#13), AML (#15) and in the different immature, megakaryoblastic and myeloid blast populations from the dual phenotype ETO2-$GLIS2^+$ AML patient (#5).
B, Percentages of GFP^+ cells within nucleated blood cells of recipient mice. Analyses were performed at 2, 4, 6, 16, 28 weeks and 40 weeks. Mean+/-SEM (n=4) is shown.
C, Histopathology of BM from recipients engrafted with CTRL or iEG HSC from ABM or FL. HE: Hematoxylin-Eosin staining, VWF: von Willebrand factor immunohistochemistry highlighting megakaryocytic features, GFP immunohistochemistry: highlight all donor-derived cells. For ABM, but not FL, it was possible to perform VWF and GFP stainings on consecutive sections. Note that the maturing megakaryocytes visible in ABM HSC transplanted mice are GFP^- and therefore most likely do not derive from GFP^+ iEG^+ donor cells.
D, Flow purified $CD41^+$ blasts from primary iEG HSC FL recipient mice were grown for 2 days *in vitro* with (Dox +) or without (Dox -) Dox to evaluate their potential to generate maturing megakaryocyte or myeloid cells. Representative flow cytometry analyses show that some recipients (e.g. #1) the leukemic blasts can differentiate toward both megakaryocytes ($CD41^+CD42^+$) and myeloid ($CD11b^+Gr1^+$) cells and other recipients generate a more predominant myeloid differentiation (e.g. #2).

E, Quantitative RT-PCR analysis of *ETO2-GLIS2* (iEG), *Erg* and *Gata1* expression in cells obtained in (C). Expression was normalized to *HPRT* mRNA level. *iEG:* p=0.0121, *Erg*: p<0.0001, *Gata1*: p=0.0725.
F, Representative flow cytometry result of the percentage of GFP^+ cells in white blood cells (WBC), erythrocytes (RBC) and platelets (Plt) of II^{ary} ABM recipients.
G, Percentage of GFP in the BM of II^{ary} recipients engrafted with cells from ABM LT-HSC-derived I^{ary} recipients. Analysis was performed at 2-weeks post-graft in Dox-exposed II^{ary} recipients. The same day, Dox was removed in Group B and maintained in Group A (Group A n=6, Group B n=6). These results indicate that the initial percentage of GFP is the same in the two groups.
H, Follow-up of the WBC and percentage of GFP in WBC upon removal of Dox treatment. D0 represents the time at which Dox was removed in Group B and maintained in Group A. Mean+/SEM (n=6 / group) is shown.
I, II^{ary} recipients in Group B (Dox -) were analysed 6 months post-Dox removal to assess residual long-term HSC potential of iEG^+ leukemic (GFP^+) cells. Histograms indicate that within a recipient the percentages of GFP^+ cells were similar in the BM, WBC, RBC and PLT indicating that GFP^+ cells are endowed with LT-HSC potential.
Statistical significance is indicated as p values (Student's t test). *: p<0.05, **: p<0.01, ***: p<0.001.

Figure S7.
MLL-AF9 expression in fetal liver LT-HSC induces megakaryoblastic features
A, Upper panel: wild-type FL (E12.5) and ABM progenitors were transduced with MLL-AF9 (rMA9)-encoding retroviruses, plated in methylcellulose for 7 days and analysed by flow cytometry for megakaryoblastic (CD41) and myeloid (Gr1) markers. A representative flow cytometry analysis at P1 is shown. Lower panel: histogram representation of the percentage of indicated markers (n=3).
B, Left panel: wild-type HSC and MPP from FL (E14.5) were transduced with MLL-AF9 (rMA9)-encoding retroviruses, plated in methylcellulose for 7 days (=P1) and analysed by flow cytometry for megakaryoblastic (CD41) and myeloid (Gr1) markers. A representative flow cytometry analysis at P1 is shown. Right panel: histogram representation of the percentage of indicated markers (n=3).
C, Retroviral MLL-AF9 (rMA9) HSC and MPP4 cells derived from methylcellulose were transplanted into wild-type recipients. Representative flow cytometry analysis of diseased recipients at time of sacrifice.
D, Histopathological analyses of spleen sections of mice described in (C).
E, Relative expression of transcription factors (*Gata1*, *Erg*, *Cebpa*, *Spi1*) in $CD41^+$ and $Gr1^+$cells derived from mice transplanted with HSC described in (C), measured by quantitative RT-PCR analysis.

B-Supplementary Tables

- Table S1: ETO2-$GLIS2^+$ patients information

- Table S2: Differentially expressed genes in scRNAseq from FL LT-HSC CTRL vs. FL LT-HSC iEG

- Table S3: Differentially expressed genes in scRNAseq from ABM LT-HSC CTRL vs. ABM LT-HSC iEG

- Table S4: Differentially expressed genes in scRNAseq from ABM MPP4 CTRL vs. ABM MPP4 EG

- Table S5: FL iEG transcription factor network
Transcription factor network inferred using the scRNAseq from FL LT-HSC from CTRL and iEG mice.

- Table S6: ABM iEG transcription factor network
Transcription factor network inferred using the scRNAseq from ABM LT-HSC and MPP4 from CTRL and iEG mice.

- Table S7: Normal transcription factor network
Transcription factor network inferred using published scRNAseq from normal murine HSPC (Nestorowa et al. Blood 2016).

- Table S8: Predicted activity of transcriptional regulators in FL LT-HSC CTRL vs. FL LT-HSC iEG (normal network)

- Table S9: Predicted activity of transcriptional regulators in ABM LT-HSC CTRL vs. ABM LT-HSC iEG (normal network)

- Table S10: Reagents and resources

- Table S11: List of oligonucleotides

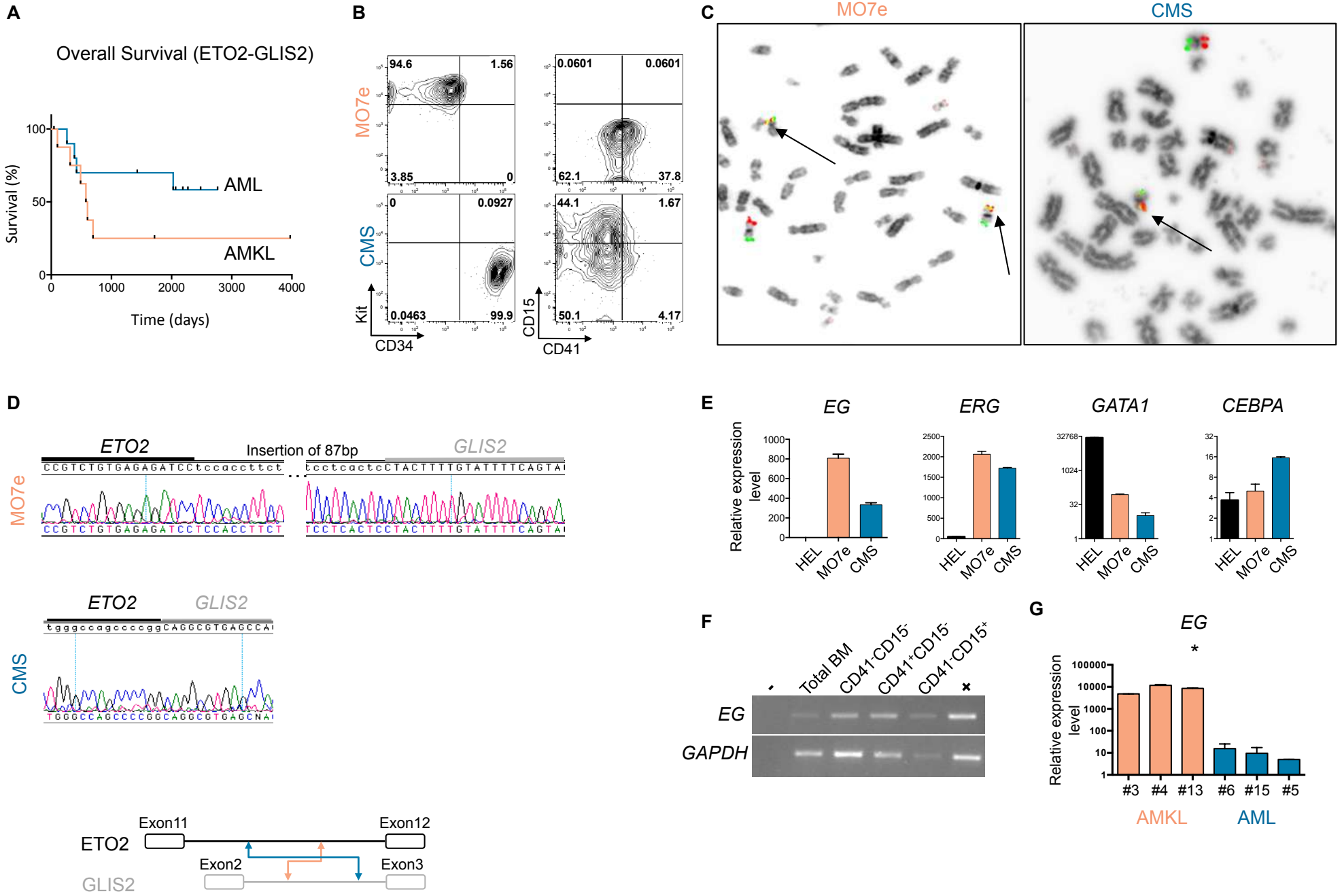
A
Overall Survival (ETO2-GLIS2)
Survival (%)
Time (days)
AML
AMKL
B
MO7e
CMS
Kit
CD34
CD15
CD41
C
MO7e
CMS
D
ETO2
GLIS2
Insertion of 87bp
MO7e
CMS
Exon11
Exon12
Exon2
Exon3
E
EG
ERG
GATA1
CEBPA
Relative expression level
HEL
MO7e
CMS
F
Total BM
CD41-CD15-
CD41+CD15-
CD41-CD15+
EG
GAPDH
G
EG
Relative expression level
#3
#4
#13
#6
#15
#5
AMKL
AML

Figure S1

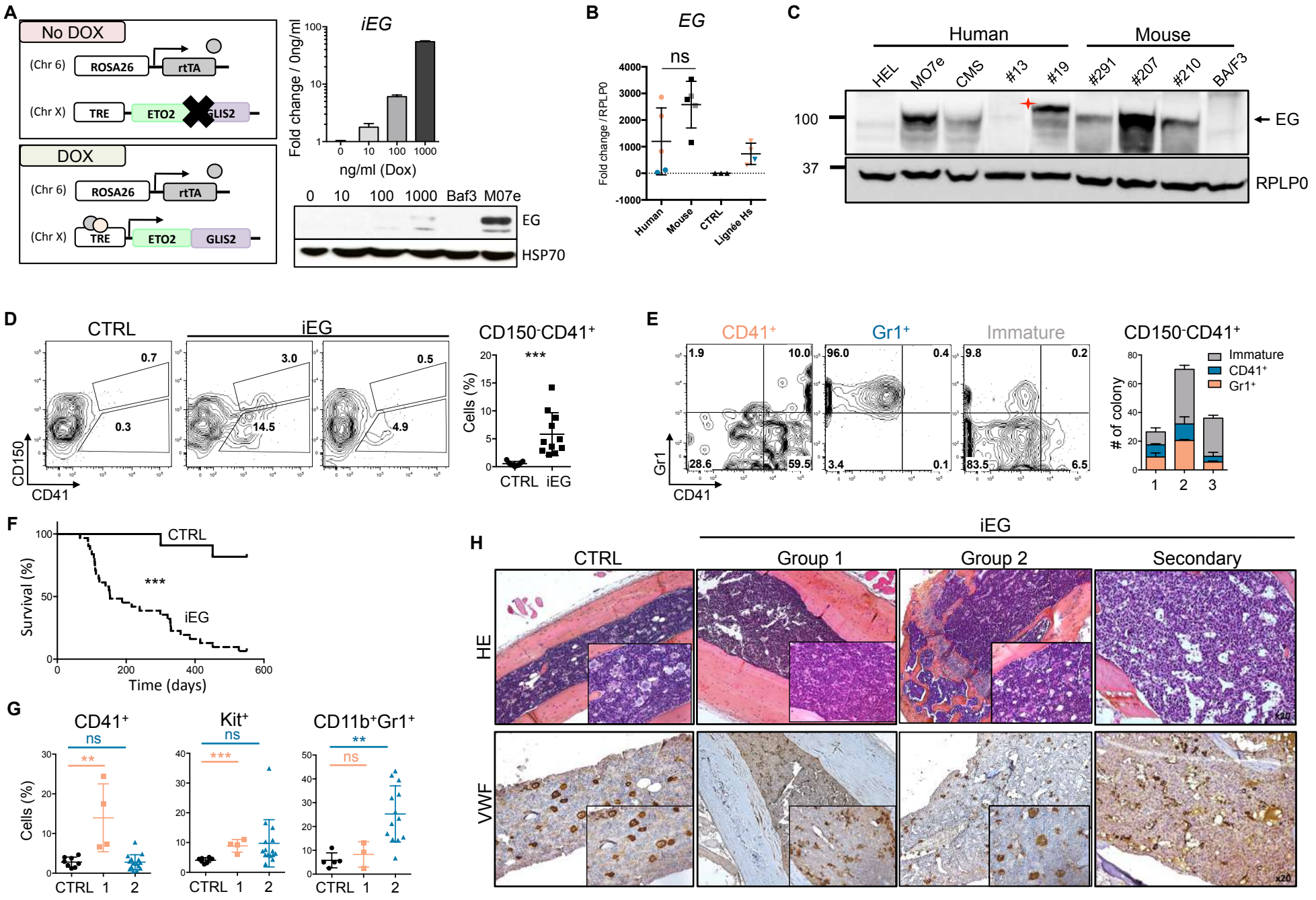
A
No DOX
(Chr 6)
ROSA26
rtTA
(Chr X)
TRE
ETO2
GLIS2
DOX
iEG
Fold change / 0ng/ml
ng/ml (Dox)
0 10 100 1000 Baf3 M07e
EG
HSP70
B
EG
ns
Fold change / RPLP0
Human
Mouse
CTRL
Lignée Hs
C
Human
Mouse
HEL
MO7e
CMS
#13
#19
#291
#207
#210
BA/F3
100
37
EG
RPLP0
D
CTRL
iEG
0.7
0.3
3.0
14.5
0.5
4.9
CD150
CD41
CD150-CD41+
***
Cells (%)
CTRL iEG
E
CD41+
Gr1+
Immature
1.9
10.0
28.6
59.5
96.0
0.4
3.4
0.1
9.8
0.2
83.5
6.5
Gr1
CD41
CD150-CD41+
Immature
CD41+
Gr1+
# of colony
1 2 3
F
CTRL
iEG
***
Survival (%)
Time (days)
G
CD41+
Kit+
CD11b+Gr1+
ns
**
***
Cells (%)
CTRL 1 2
H
iEG
CTRL
Group 1
Group 2
Secondary
HE
VWF

Figure S2

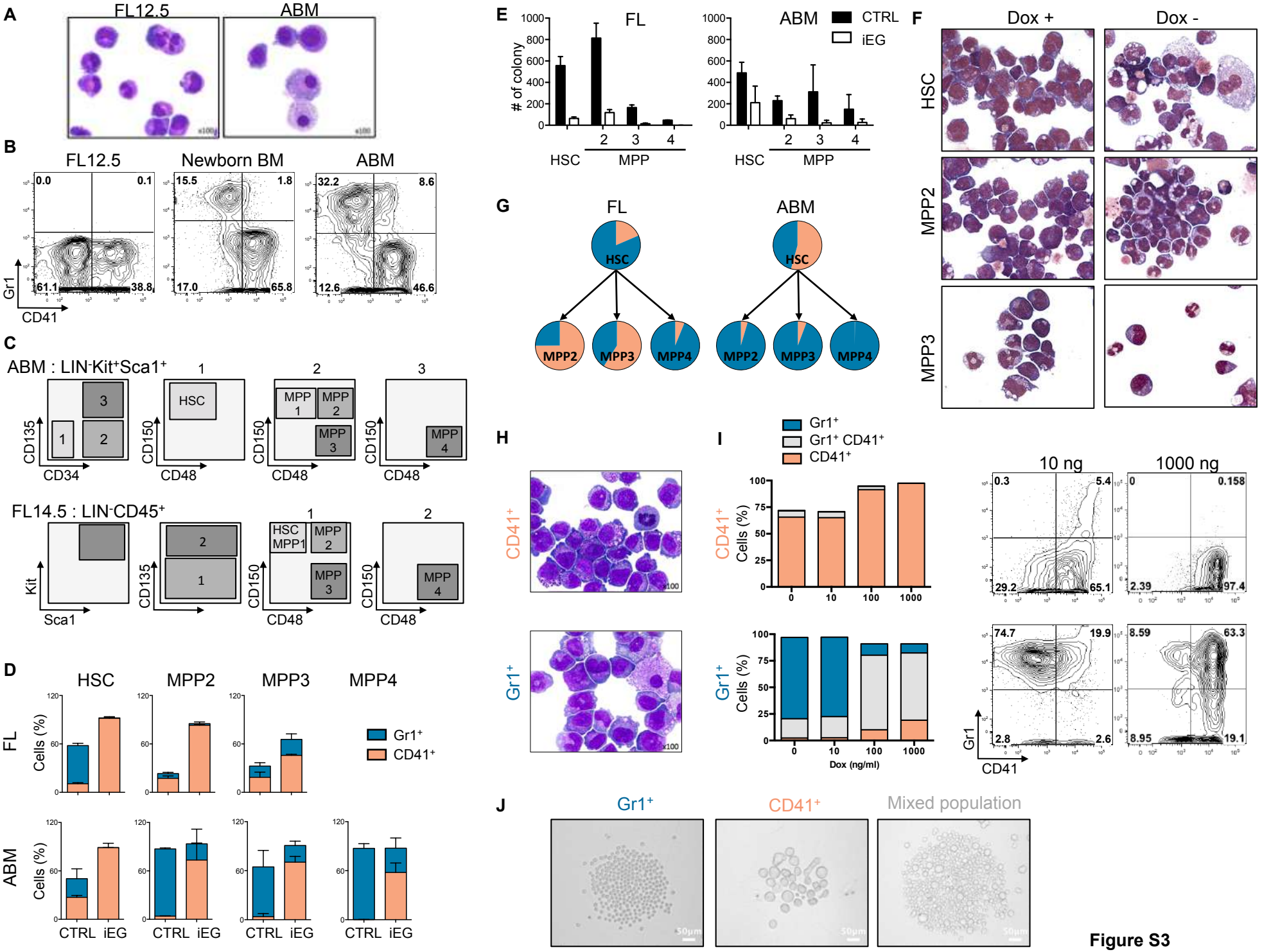
A
FL12.5
ABM
B
FL12.5
Newborn BM
ABM
0.0
0.1
61.1
38.8
15.5
1.8
17.0
65.8
32.2
8.6
12.6
46.6
Gr1
CD41
C
ABM : LIN-Kit+Sca1+
CD135
CD34
CD150
CD48
HSC
MPP 1
MPP 2
MPP 3
MPP 4
FL14.5 : LIN-CD45+
Kit
Sca1
HSC MPP1
D
HSC
MPP2
MPP3
MPP4
FL
ABM
Cells (%)
Gr1+
CD41+
CTRL
iEG
E
FL
ABM
CTRL
iEG
# of colony
HSC
MPP
F
Dox +
Dox -
HSC
MPP2
MPP3
G
FL
ABM
MPP2
MPP3
MPP4
H
CD41+
Gr1+
I
Gr1+
Gr1+ CD41+
CD41+
Cells (%)
Dox (ng/ml)
10 ng
1000 ng
0.3
5.4
29.2
65.1
0
0.158
2.39
97.4
74.7
19.9
2.8
2.6
8.59
63.3
8.95
19.1
Gr1
CD41
J
Gr1+
CD41+
Mixed population

Figure S3

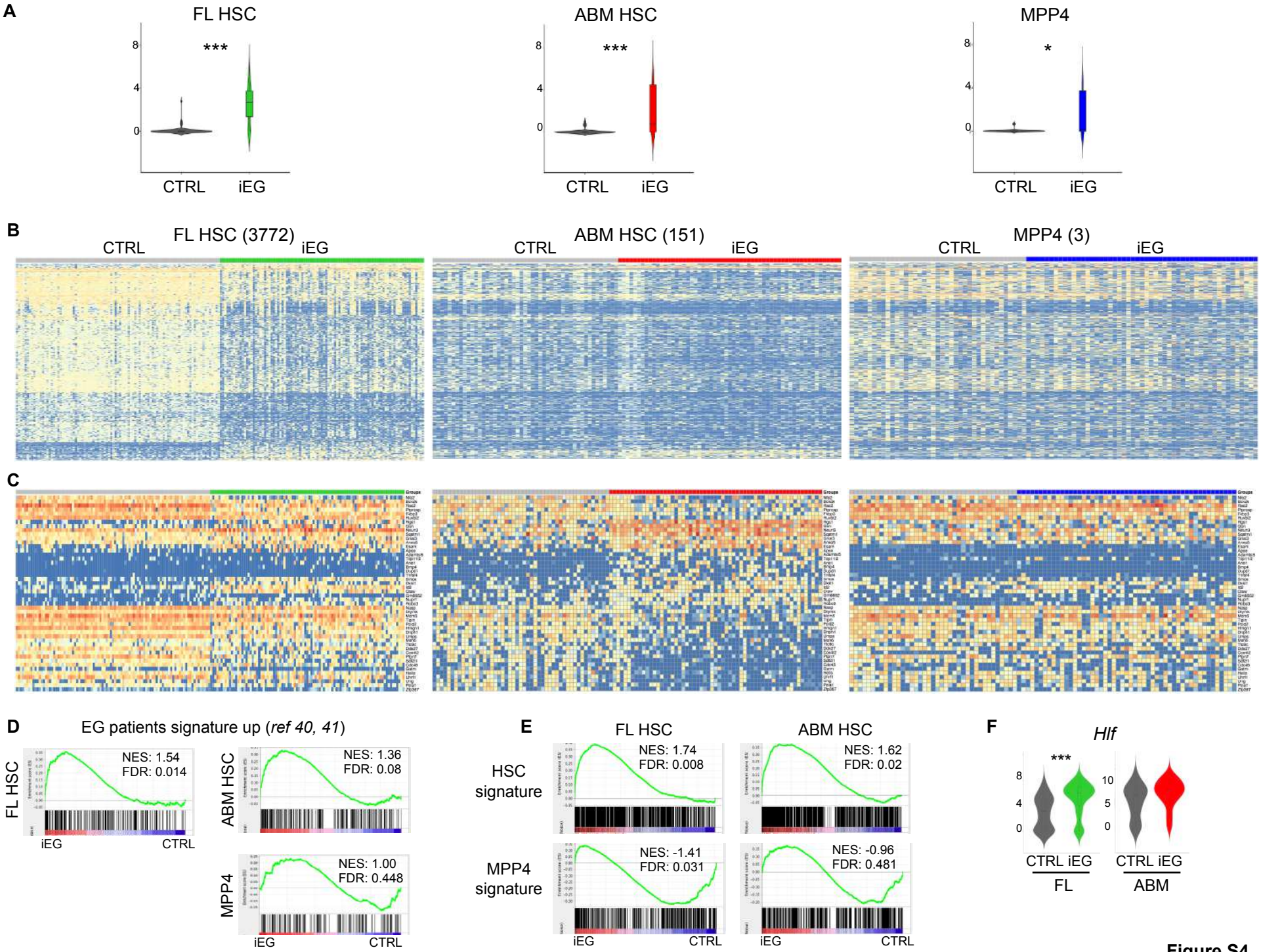
A
FL HSC
ABM HSC
MPP4
***
***
*
8
4
0
CTRL
iEG
B
FL HSC (3772)
ABM HSC (151)
MPP4 (3)
CTRL
iEG
C
D
EG patients signature up (ref 40, 41)
FL HSC
ABM HSC
MPP4
NES: 1.54
FDR: 0.014
NES: 1.36
FDR: 0.08
NES: 1.00
FDR: 0.448
iEG
CTRL
E
FL HSC
ABM HSC
HSC signature
MPP4 signature
NES: 1.74
FDR: 0.008
NES: 1.62
FDR: 0.02
NES: -1.41
FDR: 0.031
NES: -0.96
FDR: 0.481
F
Hlf
***
10
5
0
FL
ABM

Figure S4

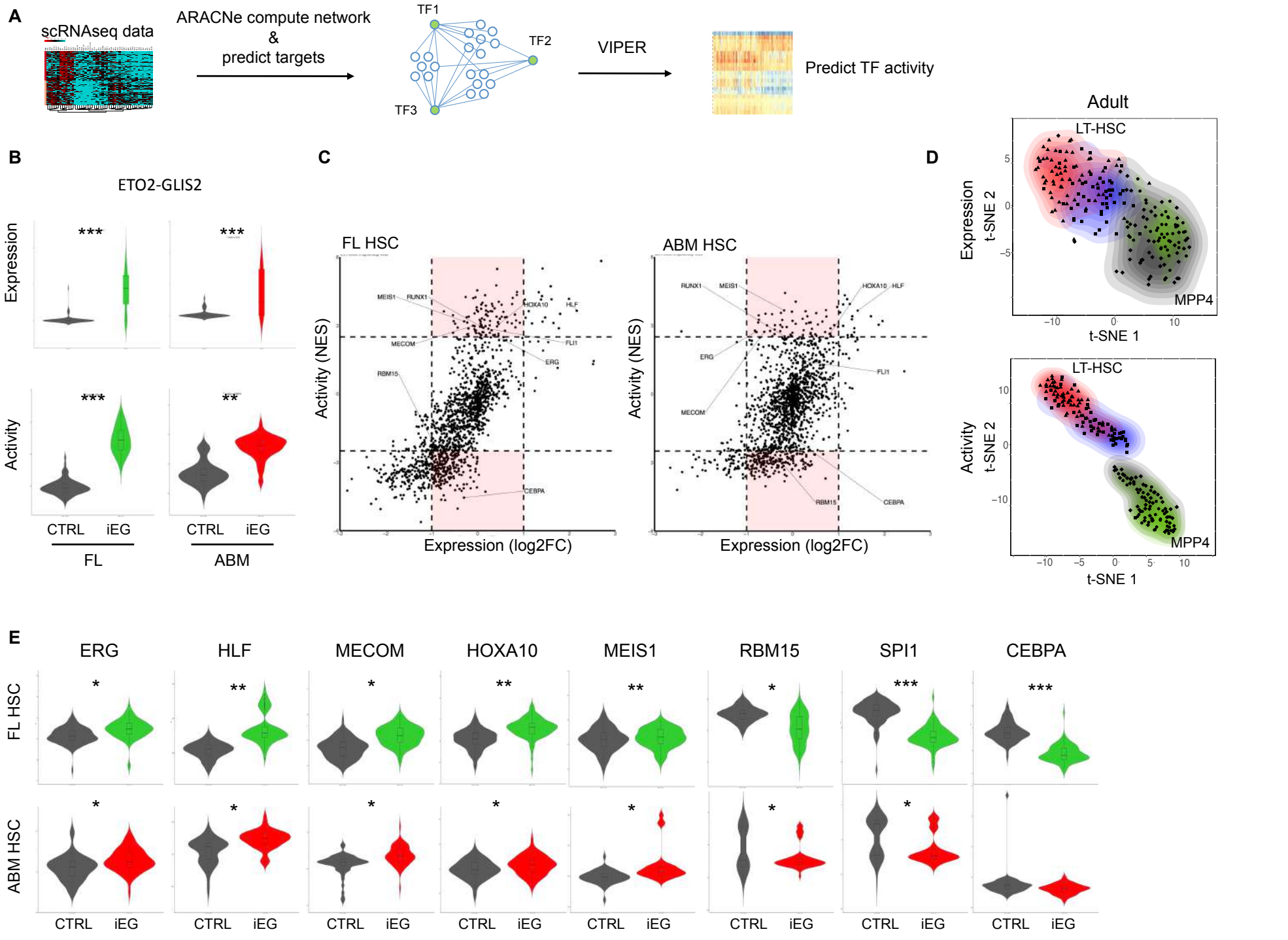
A
scRNAseq data
ARACNe compute network
&
predict targets
TF1
TF2
TF3
VIPER
Predict TF activity
B
ETO2-GLIS2
Expression
Activity
***
***
***
**
CTRL
iEG
CTRL
iEG
FL
ABM
C
FL HSC
ABM HSC
Activity (NES)
Expression (log2FC)
MEIS1
RUNX1
HOXA10
HLF
MECOM
FLI1
ERG
RBM15
CEBPA
D
Adult
LT-HSC
MPP4
Expression
t-SNE 2
t-SNE 1
Activity
E
ERG
HLF
MECOM
HOXA10
MEIS1
RBM15
SPI1
CEBPA
FL HSC
ABM HSC
CTRL
iEG

Figure S5

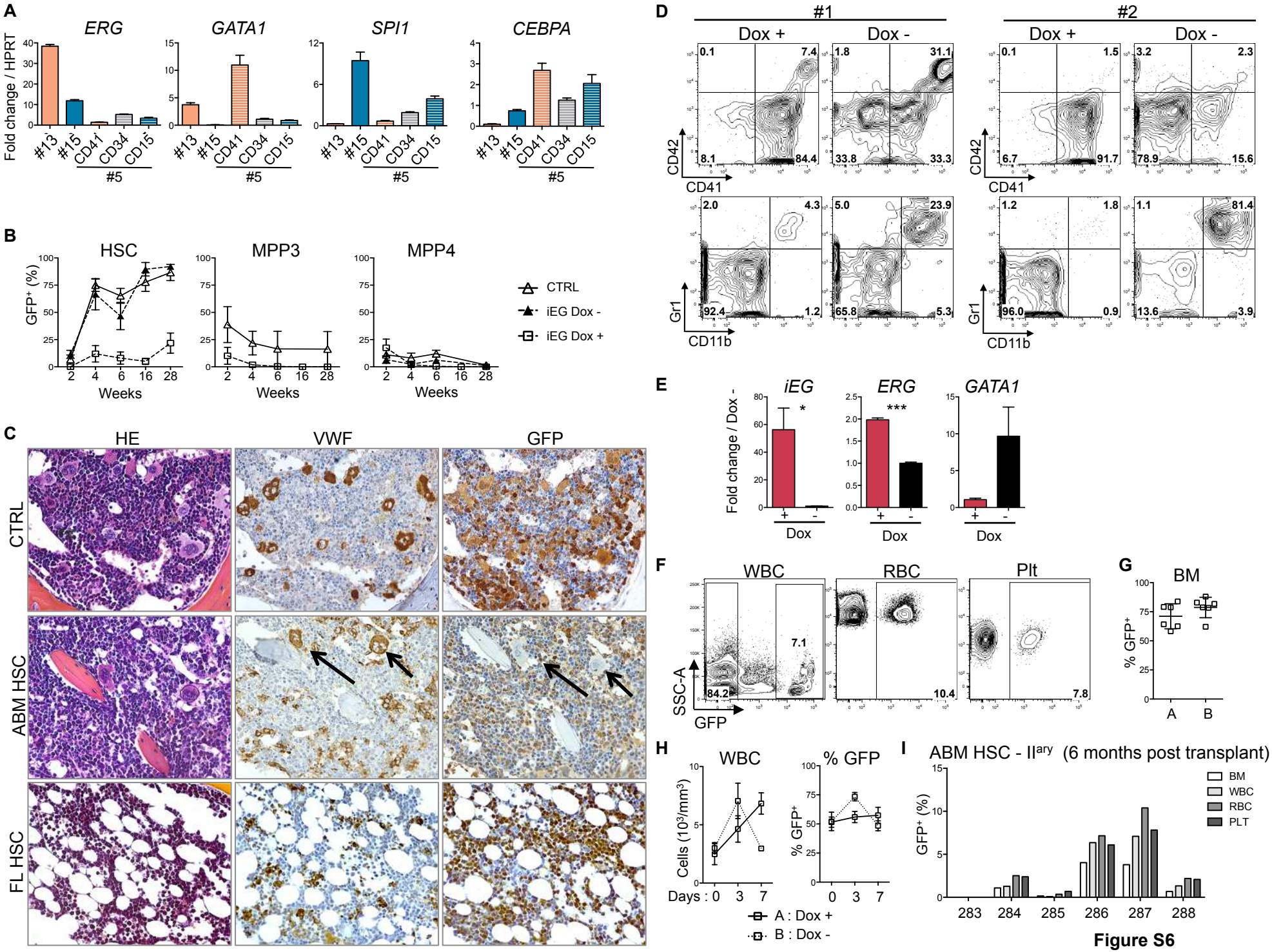
A
ERG
GATA1
SPI1
CEBPA
Fold change / HPRT
#13
#15
CD41
CD34
CD15
#5
B
HSC
MPP3
MPP4
GFP⁺ (%)
Weeks
CTRL
iEG Dox -
iEG Dox +
C
HE
VWF
GFP
CTRL
ABM HSC
FL HSC
D
#1
#2
Dox +
Dox -
CD42
CD41
Gr1
CD11b
E
iEG
ERG
GATA1
Fold change / Dox -
Dox
F
WBC
RBC
Plt
SSC-A
GFP
G
BM
% GFP⁺
A
B
H
WBC
% GFP
Cells (10³/mm³)
Days : 0 3 7
A : Dox +
B : Dox -
I
ABM HSC - II^ary (6 months post transplant)
BM
WBC
RBC
PLT
GFP⁺ (%)

Figure S6

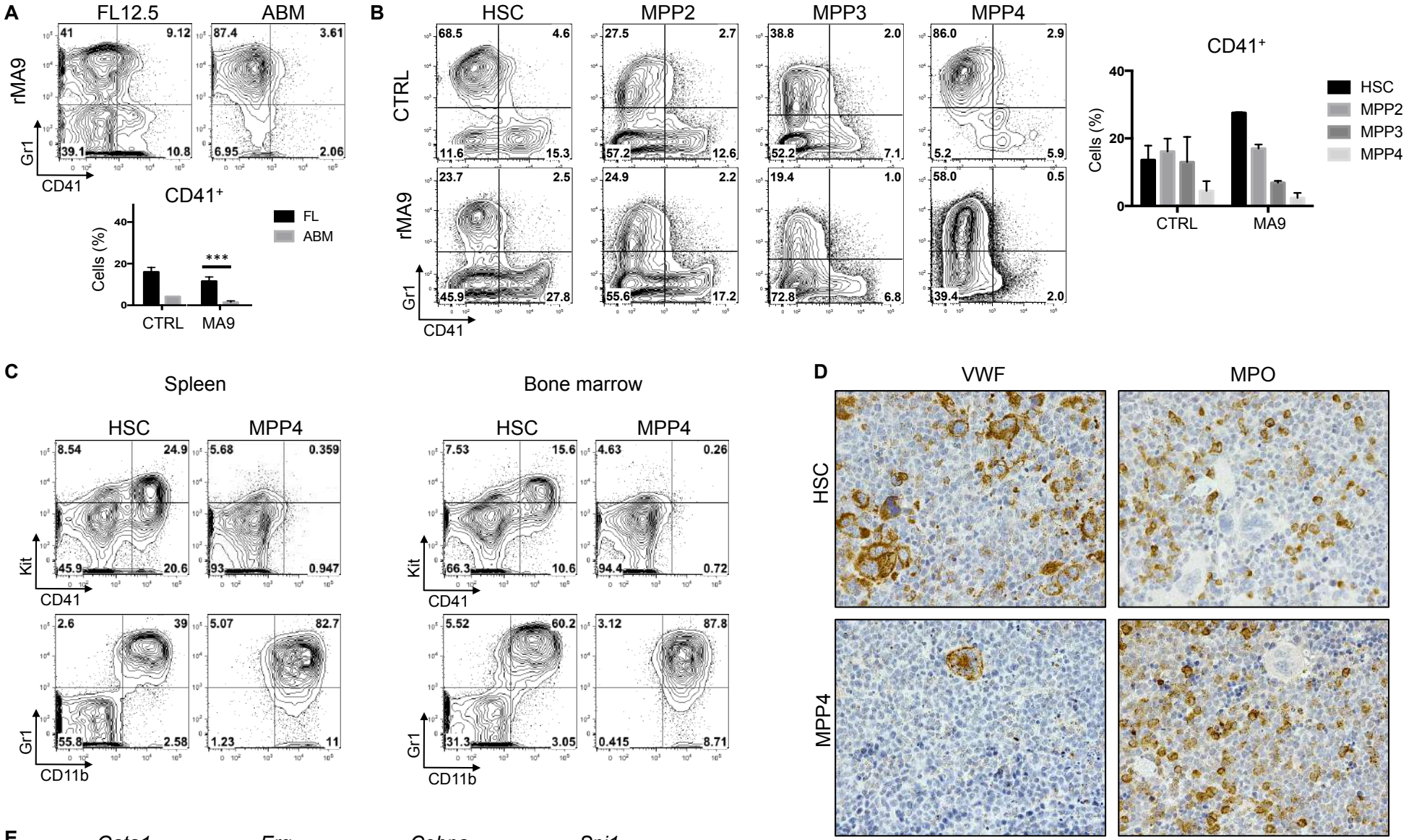
A
FL12.5
ABM
rMA9
Gr1
CD41
41
9.12
39.1
10.8
87.4
3.61
6.95
2.06
CD41+
Cells (%)
FL
ABM
***
CTRL
MA9
B
HSC
MPP2
MPP3
MPP4
CTRL
rMA9
68.5
4.6
11.6
15.3
27.5
2.7
57.2
12.6
38.8
2.0
52.2
7.1
86.0
2.9
5.2
5.9
23.7
2.5
45.9
27.8
24.9
2.2
55.6
17.2
19.4
1.0
72.8
6.8
58.0
0.5
39.4
2.0
CD41+
Cells (%)
HSC
MPP2
MPP3
MPP4
CTRL
MA9
C
Spleen
Bone marrow
HSC
MPP4
Kit
CD41
Gr1
CD11b
8.54
24.9
45.9
20.6
5.68
0.359
93
0.947
2.6
39
55.8
2.58
5.07
82.7
1.23
11
7.53
15.6
66.3
10.6
4.63
0.26
94.4
0.72
5.52
60.2
31.3
3.05
3.12
87.8
0.415
8.71
D
VWF
MPO
HSC
MPP4
E
Gata1
Erg
Cebpa
Spi1
Fold change / PPIA
P1
CD41
Gr1

Figure S7