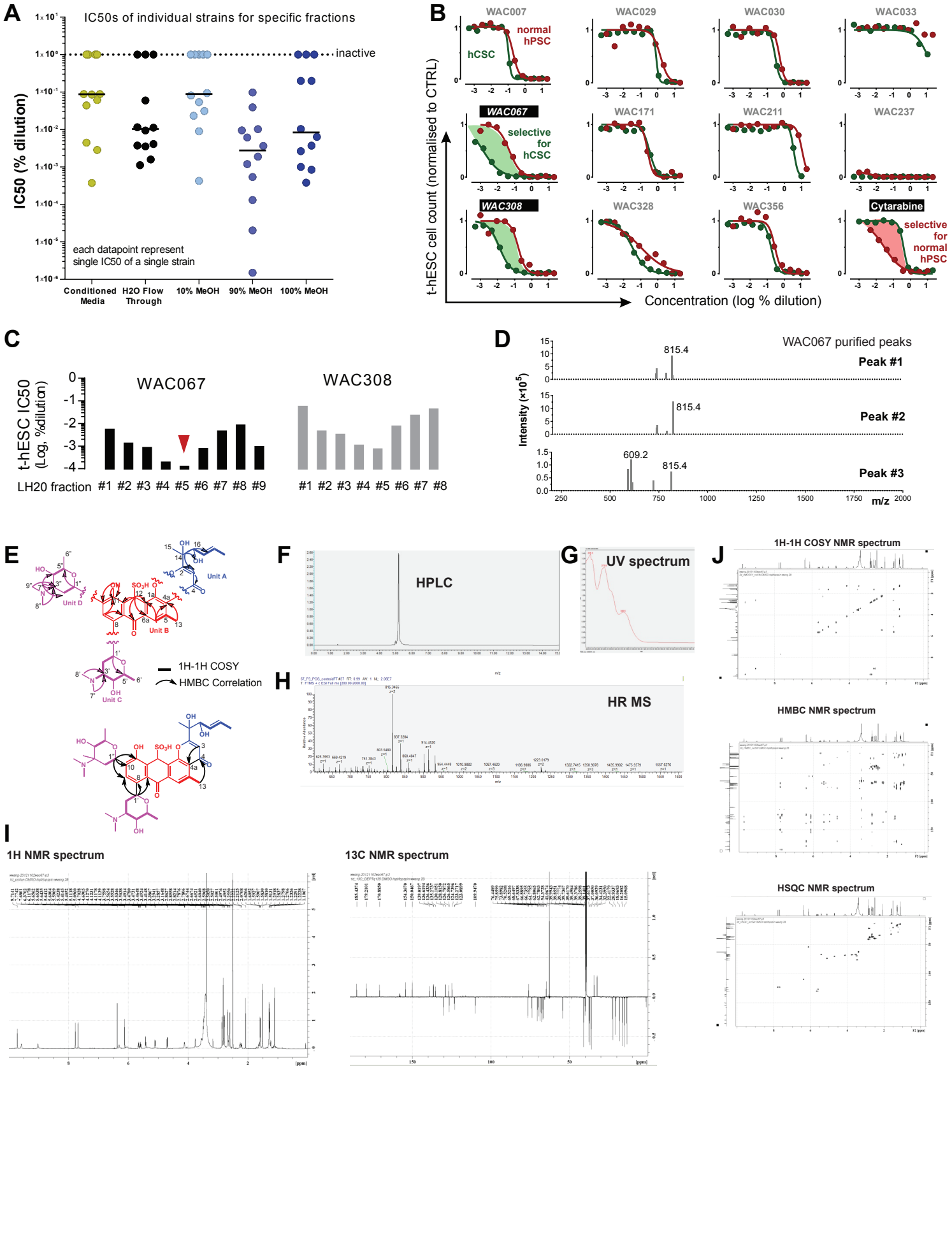
Targeting SUMOylation dependency in human cancer stem cells through a unique SAE2 motif revealed by chemical genomics

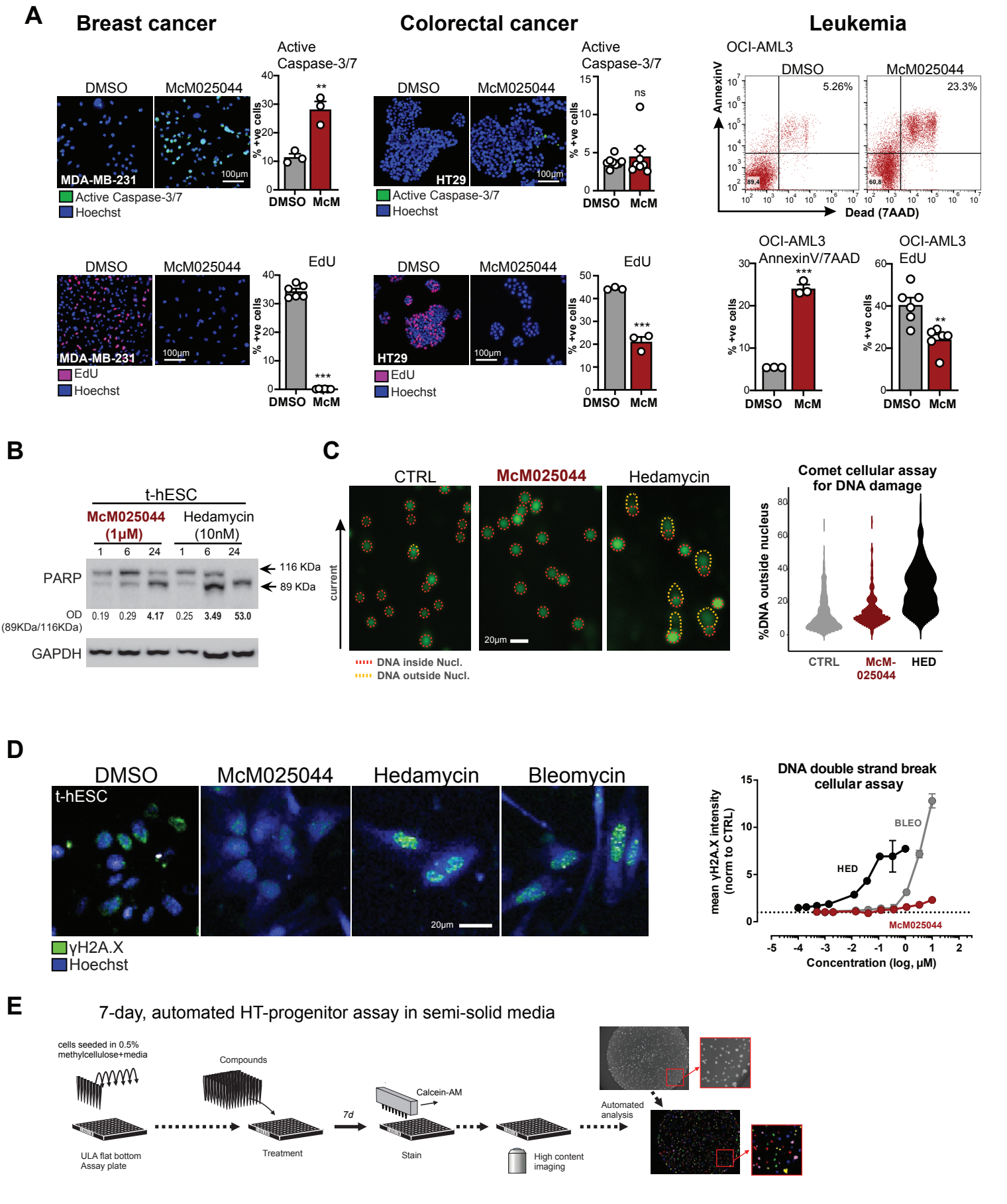
Yannick D. Benoit, Ryan R. Mitchell, Wenliang Wang, Luca Orlando, Allison L. Boyd, Borko Tanasijevic, Lili Aslostovar, Zoya Shapovalova, Meaghan Doyle, Christopher J. Bergin, Kinga Vojnits, Fanny L. Casado, Justin Di Lu, Deanna P. Porras, Juan Luis García-Rodriguez, Jennifer Russell, Aïcha Zouggar, Angelique N. Masibag, Cody Caba, Kalinka Koteva, Lakshmana Kinthada, Jagdish Suresh Patel, Sara N. Andres, Jakob Magolan, Tony J. Collins, Gerard D. Wright, and Mickie Bhatia.

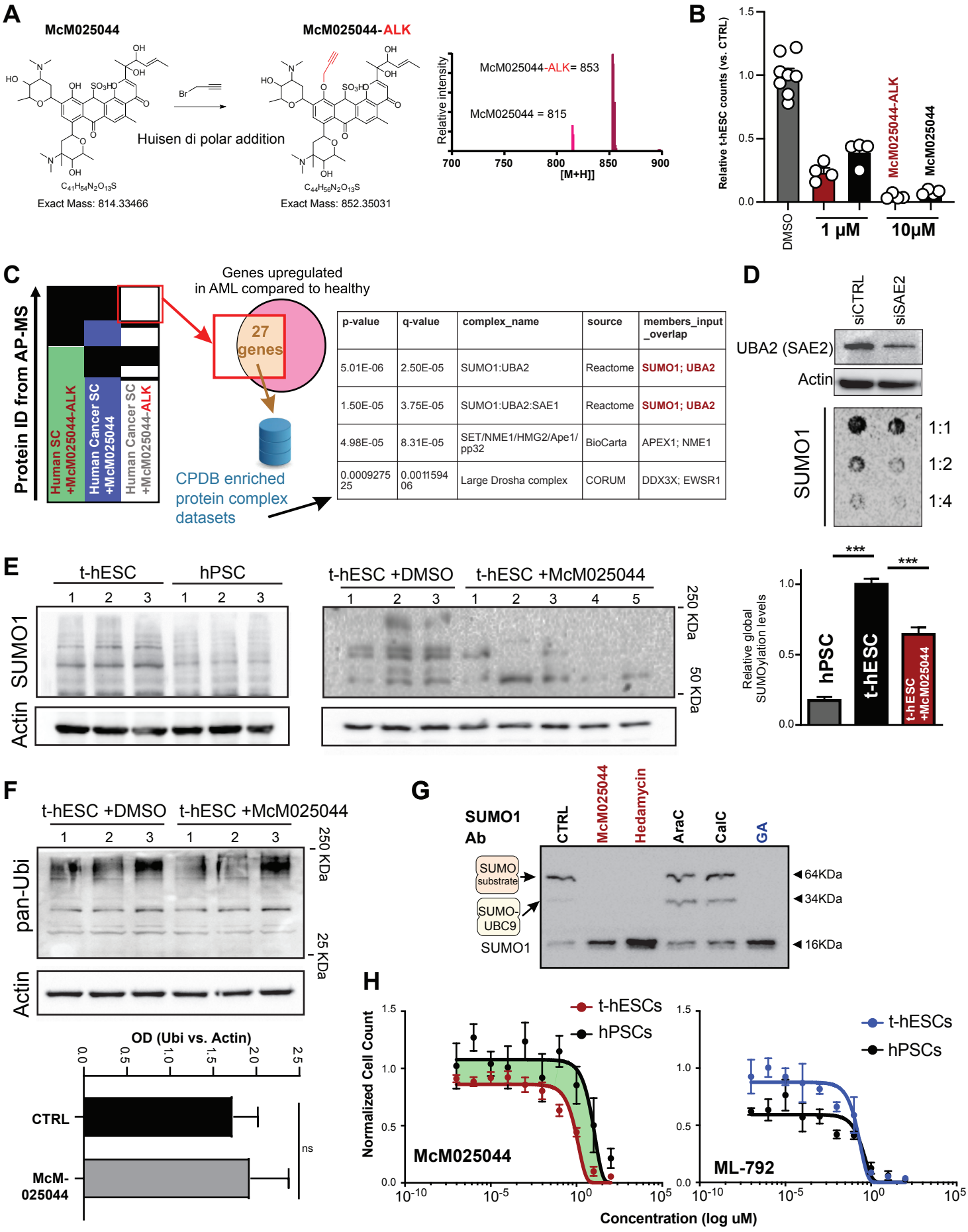
Supplemental Information

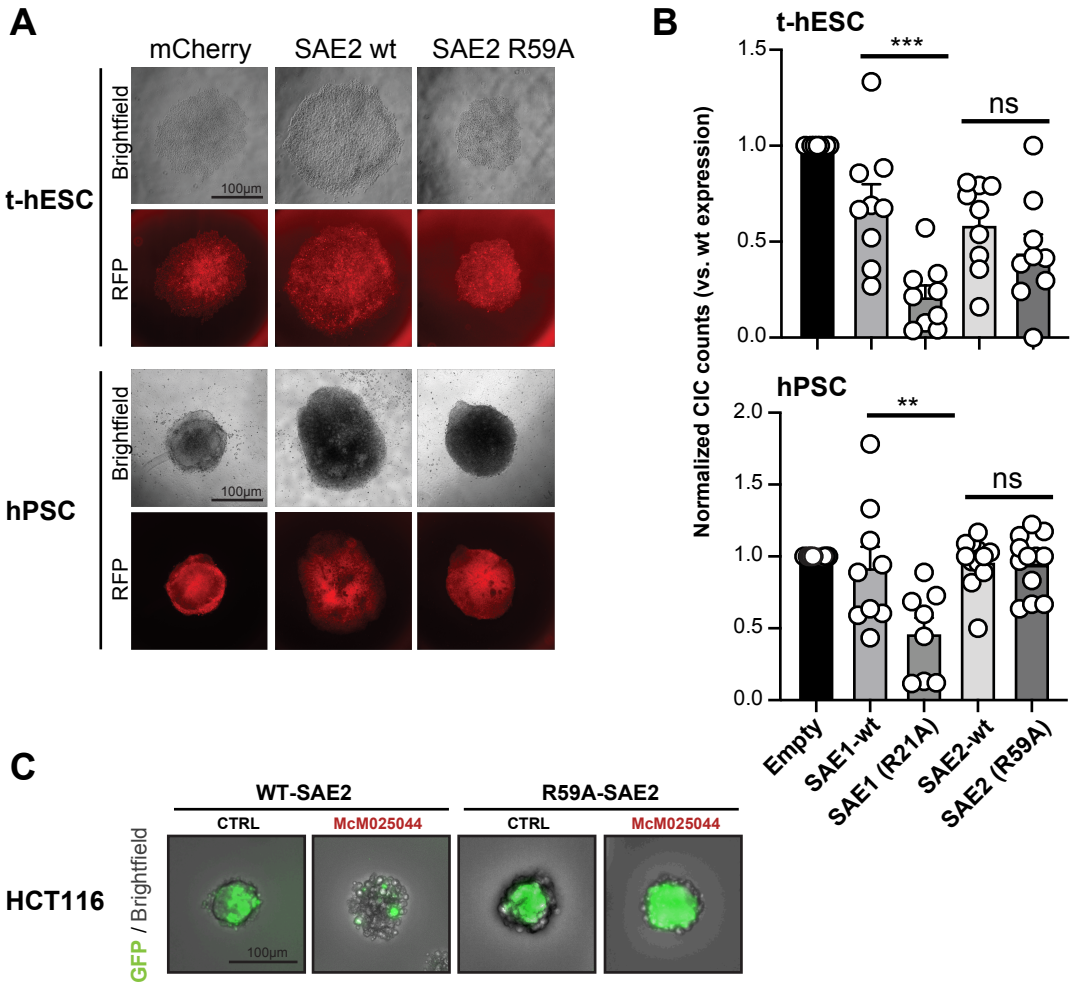
- **Supplemental Figures S1 – S5**
- **Supplemental Figure Legends**

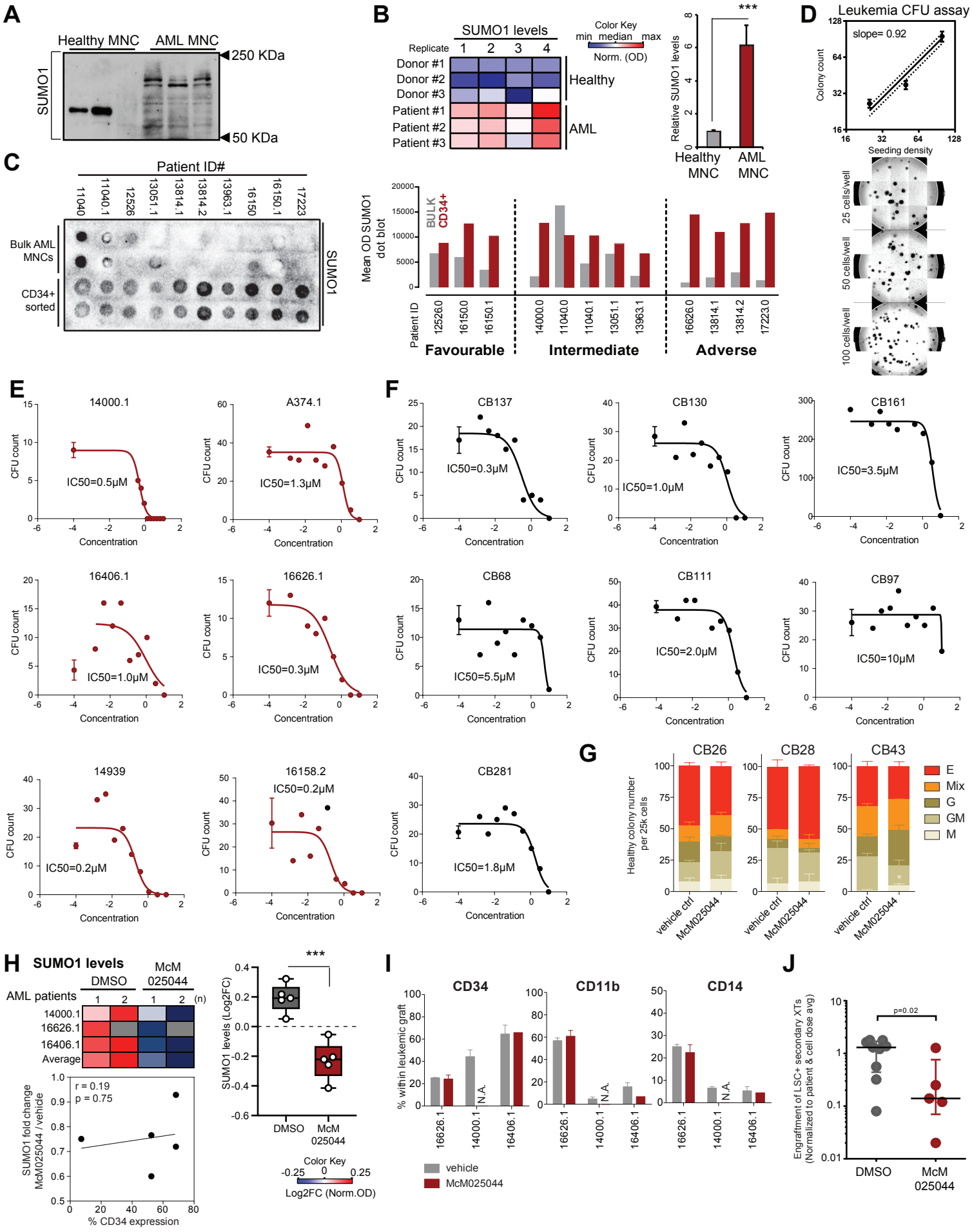
Benoit et al. Supplemental Figure 1




A
Breast cancer
Colorectal cancer
Leukemia
DMSO
McM025044
Active Caspase-3/7
MDA-MB-231
HT29
100µm
Hoechst
% +ve cells
DMSO McM
EdU
OCI-AML3
AnnexinV
5.26%
23.3%
Dead (7AAD)
AnnexinV/7AAD
B
t-hESC
McM025044 (1µM)
Hedamycin (10nM)
1 6 24 1 6 24
PARP
116 KDa
89 KDa
OD (89KDa/116KDa)
0.19 0.29 4.17 0.25 3.49 53.0
GAPDH
C
CTRL
McM025044
Hedamycin
current
20µm
DNA inside Nucl.
DNA outside Nucl.
Comet cellular assay for DNA damage
%DNA outside nucleus
CTRL
McM-025044
HED
D
DMSO
McM025044
Hedamycin
Bleomycin
t-hESC
γH2A.X
Hoechst
DNA double strand break cellular assay
mean γH2A.X intensity (norm to CTRL)
BLEO
HED
McM025044
Concentration (log, µM)
E
7-day, automated HT-progenitor assay in semi-solid media
cells seeded in 0.5% methylcellulose+media
Compounds
7d
Calcein-AM
Automated analysis
ULA flat bottom Assay plate
Treatment
Stain
High content imaging


A
McM025044
McM025044-ALK
Huisen di polar addition
Exact Mass: 814.33466
Exact Mass: 852.35031
McM025044-ALK= 853
McM025044 = 815
Relative intensity
[M+H]]
B
Relative t-hESC counts (vs. CTRL)
DMSO
1 µM
10µM
McM025044-ALK
McM025044
C
Protein ID from AP-MS
Human SC +McM025044-ALK
Human Cancer SC +McM025044
Human Cancer SC +McM025044-ALK
Genes upregulated in AML compared to healthy
27 genes
CPDB enriched protein complex datasets
p-value
q-value
complex_name
source
members_input_overlap
5.01E-06
2.50E-05
SUMO1:UBA2
Reactome
SUMO1; UBA2
1.50E-05
3.75E-05
SUMO1:UBA2:SAE1
Reactome
SUMO1; UBA2
4.98E-05
8.31E-05
SET/NME1/HMG2/Ape1/pp32
BioCarta
APEX1; NME1
0.000927525
0.001159406
Large Drosha complex
CORUM
DDX3X; EWSR1
D
siCTRL
siSAE2
UBA2 (SAE2)
Actin
SUMO1
1:1
1:2
1:4
E
t-hESC
hPSC
t-hESC +DMSO
t-hESC +McM025044
SUMO1
Actin
250 KDa
50 KDa
Relative global SUMOylation levels
***
***
hPSC
t-hESC
t-hESC +McM025044
F
t-hESC +DMSO
t-hESC +McM025044
pan-Ubi
Actin
250 KDa
25 KDa
OD (Ubi vs. Actin)
CTRL
McM-025044
ns
G
SUMO1 Ab
CTRL
McM025044
Hedamycin
AraC
CalC
GA
SUMO substrate
SUMO-UBC9
SUMO1
64KDa
34KDa
16KDa
H
Normalized Cell Count
t-hESCs
hPSCs
McM025044
ML-792
Concentration (log uM)


A
mCherry
SAE2 wt
SAE2 R59A
t-hESC
Brightfield
RFP
hPSC
100µm
B
t-hESC
hPSC
Normalized CIC counts (vs. wt expression)
***
**
ns
Empty
SAE1-wt
SAE1 (R21A)
SAE2-wt
SAE2 (R59A)
C
WT-SAE2
R59A-SAE2
CTRL
McM025044
HCT116
GFP / Brightfield


A
Healthy MNC
AML MNC
SUMO1
250 KDa
50 KDa
B
SUMO1 levels
Replicate
1 2 3 4
Donor #1
Donor #2
Donor #3
Patient #1
Patient #2
Patient #3
Healthy
AML
Color Key
min median max
Norm. (OD)
Relative SUMO1 levels
***
Healthy MNC
AML MNC
C
Patient ID#
11040
11040.1
12526
13051.1
13814.1
13814.2
13963.1
16150
16150.1
17223
Bulk AML MNCs
CD34+ sorted
SUMO1
Mean OD SUMO1 dot blot
BULK
CD34+
Patient ID
12526.0
16150.0
16150.1
14000.0
11040.0
11040.1
13051.1
13963.1
16626.0
13814.1
13814.2
17223.0
Favourable
Intermediate
Adverse
D
Leukemia CFU assay
slope= 0.92
Colony count
Seeding density
25 cells/well
50 cells/well
100 cells/well
E
14000.1
A374.1
16406.1
16626.1
14939
16158.2
IC50=0.5μM
IC50=1.3μM
IC50=1.0μM
IC50=0.3μM
IC50=0.2μM
IC50=0.2μM
CFU count
Concentration
F
CB137
CB130
CB161
CB68
CB111
CB97
CB281
IC50=0.3μM
IC50=1.0μM
IC50=3.5μM
IC50=5.5μM
IC50=2.0μM
IC50=10μM
IC50=1.8μM
G
CB26
CB28
CB43
Healthy colony number per 25k cells
E
Mix
G
GM
M
vehicle ctrl
McM025044
H
SUMO1 levels
DMSO
McM 025044
AML patients
(n)
14000.1
16626.1
16406.1
Average
SUMO1 fold change McM025044 / vehicle
r = 0.19
p = 0.75
% CD34 expression
SUMO1 levels (Log2FC)
***
DMSO
McM 025044
Color Key
Log2FC (Norm.OD)
I
CD34
CD11b
CD14
% within leukemic graft
N.A.
16626.1
14000.1
16406.1
vehicle
McM025044
J
Engraftment of LSC+ secondary XTs (Normalized to patient & cell dose avg)
p=0.02
DMSO
McM 025044

Supplemental Figure Legends

Figure S1. Identification of an anti-cancer stem cell lead candidate from crude natural product libraries. Related to Figure 1.

A) The 12 reproducibly active CM samples were fractionated by adsorption on Diaion HP20 followed by stepwise elution with increasing methanol concentrations and their potency in inhibiting proliferation of t-hESCs was quantified. The 90% methanol fraction was found to be typically the most potent fraction.

B) The cancer-cell selectivity of the 90% methanol fractions was evaluated for 11 out of 12 strains by comparing the IC50s of t-hESC (green) vs normal hPSC (red). Two strains (WAC067 and WAC308) were identified that had cancer-cell selectivity (green shading). In contrast, front line AML chemotherapy Cytarabine has an undesirable higher potency against normal hPSC over t-hESC (red shading).

C) The relative potency of (t-hESC IC50 %dilution) of LH20 sub-fractions of the 90% methanol fraction were determined. Fraction #5 of WAC067 exhibited the highest potency against t-hESC.

D) Mass spectrometry was used to determine the molecular weight of isolated peaks. Peaks 1, 2 and 3 from WAC067, 90% MeOH fraction, LH-20 fraction 5 are shown.

E) The key ^{1}H-^{1}1H COSY and HMBC correlations of McM025044.

F) The HPLC trace of McM025044.

G) The UV absorption of McM025044.

H) The HR ESI MS of McM025044.

I) The ^{1}H-NMR and ^{13}C-NMR spectra of McM025044.

J) The ^{1}H-^{1}1H COSY, HMBC and HSQC NMR spectra of McM025044. We were unable to crystallize McM025044 to assign the stereochemical configurations at C12, C14, and C16. The stereochemistry around the aminoglycoside rings is inferred based on the comparison of our NMR data to previously characterized NPs.

Figure S2. McM025044 shows anti-neoplastic activity in a broad range of human tumor models. Related to Figure 2.

A) Activated caspase-3/7 staining and outer plasma membrane AnnexinV detection was performed on leukemia (OCI-AML3, n=3), colorectal (HT29, n=8), and breast (MDA-MB-231, n=3) cancer cells to assess apoptosis in response to McM025044 treatment (vs. DMSO). Representative micrographs are composites of CellEvent 488 (green), and Hoechst (blue) channels. Proliferation rates were measured for the same experimental conditions, using EdU incorporation analysis (MDA-MB-231, n=6; HT29, n=3; OCI-AML3, n=5; **: $p \leq 0.0056$, ***: $p < 0.0001$). All data are represented as mean $\pm$ SEM. Representative micrographs are composites of iFluor 647 (red), and Hoechst (blue) channels. Scale bar: 100μm.

B) Western blot analysis of PARP cleavage in McM025044 (1 μM) and hedamycin (10 nM) treated t-hESCs. Native (116KDa) and caspase-cleaved (89KDa) fragments of PARP are indicated on the blot, and ratio of optical density (OD) signal for both fragments in presented (caspase-cleaved / native). PARP cleavage was assessed at 1, 6, and 24 hours of drug treatment and GAPDH was used as a loading control.

C) Representative images of single-cell gel electrophoresis (Comet) assays performed on t-hESC, treated for 16 hours with DMSO (n=48), McM025044 (1 μM, n=51), and hedamycin (10 nM, n=48). Red and yellow dashed lines are respectively delineating DNA inside and DNA outside of nuclear area, as quantified by high content imaging. Scale bar: 20μm. Size of comet tail (DNA outside of nucleus) is presented in violin plots of comet assay indicating DNA damage induction.

D) Representative composite images of H2A.X gamma phosphorylation (γH2A.X, green) immunostaining and Hoechst (blue) performed on t-hESC, treated with DMSO, 1 μM McM025044, 0.04 μM hedamycin, and 1 μM bleomycin (Double strand break-inducing positive control). Scale bar: 20μm. Quantitative analysis of H2A.X gamma-phosphorylation levels (Double-strand break indicator) in McM025044, hedamycin, and bleomycin-treated t-hESC. Experiments were performed in dose-response format (n=3, 24h treatment) and presented as relative to vehicle control (Dashed line).

E) Automated high-throughput progenitor assay in semisolid media used to assess human breast and colon cancer cell lines sphere formation capacity in response to McM025044 treatment.

Figure S3. McM025044 is targeting the SUMOylation pathway in cancer stem cells through interaction with SAE2. Related to Figure 3.

A) Schematic representation of OH to alkyne substitution performed on McM025044. MS spectrum reveals MW shift of the alkyne-tagged entity (McM025044-ALK) vs. parent molecule.

B) Cell growth experiments confirming preserved bioactivity of McM025044-ALK vs. McM025044 on t-hESCs (48hrs). Relative cell counts vs. DMSO controls were presented as mean ± SEM (DMSO: n=12, McM025044 and McM025044-ALK: n=4).

C) SUMO1 and SAE2 (UBA2) were identified as candidate targets for McM025044 through affinity purification using alkyne-tagged McM025044; followed by correlating candidate proteins purified exclusively in t-hESCs (red box) with upregulated genes in AML (compared to healthy control cells). Candidate genes were analysed for protein complexes with SUMO1:SAE2 emerging as the lead candidates. Output from CPDB database identifying SUMO1:SAE2 complex as the lead candidate target for McM025044.

D) Western and dot blot analysis confirmed a global decrease of SUMO1 levels in SAE2 (UBA2) knockdown AML cells 72h post transfection. Actin was used as loading control.

E) Western blot analysis of SUMOylated proteins (SUMO1 conjugated) in t-hESCs (n=3) vs. normal hPSCs (n=3) (left panel), and DMSO (n=3) and McM025044-treated (n=5) human CSCs. Actin was used as loading control. Bar graph illustrates relative amounts of global SUMO protein conjugation in normal hPSCs, human CSCs, and McM025044-treated (1 μM, 48hrs) human CSCs ($p \leq 0.0001$). Data are represented as mean ± SEM.

F) Western blot analysis of ubiquitinated proteins (Ub-conjugated) in t-hESCs treated with DMSO (n=3) and McM025044 (1 μM, 48hrs) (n=5). Actin was used as loading control. Bar graph illustrates relative amounts of global Ub protein conjugation in DMSO and McM025044-treated human CSCs (mean ± SEM).

G) Cell-free biochemical assay shows inhibition of protein target (RanGap1) SUMOylation in the presence of MCM25044 (0.5mM), hedamycin (0.5mM), Ara-C (0.5mM), calphostin C (0.5mM), and GA (0.5mM) vs. DMSO control (CTRL).

H) Cancer-cell selectivity of McM025044 (upper panel) and E1 inhibitor ML-792 was evaluated by comparing growth inhibition in dose-response experiments using t-hESC vs. normal hPSC. Only McM025044 was shown with cancer-cell selectivity (green shading), while ML-792 was equally toxic in hPSC vs. t-hESC ($n \geq 3$).

Figure S4. McM025044 binding to SAE2 is inhibiting SUMOylation and stem cell self-renewal. Related to Figure 4.

A) Representative brightfield and RFP filter micrographs of colonies induced from wt-SAE2 and R59A-SAE2 transduced t-hESCs and hPSCs cells. mCherry reporter empty vector was used as control. Scale bar: 100μm.

B) Assessment of CIC capacity for wildtype (wt) SAE1 and SAE2 versus R21A-SAE1 and R59A-SAE2 mutants overexpression in t-hESCs and hPSCs. In both cases, R59A-SAE2 showed no significant changes in the CIC capacity vs. wt, while over-expression of SAE1 R21A resulted in decreased CIC capacity compared to wt SAE1 (t-hESC: n=9, hPSC: $n \geq 8$). Empty vector transduction was presented as control. Data are represented as mean $\pm$ SEM (**: p=0.0032, ***: p=0.0007).

C) Representative brightfield and GFP (green) composite micrographs of spheres induced from wt-SAE2 and R59A-SAE2 transduced HCT116 human colorectal cancer cells. Images were acquired by high-content imaging following 5-day McM025044 treatments (5μM) vs. control DMSO. Scale bar: 100μm.

Figure S5. McM025044 selectively inhibits leukemia stem/progenitor function over healthy hematopoietic cells. Related to Figure 5.

A) Western blot analysis of SUMOylated proteins (SUMO1 conjugated) in MNCs from 3 healthy blood donors and 3 AML independent patients. Equal amounts of total protein lysate were loaded per tracks.

B) Heat map of total SUMO1 signal detected from each healthy donor and AML patient (Red: High, Blue: low, White: median value). Measurements were performed as 4 technical replicates for each primary sample. All combined replicates were presented as relative SUMO1 levels (n=3 patients, 4 replicates per patient, mean $\pm$ SEM, ***: p<0.0001).

C) Dot blot analysis of total SUMO1 abundance in bulk (unsorted) and CD34+-enriched MNC fractions from 10 independent AML patients. Total protein content for each MNC lysates was determined from actin western blot quantification. Equal amount of total protein content was loaded per dot. Mean SUMO1 OD values were scored in bulk and CD34+ fractions and patients were classified according to ELN risk stratification (Favourable, Intermediate, Adverse).

D) Relationship between colony forming capacity versus seeding cell density was evaluated for leukemia cells (OCI-AML3). Data points represent individual replicates per cell dose (25, 50 and 100 cells per well).

Dotted lines represent 95% confidence intervals. ***, P<0.0001, linear regression. Full-well images are tiled micrographs of representative calcein-stained wells of 100, 50 and 25 cells/well seeding.

E-F) Ten-point dose response curves used to derive IC50 values on AML primary patient **(E)** and healthy donor **(F)** samples presented in Figure 5D and E.

G) Healthy cord blood-derived colony scoring for lineage type based on colony morphology. E: erythroid, Mix: mixed-lineage, G: granulocytic, GM: granulocytic/macrophage, M: macrophage. Data are represented as mean ± SEM (n=3 donor samples, *: p<0.05).

H) Heat map of total SUMO1 signal detected from vehicle and McM025044-treated (1μM, 24hrs) AML patient samples prior to *in vivo* injection (Red: High, Blue: low, White: median value, Grey: not tested). Measurements were performed as 2 technical replicates for 2 out of 3 each primary sample. All combined replicates were presented as L2FC relative to the group average (n=3 patients, ***: p<0.0001). Linear regression analysis could not establish a correlation between McM025044-induced decrease of total SUMO1 and the percentage of CD34-positive cells in each individual AML patient samples.

I) Immunophenotypic analysis of human leukemic grafts in primary xenograft recipients performed by flow cytometry. Patient 16626.1: control n=4, McM025044 n=5; Patient 16406.1: control n=3, McM025044 n=1; Patient 14000.1: control n=3, McM025044 "N.A.", not applicable (no engrafted mice). Data are represented as mean ± SEM.

J) BM engraftment levels (% chimerism) in secondary recipients showing engraftment upon transplantation with cells isolated from primary transplanted mice of vehicle and McM025044-treated groups (DMSO n=10; McM025044 n=5). Data are represented as mean ± SEM.