Supplementary Materials: First critical repressive H3K27me3 marks in embryonic stem cells identified using designed protein inhibitor

Authors: James D. Moody[a,b,l,2], Shiri Levy[c,d,l], Julie Mathieu[c,d,l], Yalan Xing[c,d,l], Woojin Kim[e,f,g,l], Cheng Dong[h], Wolfram Tempel[h], Aaron Robitaille[d,i], Luke T. Dang[a,b], Amy Ferreccio[c,d], Damien Detraux[c,d], Sonia Sidhu[c,d], Licheng Zhu[h,j], Lauren Carter[b], Chao Xu[h,k], Cristina Valensisi[d,q], Yuliang Wang[d,l], David Hawkins[d,q], Jinrong Min[h,m], Randall Moon[d,i,n], Stuart Orkin[e,f,g,o,p], David Baker[b,c,n,] Hannele Ruohola-Baker[c,d]

Materials and Methods

Brief computational protein design. We used the EpiGraft tool in Rosetta, previously used to design HIV vaccine epitope scaffolds (see main text ref# 12), as follows: Residues 42-60 from the crystal structure of Ezh2 (taken from PDB ID 2qxv)(see main text ref# 13) were iteratively superimposed onto every 19-residue stretch within a set of 34467 monomeric proteins (scaffolds) taken from the Protein Data Bank (PDB)(1) using the Epigraft application in Rosetta (2, see main text red# 12). To be considered for grafting, scaffolds had to match the backbone of the Ezh2 19-residue stretch with a backbone RMSD $\leq$ 1.5 Å and had to allow the matched region to bind the Ezh2-binding pocket of EED (taken from PDB ID 2qxv) in the same position as Ezh2 without introducing significant clashes between the backbone of the scaffold and the backbone or side chains of EED. A total of 306 grafted models in 291 scaffolds were visually screened in PyMOL to remove scaffolds that were not truly monomeric, that required metals to fold, that were over 150 residues in length, that introduced unresolvable clashes with EED, whose endpoints differed too much from the aligned region of Ezh2, or that had large cavities. This yielded 8 candidates (in 6 unique scaffolds) for design.

Scaffold 1wqg (Mycobacterium tuberculosis ribosome recycling factor) consists of a 3-helix bundle with an inserted beta sheet domain at one end of the bundle (3). Homologous scaffold 3lf9 (HIV 4E10 epitode scaffold), based on PDB ID 1ise (Escherichia coli ribosome recycling factor), replaces the beta sheet domain with a 4-Glycine loop (see main text ref# 12, 4). In order to reduce the size of 1wqg and increase its chances for expression in yeast and bacteria, this beta sheet domain was replaced with 4 Glycine residues using the Rosetta Remodel application (5).

Since only side chains and not the backbone of the Ezh2 N-terminus were grafted onto the scaffold proteins, the scaffolds do not orient the grafted side chains exactly as does the N-terminus of Ezh2. This was particularly a problem with Tryptophan 60 from Ezh2, but was also observed with Phenylalanine 42 and Asparagine 45 from Ezh2 and Glutamate 26, Phe 372 Trp 373, and Arg 420 from EED. These incompatible side chain positions were adjusted to the orientations observed in the crystal structure of EED-Ezh2 complex using Foldit (6). A fixed-backbone design step was included to reverse unwanted mutations and allow redesign of other mutations (see main text ref# 12). For each binder-EED complex, a resfile was manually generated that 1) prevented movement or design of manually adjusted residues, 2) forced reversal of mutations to core residues, 3) allowed other mutated residues to design to any residue except Cysteine, Glycine, Histidine, Phenylalanine, Tryptophan, or Tyrosine, 4) removed unwanted Glycine or Cysteine residues from the scaffold, and 5) added additional binding residues taken from Ezh2 in a scaffold-dependent manner (2). This resfile was used by RosettaScripts (see main text ref# 14) to guide side-chain conformational sampling and mutation of selected binder residues in the presence of EED, without backbone flexibility. The score function was additionally biased using a position-specific scoring matrix (pssm) based on alignments of each scaffold with homologous proteins (7). Twenty independent design runs were carried out for each model, consisting of 4 cycles of side-chain conformational sampling followed by minimization of all side-chain and backbone dihedral angles and the rigid body transform between each of the protein chains in the model (see main text ref# 14). The atom-to-atom repulsive term in the Talaris2013 score function was reweighted to 8%, 20%, 50%, and 100% of its standard value during cycles 1, 2, 3, and 4, respectively. Applying iterative cycles of minimization of dihedral angles with a reduced weight on the atom-atom repulsion term tended to overly distort the protein backbone and so atom-atom distance constraints were applied to every pair of Cα atoms within each chain that were within 12 Å of each other in the input crystal structure or model and greater than 7 residues apart in primary sequence. No constraints were applied between atoms residing on different chains in the model to allow free sampling of the rigid body transform between each of the protein chains in the model. This protocol allowed the backbone of each partner in a protein-protein complex to move to better accommodate binding to the other partner while not forcing unrealistic binding interfaces to remain in the bound state. A manually-generated resfile was used during design that allowed surface residues on the binder to change to any amino acid except Cysteine, Glycine, Histidine, Phenylalanine, Proline, Tryptophan, or Tyrosine, allowed binder core residues and all EED residues to sample different side chain conformations without changing their identities, and prevented movement of selected

binding residues. The score function was additionally biased using a position-specific scoring matrix (pssm) based on alignments of each scaffold with homologous proteins (7).

For each design, the best-scoring model according to total score, protein-protein shape complementarity, binding energy, and buried unsatisfied hydrogen bonds was then subjected to greedy optimization 20 times in parallel as previously described (8). The filter criteria used were total score, shape complementarity, binding energy, and buried unsatisfied hydrogen bonds, equally-weighted. The residue positions and identities to be scored during the greedy run were defined using the same resfile used for the surface design run and the score function was additionally biased using the aforementioned pssm (7). For each candidate binder, the amino acid sequences of the models output from each of the 20 greedy runs were aligned in Geneious (Biomatters Inc. San Francisco, CA, USA) and a consensus sequence was determined. The model with the sequence closest to the consensus was modified to have the consensus sequence using Foldit (6). An electrostatics map was created using PyMOL for each modified model and non-complementary charge interactions across the binder-EED interface were flagged. These flagged interactions were manually adjusted in Foldit by mutating the residue on the binder side of the interface to the most appropriate identity and the resulting model was re-evaluated in PyMOL. At this point, N-linked glycosylation sites on the binder were identified by the consensus sequence NXS/T and removed by mutating either the Asparagine or the Serine/Threonine in Foldit. Once all non-complementary charged interactions and N-linked glycosylation sites were corrected, the modeled interface was validated by subjecting the model to side chain conformational sampling and side chain and backbone minimization without any constraints in both the bound and unbound states. None of the 8 candidate designs disassociated from the modeled complex when relaxed in the bound state with no constraints. The validated models were subjected to fragment-based ab initio structure prediction using only their primary amino acid sequences and the top scoring models were compared to the design models (9).

In-depth computational protein design

General information:

ROSETTA software can be downloaded from https://www.rosettacommons.org/ and is available free to academic users. Online documentation can be found at https://www.rosettacommons.org/manuals/archive/rosetta3.5_user_guide/index.html and instructions for RosettaScripts syntax is available at https://www.rosettacommons.org/manuals/archive/rosetta3.3_user_guide/RosettaScripts_Documentation.html.

1. Identifying candidate scaffolds:

Residues 42-60 from the crystal structure of Ezh2 (taken from PDB ID 2qxv)(See main text ref# 12) were iteratively aligned with every possible 19-residue stretch within a set of 34467 monomeric proteins (scaffolds) taken from the Protein Data Bank (PDB) using the Epigraft application in Rosetta (1, 2, main text ref# 14).

A sample commandline to run the Epigraft matching step is given below:

```
/path_to_rosetta/rosetta.intel -paths ../paths.txt -epi_graft -match -rough_match -input_file
<scaffold_list_file> -output_file out_rough -nres_Ab
<number_of_residues_in_target_protein_chain(in this case 352)> -E_align -S_align -
fluidize_takeoff -max_closure_rms 1.5 -fluidize_landing -native_complex
<pdb_structure_of_native_complex> -loop_ranges <ranges file> -compute_cbeta_neighbors -
rough_match_closure_rms 1.5 -max_intra_clash 3000 -max_inter_clash 10000 -termini_residue_skip 4
```

An example “paths.txt” file is given below:

```
Rosetta Input/Output Paths (order essential)
path is first '/', './',or  '../' to next whitespace, must end with '/'
INPUT PATHS:
pdb1                            ./
pdb2                            /net/pdb/
alternate data files            /net/shared/rosetta_database/
fragments                       ./
structure dssp,ssa (dat,jones)  ./
sequence fasta,dat,jones        ./
constraints                     ./
```

```
starting structure            ./
data files                      /scratch/ROSETTA/rosetta_database/
OUTPUT PATHS:
movie                           ./
pdb path                        ./output/
score                           ./
status                          ./
user                            ./
FRAGMENTS: (use '*****' in place of pdb name and chain)
2                                       number of valid fragment files
3                                       frag file 1 size
aa*****03_05.200_v1_3                             name
9                                       frag file 2 size
aa*****09_05.200_v1_3                             name
-------------------------------------------------------------------------
CVS information:
$Revision: 1.10 $
$Date: 2002/07/16 16:47:31 $
$Author: rohl $
-------------------------------------------------------------------------
```

An example “scaffold list file” is given below: (“~~~” denotes portions omitted for brevity).

```
path_to_scaffolds/01/101ma.pdb
path_to_scaffolds/01/102la.pdb
path_to_scaffolds/01/102ma.pdb

~~~

path_to_scaffolds/ra/9rata.pdb
path_to_scaffolds/ra/9rnta.pdb
path_to_scaffolds/ra/9rnta.pdb
```

An example ranges file is given below:

```
loop: 1
full_range: 353 381
nranges: 1
range: 355 373
```

An example of the output from the Epigraft matching step is given below: (“~~~” denotes portions omitted for brevity).

```
a1/2a1jB.pdb           1             E             6          23                   355
373         5:0:-64.951065         5:1:-44.615498          6:0:-72.98307         6:1:-49.676495
23:0:-85.84198        23:1:-5.6372075        24:0:-69.804115        24:1:-39.077366
0.44630823      -0.87222779      -0.20006934       0.89104259       0.41246533       0.18951385
-0.08277757      -0.26285189       0.96127874      -18.22481346       51.68355942      -
30.45912743           1.109              1.220              1.073                 inf
907.089        3054.459        61.607        30.799                  13                    -

~~~

k5/1k53A.pdb           1             S            22          40                   355
373                     -                     -                     -                    -
-                       -                    -                     -      -0.79238760      -
0.34577596       0.50255436       0.55912358      -0.08224291       0.82499516      -0.24393195
0.93470591       0.25849974      115.39505005        8.01579285      -40.18924713           1.095
1.606              2.193              0.058         372.293        8511.128        10.285
22.899                   9                     -
```

2. Grafting key side chains onto candidate scaffolds:

To be considered for grafting, scaffolds had to match the backbone of the Ezh2 19-residue stretch with a backbone RMSD ≤ 1.5 Å and had to allow the matched region to bind the Ezh2-binding pocket of EED (taken from

PDB ID 2qxv) in the same position as Ezh2 without introducing significant clashes between the backbone of the scaffold and the backbone or side chains of EED.

An example formatted list of matches is given below: ("~~~" denotes portions omitted for brevity).

```
1le2_1_rlx.pdb 1 S 73 91 355 373 $ keep_bb graft_sc
1le2_1_rlx.pdb 1 S 77 95 355 373 $ keep_bb graft_sc
1le2_1_rlx.pdb 1 S 84 102 355 373 $ keep_bb graft_sc

~~~

3ef8_1_rlx.pdb 1 S 7 25 355 373 $ keep_bb graft_sc
3f4m_1_rlx.pdb 1 S 24 42 355 373 $ keep_bb graft_sc
3lf9_1_rlx.pdb 1 S 47 65 355 373 $ keep_bb graft_sc
```

An example commandline to run the Epigraft grafting step is given below:

```
/path_to_rosetta/rosetta.intel -epi_graft -multigraft -nres_Ab 352 -native_complex <pdb structure
of the native complex> -loop_ranges <ranges filename> -input_file <formatted_list_of_matches> -
use_non_monotone_line_search -atom_vdw_set highres -ex1 -ex1aro -ex2 -extrachi_cutoff 0 -
try_both_his_tautomers -dump_predesign <output_design_filename> -output_file
<output_data_filename> -Ab_epitope_optimize -repack_Ab -design_attempts 10 -store_n_best_designs
1 -paths ./paths.txt -keep_natro <list_of_residues_to_graft> -design_after_closure
```

An example list of residues to graft is given below:

```
355
358
362
365
366
369
373
```

3. Screening Epigraft match output for candidate designs:

A total of 306 grafted models in 291 scaffolds were visually screened in PyMOL to remove scaffolds that were not truly monomeric, that required metals to fold, that were over 150 residues in length, that introduced unresolvable clashes with EED, whose endpoints differed too much from the aligned region of Ezh2, or that had large cavities, yielding 8 candidates (in 6 scaffolds) for design.

4. Remodeling selected scaffolds for improved experimental properties:

Scaffold 1wqg (Mycobacterium tuberculosis ribosome recycling factor)(3) consists of a 3 helix bundle with an inserted beta sheet domain at one end of the bundle. Homologous scaffold 3lf9 (HIV 4E10 epitode scaffold), based on PDB ID 1ise (Escherichia coli ribosome recycling factor), replaces the beta sheet domain with a 4-Glycine loop (see main text ref# 14). In order to reduce the size of 1wqg and increase it's chances for expression in yeast and bacteria, this beta sheet domain was removed in PyMOL and replaced with 4 Glycine residues using the Rosetta Remodel application (5).

An example commandline to run Remodel is given below:

```
/path_to_rosetta/remodel.static.linuxiccrelease -database
/path_to_rosetta_database/rosetta_database/ -s <input_structure> -remodel:blueprint <blueprint
file> -ex1 -ex2 -correct -num_trajectory 1 -remodel:use_blueprint_sequence -use_input_sc -
save_top 1 -chain A -remodel:quick_and_dirty -remodel:use_pose_relax -run_confirmation true -
preserve_header true -use_clusters false -overwrite -score:weights score12prime -hbond_params
sp2_params -corrections:score:hb_sp2_chipen -lj_hbond_hdis 1.75 -lj_hbond_OH_donor_dis 2.6 -
ignore_zero_occupancy false
```

An example blueprint file is given below: ("~~~" denotes portions omitted for brevity).

```
1 I .
2 D .
3 E .

~~~

25 T .
26 I .
27 R .
28 T L PIKAA T
0 x L PIKAA G
0 x L PIKAA G
0 x L PIKAA G
0 x L PIKAA G
29 T L PIKAA T
30 E .
31 E .
32 R .

~~~

106 L .
107 E .
108 V .
```

5. Preparing Epigraft output models for surface design:

Since only select side chains and not the backbone of the Ezh2 N-terminnus are grafted onto the scaffold protein, the scaffold cannot orient the grafted side chains exactly as does the N-terminus of Ezh2. As the Epigraft grafting step automatically adjusts side chain orientations during the grafting step, these residues were sometimes modeled in positions incompatible with binding to EED. This was particularly a problem with Tryptophan 60 from Ezh2, but was also observed with Phenylalanine 42 and Asparagine 45 from Ezh2 and Glutamate 26, Phenylalanine 284, Tryptophan 285, and Arginine 332 from EED. These incompatible side chain positions were manually adjusted to the orientations observed in the crystal structure of EED-Ezh2 complex using Foldit (6).

Further, as the Epigraft grafting step alters the identities of scaffold residues using an older score function, a fixed-backbone design step was included to reverse unwanted mutations and allow redesign of other mutations. For each binder-EED complex, a resfile was manually generated that 1) prevented movement or design of manually adjusted residues, 2) forced reversal of mutations to core residues, 3) allowed other mutated residues to design to any residue except Cysteine, Glycine, Histidine, Phenylalanine, Tryptophan, or Tyrosine, 4) removed unwanted Glycine or Cysteine residues from the scaffold, and 5) added additional binding residues taken from Ezh2 in a scaffold-dependent manner. This resfile was used by RosettaScripts to guide side-chain conformational sampling and mutation of selected binder residues in the presence of EED, without backbone flexibility. The score function was additionally biased using a position-specific scoring matrix (pssm) based on alignments of each scaffold with homologous proteins.

A sample commandline to run the fixed-backbone design step is given below:

```
/path_to_rosetta/rosetta_scripts.static.linuxiccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -ignore_unrecognized_res -
overwrite -out:file:renumber_pdb false -ex1 -ex2 -nstruct 1 -parser:script_vars
resfile=<resfile_filename> pssm=<pssm_filename> -s <input_structure> -parser:protocol
./fixbb_dsn.xml -score:weights talaris2013 \
```

A sample xml protocol to run the fixed-backbone design step is given below:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <sfxn_hard weights=talaris2013/>
        </SCOREFXNS>
        <TASKOPERATIONS>
                <LimitAromaChi2 name=arochi2/>
                <DisallowIfNonnative name=nocys disallow_aas=CGH/>
                <RestrictChainToRepacking name=rctr chain=1/>
```

```
		<ReadResfile name=resfile filename="%%resfile%%"/>
	</TASKOPERATIONS>
	<MOVERS>
		<FavorSequenceProfile name=fsp_pssm scaling=global chain=2 pssm="%%pssm%%"
weight=1.0 scorefxns=sfxn_hard/>
		<AtomTree name=ftree simple_ft=1/>
		<PackRotamersMover name=packdsn scorefxn=sfxn_hard
task_operations=rctr,arochi2,nocys,resfile/>
		<MinMover name=min_bb_sc bb=1 chi=1 jump=0 scorefxn=sfxn_hard>
			<MoveMap name=pin_down_hs>
				<Chain number=2 chi=1 bb=1/>
				<Chain number=1 chi=1 bb=1/>
				<Jump number=1 setting=0/>
			</MoveMap>
		</MinMover>
	</MOVERS>
	<FILTERS>
		<Ddg name=binding_energy threshold=0 scorefxn=sfxn_hard confidence=0 jump=1
repack=1 relax_mover=min_bb_sc repeats=3/>
		<Sasa name=dsasa threshold=500 confidence=0/>
		<ShapeComplementarity name=shape_comp jump=1 verbose=0 min_sc=0.60 confidence=0/>
		<SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>
	</FILTERS>
	<PROTOCOLS>
		<Add mover_name=ftree/>
		<Add mover_name=fsp_pssm/>
		<Add mover_name=packdsn/>
		<Add filter_name=binding_energy/>
		<Add filter_name=dsasa/>
		<Add filter_name=shape_comp/>
		<Add filter_name=unsat/>
	</PROTOCOLS>
</ROSETTASCRIPTS>
```

A sample resfile is used during the fixed-backbone design step given below: ("~~~" denotes portions omitted for brevity).

```
NATAA
START
332 A NATRO
284 A NATRO
285 A NATRO
364 B NOTAA CGHWYF
368 B PIKAA T
394 B PIKAA V
395 B NOTAA CGHWYF
396 B NOTAA CGHWYF
398 B NOTAA CGHWYF
399 B NATRO
400 B NOTAA CGHWYF
403 B PIKAA R
404 B NOTAA CGHWYF
405 B NOTAA CGHWYF
406 B NATRO
407 B NOTAA CGHWYF
409 B NATRO
410 B NATRO
413 B NATRO
414 B NOTAA CGHWYF
417 B NATRO
418 B NOTAA CGHWYF
424 B NOTAA CGHWYF
437 B PIKAA L
441 B NOTAA CGHWYF
444 B PIKAA T
448 B NOTAA CGHWYF
452 B NOTAA CGHWYF
455 B PIKAA L
459 B NOTAA CGHWYF
```

A sample pssm file is given below: ("~~~" denotes portions omitted for brevity).

```
Last position-specific scoring matrix computed, weighted observed percentages rounded down,
information per position, and relative weight of gapless real matches to pseudocounts
           A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V   A   R   N   D   C   Q   E
G   H   I   L   K   M   F   P   S   T   W   Y   V
    1 M   -2 -3 -3 -4 -2 -2 -3 -4 -3  2  2 -2  8 -1 -4 -3 -2 -2 -2  1    0   0   0   0   0   0
0   0   0  12  10   0  75   0   0   0   0   0   0   3  0.78 0.15
    2 I   -2 -4 -4 -4 -2 -3 -4 -5 -4  6  1 -3  1 -1 -4 -3 -1 -3 -2  3    0   0   0   0   0   0
0   0   0  85   3   0   0   0   0   0   2   0   0  10  0.69 0.13
    3 N   -2 -1  7  0 -3 -1 -1 -1  0 -4 -4  0 -3 -4 -3  1  1 -4 -3 -3    0   0  80   0   0   0
1   0   0   0   0   3   0   0   0   6  10   0   0   0  0.90 0.15

~~~

  118 H   -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3    0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0  0.00    0.00
  119 H   -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3    0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0  0.00    0.00
  120 H   -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3    0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0   0   0   0   0  0.00    0.00

                      K         Lambda
Standard Ungapped    0.1301     0.3128
Standard Gapped      0.0410     0.2670
PSI Ungapped         0.1883     0.3179
PSI Gapped           0.0576     0.2670
```

6. Optimizing the identities of surface residues of the designed binders:

The models output from fixed_backbone design were then redesigned 20 times in parallel using the fastdesign_with_atompair constraints protocol. A resfile was manually generated that allowed surface residues on the binder to change to any amino acid except Cysteine, Glycine, Histidine, Phenylalanine, Proline, Tryptophan, or Tyrosine, allowed binder core residues and all EED residues to sample different side chain conformations without changing their identities, and prevented movement of selected binding residues. The score function was additionally biased using a position-specific scoring matrix (pssm) based on alignments of each scaffold with homologous proteins.

A sample commandline used for the surface design step is given below:

```
/path_to_rosetta/rosetta_scripts.default.linuxgccrelease -database
/path_to_rosetta_databse/database -ignore_zero_occupancy false -ignore_unrecognized_res -
overwrite -out:file:renumber_pdb false -ex1 -ex2 -nstruct 1 -parser:script_vars
resfile=<resfile_filename> pssm=<pssm_filename> -s <input_pdb_structure> -parser:protocol
./flxbb_fstdsn_resfile_atompaircst.xml -score:weights talaris2013
```

A sample xml protocol used for the surface design step is given below:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <sfxn_hard weights=talaris2013>
                        <Reweight scoretype=atom_pair_constraint weight=1.0/>
                </sfxn_hard>
        </SCOREFXNS>
        <TASKOPERATIONS>
                <OperateOnCertainResidues name=chainA>
                      <PreventRepackingRLT/>
                        <ChainIs chain=A/>
                </OperateOnCertainResidues>
                <OperateOnCertainResidues name=chainB>
                      <PreventRepackingRLT/>
                        <ChainIs chain=B/>
                </OperateOnCertainResidues>
                <LimitAromaChi2 name=arochi2/>
                <DisallowIfNonnative name=nocys disallow_aas=CGHWYF/>
```

```
		<ProteinInterfaceDesign name=pido repack_chain1=1 repack_chain2=1 design_chain1=0
design_chain2=1 jump=1 interface_distance_cutoff=10 allow_all_aas=1/>
		<RestrictChainToRepacking name=rctr chain=1/>
		<ReadResfile name=resfile filename="%%resfile%%"/>
	</TASKOPERATIONS>
	<MOVERS>
		<FavorSequenceProfile name=fsp_pssm scaling=global chain=2 pssm="%%pssm%%"
weight=1.0 scorefxns=sfxn_hard/>
		<AtomTree name=ftree simple_ft=1/>
		<MinMover name=min_bb_sc bb=1 chi=1 jump=0 scorefxn=sfxn_hard>
			<MoveMap name=minmvrmap>
				<Chain number=2 chi=1 bb=1/>
				<Chain number=1 chi=1 bb=1/>
				<Jump number=1 setting=1/>
			</MoveMap>
		</MinMover>
		<FastRelax name=fstrlx_dsn scorefxn=sfxn_hard repeats=1
task_operations=pido,arochi2,rctr,nocys,resfile>
			<MoveMap name=fstrlxmap>
				<Chain number=2 chi=1 bb=1/>
				<Chain number=1 chi=1 bb=1/>
				<Jump number=1 setting=1/>
			</MoveMap>
		</FastRelax>
		<AddConstraintsToCurrentConformationMover name=add_pair_chainA_cst
use_distance_cst=1 coord_dev=0.5 bound_width=0.1 min_seq_sep=8 max_distance=12.0 cst_weight=1.0
task_operations=chainA/>
		<AddConstraintsToCurrentConformationMover name=add_pair_chainB_cst
use_distance_cst=1 coord_dev=0.5 bound_width=0.1 min_seq_sep=8 max_distance=12.0 cst_weight=1.0
task_operations=chainB/>
		<ClearConstraintsMover name=clear_cst/>
	</MOVERS>
	<FILTERS>
		<Ddg name=binding_energy threshold=0 scorefxn=sfxn_hard confidence=0 jump=1
repack=1 relax_mover=min_bb_sc repeats=3/>
		<Sasa name=dsasa threshold=500 confidence=0/>
		<ShapeComplementarity name=shape_comp jump=1 verbose=0 min_sc=0.60 confidence=0/>
		<SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>
		<ScoreType name=total_score_complex scorefxn=sfxn_hard score_type=total_score
confidence=0 threshold=0/>
	</FILTERS>
	<PROTOCOLS>
		<Add mover_name=ftree/>
		<Add mover_name=fsp_pssm/>
		<Add mover=add_pair_chainA_cst/>
		<Add mover=add_pair_chainB_cst/>
		<Add mover_name=fstrlx_dsn/>
		<Add mover=clear_cst/>
		<Add filter_name=binding_energy/>
		<Add filter_name=dsasa/>
		<Add filter_name=shape_comp/>
		<Add filter_name=unsat/>
		<Add filter_name=total_score_complex/>
	 </PROTOCOLS>
</ROSETTASCRIPTS>
```

A sample resfile used for the surface design step is given below: (“~~~” denotes portions omitted for brevity).

```
NATAA
START
332 A NATRO
284 A NATRO
285 A NATRO
355 B NOTAA CGHWYFP
356 B NOTAA CGHWYFP
358 B NOTAA CGHWYFP

~~~

464 B NOTAA CGHWYFP
```

```
466 B NOTAA CGHWYFP
467 B NOTAA CGHWYFP
399 B NATRO
406 B NATRO
409 B NATRO
410 B NATRO
413 B NATRO
417 B NATRO
```

7. Refining the designed binders using greedy optimization:

For each binder, the best-scoring model according to total score, shape complementarity, binding energy, and buried unsatisfied hydrogen bonds was then subjected to greedy optimization 20 times in parallel. The greedy algorithm individually scores all allowed mutations according to predefined combination of filters and then attempts to incorporate all of the mutations into the design in order of best scoring to worst scoring. A mutation is only incorporated if it improves the combined score of all filter criteria. Otherwise that mutation is skipped (8). Running the greedy algorithm 20 X in parallel allows the time-consuming step of scoring all allowed mutations to be split up among 20 different processors, which then share the resulting data during the incorporation step. The filter criteria used here were total score, shape complementarity, binding energy, and buried unsatisfied hydrogen bonds, where the value of each filter criteria was divided by a factor equal to that filter's value in the input model. The filters total score, shape complementarity, and binding energy were additionally multiplied by a factor of negative 1. These factors had the effect of giving each of the four filters equal weight in determining whether a mutation should be incorporated or skipped during the design step. The residue positions and identities to be scored during the greedy run were defined using the same resfile used for the surface design run and the score function was additionally biased using the aforementioned pssm.

A sample script used to automatically set the factors for the combined filters and launch the greedy MPI run is given below:

```
#!/bin/bash

pssm=<pssm_filename>
resfile=<resfile_filename>
pdb=<input_structure>
total_score=`grep ^total_score $pdb | awk '{print 1/$2*(0-1)}'`
shape=`grep ^shape $pdb | awk '{print 1/$2*(0-1)}'`
unsat=`grep ^unsat $pdb | awk '{print 1/$2}'`
binding=`grep ^binding $pdb | awk '{print 1/$2*(0-1)}'`

mpirun -np 20 /path_to_rosetta/rosetta_scripts.mpi.linuxgccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -ignore_unrecognized_res -
overwrite -out:file:renumber_pdb false -ex1 -ex2 -nstruct 19 -parser:script_vars
resfile=${resfile} pssm=${pssm} total_score=${total_score} shape=${shape} binding=${binding}
unsat=${unsat} -s $pdb -parser:protocol ./greedy_no_min.xml -score:weights talaris2013
```

A sample xml protocol used for the MPI greedy run is given below:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <sfxn_hard weights=talaris2013/>
        </SCOREFXNS>
        <TASKOPERATIONS>
                <LimitAromaChi2 name=arochi2/>
                <DisallowIfNonnative name=nocys disallow_aas=CGHWYFP/>
                <ProteinInterfaceDesign name=pido repack_chain1=1 repack_chain2=1 design_chain1=0
design_chain2=1 jump=1 interface_distance_cutoff=10 allow_all_aas=1/>
                <RestrictChainToRepacking name=rctr chain=1/>
                <ReadResfile name=resfile filename="%%resfile%%"/>
        </TASKOPERATIONS>
        <MOVERS>
                <FavorSequenceProfile name=fsp_pssm scaling=global chain=2 pssm="%%pssm%%"
weight=1.0 scorefxns=sfxn_hard/>
                <AtomTree name=ftree simple_ft=1/>
                <MinMover name=min_bb_sc bb=0 chi=1 jump=1 scorefxn=sfxn_hard>
                        <MoveMap name=minmvrmap>
```

```
                                <Chain number=2 chi=1 bb=0/>
                                <Chain number=1 chi=1 bb=0/>
                                <Jump number=1 setting=1/>
                        </MoveMap>
                </MinMover>
        </MOVERS>
        <FILTERS>
                <Ddg name=binding threshold=0 scorefxn=sfxn_hard confidence=0 jump=1 repack=1
relax_mover=min_bb_sc repeats=3/>
                <Sasa name=dsasa threshold=500 confidence=0/>
                <ShapeComplementarity name=shape jump=1 verbose=0 min_sc=0.60 confidence=0/>
                <SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>
                <ScoreType name=total_score scorefxn=sfxn_hard score_type=total_score confidence=0
threshold=0/>
                <CombinedValue name=combo>
                        <Add filter_name=total_score factor="%%total_score%%"/>
                        <Add filter_name=unsat factor="%%unsat%%"/>
                        <Add filter_name=binding factor="%%binding%%"/>
                        <Add filter_name=shape factor="%%shape%%"/>
                </CombinedValue>
        </FILTERS>
        <MOVERS>
                <GreedyOptMutationMover name=greedy
task_operations=pido,arochi2,rctr,nocys,resfile filter=combo scorefxn=sfxn_hard
relax_mover=min_bb_sc sample_type=low rtmin=0 design_shell=-1 repack_shell=8.0 parallel=1/>
        </MOVERS>
        <PROTOCOLS>
                <Add mover_name=ftree/>
                <Add mover_name=fsp_pssm/>
                <Add mover_name=greedy/>
                <Add filter_name=binding/>
                <Add filter_name=dsasa/>
                <Add filter_name=shape/>
                <Add filter_name=unsat/>
                <Add filter_name=total_score/>
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

For each candidate binder, the amino acid sequences of the models output from each of the 20 greedy runs were aligned in Geneious (Biomatters Inc. San Francisco, CA, USA) and a consensus sequence was determined. The model with the sequence closest to the consensus was modified to have the consensus sequence using Foldit. An electrostatics map was created using PyMOL for each modified model and non-complementary charge interactions across the binder-EED interface were flagged. These flagged interactions were manually adjusted in Foldit by mutating the residue on the binder side of the interface to the most appropriate identity and the resulting model was re-evaluated in PyMOL. At this point, N-linked glycosylation sites on the binder were identified by the consensus sequence NXS/T and removed by mutating either the Asparagine or the Serine/Threonine in Foldit.

Once all non-complementary charged interactions and N-linked glycosylation sites were corrected, the modeled interface was validated by subjecting the model to side chain conformational sampling and side chain and backbone minimization without any constraints using the fast design with atompair constraints protocol in both the bound and unbound states. None of the 8 candidate designs saw dissociation of the modeled complex when relaxed in the bound state with no constraints. The validated models were subjected to fragment-based ab initio structure prediction using only their primary amino acid sequences and the top scoring models were compared to the design models.

A sample script to generate an unbound binder model, identify the grafted region, and relax the binder in the unbound state is given below:

```
#!/bin/bash

hs_finder=`/path_to_residue_finder/residue_finder.sh <input_model> PHE 356 CZ A PHE CZ B`
hs_list=`echo $hs_finder | awk '{print $4","$4+3","$4+7","$4+10","$4+11","$4+14","$4+18}'`
hs_range_start=`echo $hs_finder | awk '{print $4-352}'`
hs_range_end=`echo $hs_finder | awk '{print $4+18-352}'`
mnmr_flank_start=`echo $hs_finder | awk '{print $4-7-352}'`
mnmr_flank_end=`echo $hs_finder | awk '{print $4+25-352}'`
pdb=`echo <input_model> | sed 's|\.pdb||'`
```

```
grep ^ATOM <input_model> | grep ' B ' > ${pdb}_chain_B.pdb
/path_to_scripts/convpdb.pl -renumberAcrossChains ${pdb}_chain_B.pdb > ${pdb}_chain_B_rlx.pdb
rm -f ${pdb}_chain_B.pdb

/path_to_rosetta/rosetta_scripts.default.linuxgccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -ignore_unrecognized_res -
overwrite -out:file:renumber_pdb false -ex1 -ex2 -nstruct 1 -parser:script_vars
hs_list="$hs_list" mnmr_flank_start=${mnmr_flank_start} mnmr_flank_end=${mnmr_flank_end}
hs_range_start=${hs_range_start} hs_range_end=${hs_range_end} -s ${pdb}_chain_B_rlx.pdb -
in:file:native ${pdb}_chain_B_rlx.pdb -parser:protocol ./validate_monomer.xml -score:weights
talaris2013
```

A sample xml protocol to relax and score the binder in the unbound state is given below:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <sfxn_hard weights=talaris2013/>
        </SCOREFXNS>
        <TASKOPERATIONS>
                <LimitAromaChi2 name=arochi2/>
                <RestrictChainToRepacking name=rctr2 chain=1/>
        </TASKOPERATIONS>
        <MOVERS>
                <FastRelax name=fstrlx_monomer scorefxn=sfxn_hard repeats=8
task_operations=rctr2,arochi2/>
                <Superimpose name=super ref_start=%%mnmr_flank_start%% ref_end=%%mnmr_flank_end%%
target_start=%%mnmr_flank_start%% target_end=%%mnmr_flank_end%%/>
        </MOVERS>
        <FILTERS>
                <ScoreType name=total_score_monomer_after scorefxn=sfxn_hard
score_type=total_score confidence=0 threshold=0/>
                <Rmsd name=rmsd_all_rlx chains=B threshold=5 confidence=0 superimpose=1/>
                <Rmsd name=rmsd_graft_rlx threshold=5 confidence=0 superimpose=1>
                        <span begin_res_num=%%hs_range_start%% end_res_num=%%hs_range_end%%/>
                </Rmsd>
        </FILTERS>
        <PROTOCOLS>
                <Add mover_name=fstrlx_monomer/>
                <Add mover_name=super/>
                <Add filter_name=total_score_monomer_after/>
                <Add filter_name=rmsd_all_rlx/>
                <Add filter_name=rmsd_graft_rlx/>
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

A sample script to identify the grafted region and relax the binder and EED in the bound state is given below:

```
#!/bin/bash

hs_finder=`/ path_to_residue_finder/residue_finder.sh <input_model> PHE 356 CZ A PHE CZ B`
hs_list=`echo $hs_finder | awk '{print $4","$4+3","$4+7","$4+10","$4+11","$4+14","$4+18}'`
hs_range_start=`echo $hs_finder | awk '{print $4}'`
hs_range_end=`echo $hs_finder | awk '{print $4+18}'`

/path_to_rosetta/rosetta_scripts.default.linuxgccrelease -database
/path_to_rosetta_database/database -ignore_zero_occupancy false -ignore_unrecognized_res -
overwrite -out:file:renumber_pdb false -ex1 -ex2 -nstruct 1 -s <input_model>  -parser:script_vars
hs_list=${hs_list} hs_range_start=${hs_range_start} hs_range_end=${hs_range_end} -parser:protocol
./validate_complex.xml -score:weights talaris2013
```

A sample xml protocol to relax and score the binder and EED in the bound state is given below:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <sfxn_hard weights=talaris2013/>
        </SCOREFXNS>
        <TASKOPERATIONS>
                <LimitAromaChi2 name=arochi2/>
                <RestrictChainToRepacking name=rctr chain=1/>
```

```
		<RestrictChainToRepacking name=rctr2 chain=2/>
	</TASKOPERATIONS>
	<MOVERS>
		<AtomTree name=ftree simple_ft=1/>
		<MinMover name=min_bb_sc bb=0 chi=0 jump=0 scorefxn=sfxn_hard>
			<MoveMap name=minmvrmap>
				<Chain number=2 chi=0 bb=0/>
				<Chain number=1 chi=0 bb=0/>
				<Jump number=1 setting=0/>
			</MoveMap>
		</MinMover>
		<FastRelax name=fstrlx_dsn scorefxn=sfxn_hard repeats=1
task_operations=arochi2,rctr,rctr2>
			<MoveMap name=fstrlxmap>
				<Chain number=2 chi=1 bb=0/>
				<Chain number=1 chi=1 bb=0/>
				<Jump number=1 setting=1/>
			</MoveMap>
		</FastRelax>
	</MOVERS>
	<FILTERS>
		<Ddg name=binding_energy threshold=0 scorefxn=sfxn_hard confidence=0 jump=1
repack=1 relax_mover=min_bb_sc repeats=3/>
		<Sasa name=dsasa threshold=500 confidence=0/>
		<ShapeComplementarity name=shape_comp jump=1 verbose=0 min_sc=0.60 confidence=0/>
		<SymUnsatHbonds name=unsat jump=1 cutoff=1000 confidence=0/>

		<ScoreType name=total_score_complex scorefxn=sfxn_hard score_type=total_score
confidence=0 threshold=0/>
		<Rmsd name=rmsd_all chains=B threshold=5 confidence=0 superimpose=1/>
		<Rmsd name=rmsd_graft threshold=5 confidence=0 superimpose=1>
		  <span begin_res_num=%%hs_range_start%% end_res_num=%%hs_range_end%%/>
		</Rmsd>
	</FILTERS>
	<PROTOCOLS>
		<Add mover_name=ftree/>
		<Add mover_name=fstrlx_dsn/>
		<Add filter_name=binding_energy/>
		<Add filter_name=dsasa/>
		<Add filter_name=shape_comp/>
		<Add filter_name=unsat/>
		<Add filter_name=rmsd_all/>
		<Add filter_name=rmsd_graft/>
		<Add filter_name=total_score_complex/>
	</PROTOCOLS>
</ROSETTASCRIPTS>
```

8. Automatically identifying positions of key interacting residues:

When optimizing the residue identities of designed binding proteins, its necessary to hold fixed the identities of those residues corresponding to the key binding residues from the native binder, Ezh2 in this case. RosettaScripts includes functions to prohibit sequence design at specified residue positions, but identifying the positions of key binding residues becomes nontrivial when using a wide variety of host proteins where the key residues may occupy different numerical positions from host to host. To solve this problem, a bash script, residue_finder.sh, was written that identifies the residue position of that residue on the design molecule closest to a specified residue on the target molecule. In the case of designs against EED, the script was instructed to look for all Cζ on Phenylalanine residues of the design (chain B) and measure the distance between each and the Cζ on Phenylalanine 356 of the target chain A (EED). The Phenylalanine Cζ closest to the Phenylalanine 356 Cζ can only correspond to the key interacting Phenylalanine taken from Ezh2. The residue positions of other key residues are then found by adding 3, 7, 10, 11, 14, or 18 to the value reported for the key Phenylalanine residue. This script was automatically run immediately before running any Rosetta application that carried out sequence optimization, unless the key interacting residues were otherwise denoted using a resfile.

The script is reproduced here:

```
#!/bin/bash
```

```
### This script takes a PDB file and given a specific residue type, residue number, atom name,
and chain ID, will find the closest instance of another given residue type, printing out to
standard output the found residue's atom name, residue type, chain ID, residue number, and the
distance between the anchor atom and the stub atom.
### Usage: ./get_chn_B_stubs.sh <pdbfilename> <anchor_restype> <anchor_resnum> <anchor_atomname>
<anchor_chnid> <stub_restype> <stub_atomname> <stub_chnid>
basename=`echo $1 | sed 's|\.pdb||'`
anchor_restype=$2
anchor_resnum=$3
anchor_atomname=$4
anchor_chnid=$5
stub_restype=$6
stub_atomname=$7
stub_chnid=$8
grep 'ATOM' ${basename}.pdb | grep $stub_restype | grep $stub_atomname | grep $stub_chnid >
${basename}_stub_list.tmp
lines=`wc ${basename}_stub_list.tmp | awk '{print $1}'`
for i in `seq 1 $lines`; do
       anchor_xcoord=`grep 'ATOM' ${basename}.pdb | grep " $anchor_restype " | grep "
$anchor_resnum " | grep " $anchor_atomname " | grep " $anchor_chnid " | awk '{print $7}'`
       anchor_ycoord=`grep 'ATOM' ${basename}.pdb | grep " $anchor_restype " | grep "
$anchor_resnum " | grep " $anchor_atomname " | grep " $anchor_chnid " | awk '{print $8}'`
       anchor_zcoord=`grep 'ATOM' ${basename}.pdb | grep " $anchor_restype " | grep "
$anchor_resnum " | grep " $anchor_atomname " | grep " $anchor_chnid " | awk '{print $9}'`
       awk 'NR=='$i'{print $0}' ${basename}_stub_list.tmp > ${basename}_stub_list_${i}.tmp
       awk '{stub_xcoord += $7} {stub_ycoord += $8} {stub_zcoord += $9} {anchor_xcoord +=
'$anchor_xcoord'} {anchor_ycoord += '$anchor_ycoord'} {anchor_zcoord += '$anchor_zcoord'} END
{print $3,$4,$5,$6,((stub_xcoord-anchor_xcoord)^2+(stub_ycoord-anchor_ycoord)^2+(stub_zcoord-
anchor_zcoord)^2)^0.5}' ${basename}_stub_list_${i}.tmp  >> ${basename}_stub_list_dist.tmp
done
sort -nk5 ${basename}_stub_list_dist.tmp | head -n1
rm ${basename}_stub_list*tmp
```

Purification and Biotinylation of EED3. Soluble human EED3 was expressed in E. coli, purified over Nickel-NTA resin, and desalted into PBS buffer lacking magnesium and calcium (10 mM disodium phosphate, 1.8 mM monopotassium phosphate, pH 7.4, 137 mM NaCl, 2.7 mM KCl). The EED was then concentrated, supplemented with glycerol to a final concentration of 10% v/v, and snap frozen in liquid nitrogen for storage. The EED was thawed and desalted into 20 mM HEPES, 250 mM potassium glutamate, pH 7.5, 1 mM TCEP. The proteins were concentrated to 3.1 mg/mL (57 μM) and enzymatically biotinylated with BirA biotin ligase using the kit from Avidity (Avidity LLC, Aurora, Colorado, USA). EED3 was separated from the BirA enzyme by re-purification over Nickel-NTA resin (Qiagen, Venlo, Limburg, Netherlands), concentrated to a volume of 500 μL and further purified by size exclusion chromatography over a Superdex 200 column into 20 mM HEPES, pH 7.5, 150 mM NaCl, 5% v/v glycerol, and 1 mM TCEP (GE Healthcare, Little Chalfont, UK). The proteins were concentrated to 0.54 mg/mL (10 μM), aliquoted, and snap frozen in liquid nitrogen for storage at -80°C. Thawed aliquots of EED3 were centrifuged to remove insoluble precipitates prior to incubation with yeast cells.

Yeast transformation. The gene for the N-terminal EED-binding helix of Ezh2 was synthesized and cloned into pETCON, a modified version of the pCTCON2 yeast display vector (Genescript, Piscataway, NJ, USA)(see main text ref# 15). Genes for EB15-22 and the 1le2 and 3lf9 controls were synthesized as linear double-stranded DNA fragments (gBlocks)(Integrated DNA technologies, Coralville, IA). The cloned genes were transformed directly into chemically competent EBY100 yeast while the linear gene fragments were transformed along with linearized pETCON (10, see main text ref# 15).

Yeast Surface Titration. Due to the limited quantity of soluble biotinylated EED3 available, the yeast surface titration method of Chao, et al. was modified to work with smaller labeling volumes, generally 5 uL (see main text ref# 15, 44). Data was processed using FlowJo (FlowJo, LLC, Ashland, OR, USA) and Excel, and curves were fit using an online nonlinear least-squared algorithm (http://statpages.info/nonlin.html).

SSM library construction. In one PCR reaction, a long internal site-specific forward primer containing a degenerate NNK codon and a common reverse primer were used to amplify a C-terminal fragment of the gene of interest from the mutation site to the C-terminus of the gene. In a separate, concurrent PCR reaction, a short internal site-specific reverse primer and a common forward primer were used to amplify a second, N-terminal fragment of

the gene from the N-terminus to just before the mutation site. In a third PCR reaction, the N-terminal fragment and C-terminal fragment were joined and further amplified using common forward and reverse primers flanking the gene of interest. The product was the full-length gene with a degenerate NNK codon at the desired position (11). This process was carried out in parallel for every residue to be mutated in the gene. The joined products from the 3rd PCR reaction were pooled and purified through gel extraction. The mutagenic genes fragments were then electroporated along with linearized vector into yeast (10). The NNK codon was only contained in the long internal site-specific forward primer. The forward and reverse internal primers for each position to be mutated were automatically designed from the parent gene sequence using a Python script developed by Dr. Eva Strauch.

FACS sorting. The EB15 and EB22 single site mutant and combinatorial libraries were FACS sorted for a number of criteria, including maximal affinity for EED3, maximal on-rate, minimal off-rate, maximal stability, and minimal nonspecific binding. Variants with maximal affinity were isolated by labeling the libraries with the lowest possible concentration of soluble biotinylated EED3 under non-avid conditions (see main text ref# 15, 12). Clones with maximal on-rate were isolated by labeling the libraries for short periods of time on ice. Clones with maximal stability were isolated by labeling the libraries for short lengths of time at 42°C. Note that the conditions for maximal stability and maximal on-rate conflicted and could not be used for the same library in a single sort round. The EB15 library was selected for maximal on-rate while the EB22 library was selected for maximal stability. Clones with minimal off-rate were isolated by labeling the libraries at saturating levels of EED3, washing away the unbound EED3, and incubating with a high concentration of soluble Ezh2 competitor for lengthy periods at 37°C. The Ezh2 competitor was intended to bind any molecules of EED3 that dissociated from the yeast surface and thus prevent their re-binding. Clones with minimal nonspecific binding were isolated by alternating between 3 different secondary fluorophores used for each round of sorting: streptavidin-R-phycoerythrin conjugate, streptavidin-allophycocyanin-Alexa-fluor-750 conjugate, and neutravidin-R-phycoerythrin conjugate (Life Technologies, Carlsbad, CA, USA).

DNA sequencing. 1 x 10^7 yeast cells were stored as pellets at -80°C. The cell walls were removed using zymolase and the cells were lysed by a freeze-thaw cycle followed by an alkaline lysis miniprep procedure. Sheared genomic DNA and ssDNA are partially cleaned up from the plasmid DNA by an exonuclease processing step and then a PCR step amplified the gene and appended pool-specific barcodes. A 2nd PCR step appended Illumina flow-cell adaptors. PCR products were purified by extraction from an agarose gel and quantified using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). High-throughput sequencing was carried out on an Illumina MiSeq (Illumina, San Diego, CA, USA). Each run generated roughly 4 million paired-end sequences which were output along with their associated quality scores in compressed .fastq files. The resulting data was analyzed using a modified version of the Enrich package (see main text ref# 16, 13). The resulting sequence counts were imported into Microsoft Excel for Mac 2011 (Microsoft Corporation, Redmond, WA, USA), sequences occurring less than 30 times in at least one pool were removed, and the Log2 enrichment or depletion of each unique sequence relative to its frequency in the unselected pool was calculated.

Expression and Purification of EB15 and EB22 variants. EB15 and EB22 were subcloned into pET29b (Novagen, Madison, Wisconsin, USA) and transformed into BL21-star E. coli cells (Thermo Fisher Scientific, Waltham, MA, USA). For each gene, a single colony was inoculated into 32 mL LB media supplemented with 30 µg/mL kanamycin and grown for 16 hours at 37°C with shaking at 250 rpm. The following morning, 28 mL of the EB22 overnight culture was diluted into 1 L TB media supplemented with 30 µg/mL kanamycin and grown to an O.D. of 0.6 at 37°C with shaking at 220 rpm. At an O.D. of 0.6, the temperature was reduced to 18°C. One hour later, isopropyl-thio-galacto-pyranoside (IPTG) was added to the culture to a final concentration of 0.13 mM and the cultures were incubated for an additional 20 hours at 18°C with shaking at 220 rpm. In the case of EB15, 28 mL of the overnight culture was diluted into 1 L LB media supplemented with 30 µg/mL kanamycin and grown to an O.D. of 0.9 at 37°C with shaking at 220 rpm. At an O.D. of 0.9, IPTG was added to the culture to a final concentration of 0.15 mM and the cultures were incubated for an additional 4.5 hours at 37°C with shaking at 220 rpm. The cells were collected by centrifugation, the media was removed, and the cell pellets were resuspended in wash buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 30 mM imidazole) prior to storage at -80°C. Upon thawing the cells, Phenyl-methyl-sulfonyl-fluoride (PMSF) and benzamidine were added to the cell suspension to final concentrations of 1 mM. Lysozyme and deoxyribonuclease I (DNAse I) were also added to final concentrations of 1 mg/mL. The cells were sonicated in 15 x 20-second cycles and then centrifuged at 20500 x g for 40 minutes. The supernatant was passed through a 0.45 µM filter and applied 3 times in series to a column containing 1 mL Nickel-NTA-Superflow resin (Qiagen, Venlo, Limburg, Netherlands) pre-equilibrated with wash buffer. The resin was washed with 25 mL

of wash buffer, next with 30 mL of 20 mM HEPES, pH 7.5, 1 M NaCl, and then with another 25 mL of wash buffer. The protein was eluted from the resin with 2.5 mL of 20 mM Tris, pH 7.5, 500 mM NaCl, 250 mM Imidazole in two steps and initially desalted into 20 mM Tris, pH 7.5, 50 mM NaCl using a PD-10 desalting column (GE Healthcare, Little Chalfont, UK) following the vendor protocol. Protein purity and molecular weight were determined using SDS-PAGE and mass spectrometry.

Circular Dichroism. EB22 was buffer exchanged into 20 mM NaPO4, 150 mM NaCl and diluted 1:10 into 20 mM NaPO4. EB22 was analyzed at a final concentration of 18 μM. Circular dichroism (CD) spectra were collected on a Jasco J-1500 CD Spectrophotometer (Jasco, Easton, Maryland, USA) using a 1 mm pathlength quartz cuvette. Scans were collected at 25°C and were taken from 195 to 260 nm in 0.1 nm steps with 1 nm bandwidth at a scanning speed of 100 nm/minute. Three independent scans were averaged and buffer subtracted against a cuvette holding 20 mM NaPO4. Temperature melts were carried out with the same parameters from 25°C to 95°C in 1°C steps reading at 222 nm. During the melt, a full wavelength scan was taken at 25, 35, 45, 55, 65, 75, 85, and 95°C using the parameters above. EB15.2 and EB22.2 were dialyzed into PBS (20mM NaPO4, 150mM NaCl). EB22.2 was analyzed at a final concentration of 9.69 μM while EB15.2 as at 10.7 μM. Circular dichroism (CD) spectra were collected on an AVIV Model 420 CD spectrometer (AVIV Biomedical, Inc, Lakewood, NJ, USA) using a 1 mm pathlength quartz cuvette. Scans were collected at 25°C and were taken from 195 to 265 nm in 1 nm steps with 1 nm bandwidth at a scanning speed of 10 nm/minute. Three independent scans were averaged and buffer subtracted against a cuvette holding PBS. Temperature melts were carried out with the same parameters from 25°C to 95°C in 1°C steps reading at 222 nm. During the melt, a full wavelength scan was taken at 25, 35, 45, 55, 65, 75, 85, and 95°C using the parameters above.

Biolayer Interferometry. Data was collected with an Octet RED96 (FortéBio, Menlo Park, CA, USA) instrument and analyzed with the ForteBio data analysis package. All experiments were performed at room temperature in HBS-EP Buffer (GE Biosciences) with bovine serum albumin (BSA) blocking agent added (0.01 M HEPES, pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, 1% BSA). Dip and Read Streptavidin Biosensors (ForteBio) were activated for 30 minutes in buffer prior to loading with biotinylated EED3 (at 12.5 nM). After baseline reference collection, biosensors were dipped in analyte binder solutions to measure association and then returned to the empty buffer-containing baseline well to measure dissociation. Kinetic binding constants were determined after reference subtraction utilizing a 1:1 binding model.

Crystallography. Cloning, expression and purification of EED protein (76-441) were performed as described previously (14). EB22 was cloned into pET29b (Novagen, Madison, Wisconsin, USA), and transformed into Escherichia coli BL21 (DE3) competent cells (Thermo Fisher Scientific, Waltham, MA, USA) for induced expression with 0.2 mM IPTG at 16°C. EB22 was purified by Ni-NTA affinity chromatography, further purified via cation exchange column and gel filtration. EED-EB22 complex was obtained by mixing purified EED and EB22, then applied to gel filtration again. The gel filtration buffer contained 20mM Tris-HCl pH7.5, 150mM NaCl and 0.5mM TECP. The crystals of EED in complex with EB22 were grown at 18°C using the sitting-drop vapour-diffusion by mixing 1 μl protein and 1 μl reservoir solution. The crystallization conditions consisted of 0.2 M Sodium bromide and 20% PEG3350. The crystals were cryo-protected by reservoir solution plus 20% (v/v) glycerol and frozen in liquid nitrogen before data collection. Diffraction images, which revealed multiple lattices, were recorded at Advanced Photon Source beamline 24-ID-E and reduced to intensities with XDS (15). Symmetry-related intensities were merged with AIMLESS (16). Intensity statistics indicated significant anisotropy. The structure was solved by molecular replacement with PHASER (17) and search models based on PDB entries 2QXV (see main text ref# 13) and 3LF9 (see main text ref# 12), where coordinates from entry 3LF9 were modified with CHAINSAW (18) to match the EB22 amino acid sequence. Restrained coordinate and B-factor refinement was performed with REFMAC (19), a PHENIX (20) implementation of Rosetta (phenix.rosetta_refine)(21) and BUSTER (22)(BUSTER version 2.10.2. Cambridge, United Kingdom: Global Phasing Ltd.), where the geometry of the EED side chains was restrained (23) to the EED chain of a phenix.rosetta_refine model. The model was interactively rebuilt in COOT (24), and its geometry analyzed with MOLPROBITY (25). Higher resolution EED structures, such as PDB entry 3K26 (14) aided in the interpretation of poor electron density. Some EED or EB22 side chains were omitted from the model deposited in the Protein Data Bank due to missing or uninterpretable electron density. For selected analyses, the conformations of missing side chains were estimated by phenix.rosetta_refine. Atomic displacement parameters were analyzed on the PARVATI server (26). PDB_EXTRACT (27), CCP4 (28) and PHENIX programs, and the IOTBX (29) library were used to compile statistics in Table S4.

The script used to run Phenix.rosetta_refine is given below:

```
<pre><code>      1 sed -e 's/\(^HETATM.\{6\}\)UNK  UNX \(.\{56\}\)N/\1 O   HOH \2O/' \
      2     -e 's/\(^HETATM.\{6\}\) UNK UNX \(.\{56\}\)N/\1 O   HOH \2O/' \
      3     -e 's/\(^ATOM.\{9\}N.\{7\}A 289\)/TER\n\1/' \
      4     -e 's/\(^ATOM.\{9\}O.\{7\}B  33\)/TER\n\1/' \
      5     -e 's/\(^ATOM.\{9\}O.\{7\}B  77\)/TER\n\1/' \
      6   /home/disk7/tempel/data/EED/x255207/coot/201611031316.pdb \
      7   > 201611031316-mod.pdb
      8
      9 phenix.pdbtools 201611031316-mod.pdb output.file_name=forRosetta25.pdb  \
     10   keep="name C or name CA or name CB or name N or name O or resname HOH"
     11
     12
     13 phenix.rosetta_refine forRosetta25.pdb \
     14   /home/disk7/tempel/data/EED/x255207/ccp4/x255207_scaled5.mtz \
     15   input.xray_data.high_resolution=2.55 \
     16   input.xray_data.labels="F_xds,SIGF_xds" \
     17   input.xray_data.r_free_flags.label="FreeR_flag" \
     18   input.xray_data.r_free_flags.test_flag_value=0 \
     19   rosetta.density_sampling="thorough" \
     20   rosetta.protocol="hires" \
     21   rosetta.number_of_models=20 \
     22   runtime.nproc=20
</code></pre>

Line comments:
>1.-2. Replace UNX dummy atoms with water atoms, which could be interpreted by ROSETTA without
additional software configuration. Outright removal of the atoms could have allowed side chains
to be automatically modeled into electron density that likely did not arise from those side
chains.
3.-5. Prevent automatic linking of residues around main chain gaps.
9.-10. Remove side chains and allow Rosetta to select side chain conformations automatically.
```

Construction of Ezh1-EED homology model. All side chain torsion angles in the crystal structure of mouse Ezh2 bound to EED were optimized prior to construction of the homology model. The amino acid sequence of residues 46-74 of human Ezh1 were manually threaded onto the structure of residues 40-68 of mouse Ezh2 and all side chain torsion angles were again optimized to yield the final model.

Establishment of cancer cell lines expressing EB22.2/EB22.2NC. EB22.2/EB22.2NC and GFP were cloned into tet-inducible lentiviral vector PCW57.1 (Addgene, 41393) using gateway dual cloning system (Life technologies, 12537-102). We produced lentivirus using TransIT 293 transfection reagent (Mirus, MIR2700) and infected K562, WSU-DLCL2 and Pfeiffer cells by spin infection (2 hours at 2000 rpm). After puromycin selection and 18 hours of doxycycline induction (0.5 μg/ml), cells were sorted by Hematologic Neoplasia Flow Cytometry Core at Dana-Farber Cancer Institute using BD FACS Aria II to ensure that cells express EB22.2 and EB22.2NC at a similar level. We maintained K562, WSU-DLCL2, Pfeiffer and G401 cell cultures with RPMI (10% FCS, 0.5 μg/ml puromycin.)

Analysis of proteins interacting with EB22.2 in K562 cells. K562 cells were induced for the expression of EB22.2-GFP, EB22.2NC-GFP and GFP with 0.5 μg/ml doxycycline for 48 hours. Cells were lysed (Lysis buffer: 10mM Tris/Cl pH7.5; 150 mM NaCl; 0.5 mM EDTA; 0.5% NP-40; 0.02% Thimerosal) and pre-cleared using binding control magnetic agarose beads (Chromotek, bmab-20) at room temperature for 30 min. Pre-cleared lysates were incubated with GFP-Trap MA beads at 4°C for 3 hours. After wash, co-immunoprecipitated proteins were eluted by heating GFP-Trap MA beads at 100 °C for 10 min. Eluted proteins were subjected to SDS-PAGE gel run and western blot detection using EED antibody (Millipore, 09-774) or mass spectrometry analysis (Taplin Mass-spectrometry Facility, Harvard Medical School). We subtracted major proteins co-immunoprecipitated with GFP from the lists of proteins co-immunoprecipitated with EB22.2-GFP or EB22.2NC-GFP to exclude unspecific bindings to GFP or beads. Mass spectrometry data was analyzed in a semi-quantitative manner by calculating relative amount of each protein to the GFP fusion protein in each sample.

Sucrose gradient fractionation. Sucrose solutions were prepared by dissolving sucrose (Simga S9378) in 1X phosphate buffered saline (Gibco, 10010-023) with 1mM EDTA (Sigma E7889) and 0.5% Triton X-100 (Sigma T8787). To minimize differences between sucrose gradient tubes, multiple sucrose gradient tubes were made

simultaneously using freeze-thaw method (30). Six layers of sucrose gradient were prepared by sequential adding and snap-freezing of each layer in polyallomer centrifuge tube (Beckman Coulter, 331374) from 35% (bottom) to 10% (top). Sucrose gradient tubes were stored at -80 °C and thawed at 4 °C for overnight before use. K562 cells were induced for the expression of EB22.2/EB22.2NC by treating cells with 0.5 μg/ml doxycycline for 48 hours and harvested (4~5X107 cells). Nuclear fractions were prepared using NE-PER nuclear and cytoplasmic extraction reagents (Thermo Scientific, 78833). 1 mg of nuclear extracts were loaded on top of sucrose gradient tubes and fractionated by centrifugation at 40,000 RPM for 18 hours (Beckman Coulter, Optima XL-100K ultracentrifuge, SW 40 Ti rotor). Proteins in each fraction were precipitated by 20% TCA (Sigma, T9159).

Proliferation assay of Cancer cells. WSU-DLCL2 and Pfeiffer cells were plated in 96 well plates at the density of 2,500 cells/well and 10,000 cells/ well, respectively. Doxycycline was added at final concentrations of 0.0, 0.063, 0.125, 0.25, 0.5 and 1.0 μg/ml. At the day of measurement, the same volume of media (100 μl) was added to each well and removed to measure proliferation using Cell titer glo (Promega, G7572) and FLUOstar Omega (BMG Labtech). To measure collaborative effects of doxycycline and GSK126, cells were treated with doxycycline and GSK126 simultaneously by adding varying concentrations of doxycycline at a given column (0.0, 0.063, 0.125, 0.25, 0.5 and 1.0 μg/ml) and adding varying concentrations of GSK126 at a given row (0.0, 0.004, 0.008, 0.015, 0.031, 0.063, 0.125, 0.25, 0.5 and 1 μM). Proliferation of cells was measured as described above. Doxycycline and GSK126 were replenished after addition of fresh media.

Cell cycle and Apoptosis analysis. WSU DLCL2 cells expressing EB22.2/EB22.2NC were treated with doxcycyline at 0.5 μg/ml for 4 days (cell cycle analysis) and 8 days (Apoptosis analysis). For cell cycle analysis, BrdU was added to cell cultures at 10 μM final concentration. Cell cultures were incubated for additional 30 min for the incorporation of BrdU to newly synthesized DNA. Cells were stained using APC BrdU Flow kit (BD Biosciences, 552598) and analyzed using BD FACSCalibur. For apoptosis analysis, cells were stained with Annexin V-PE Apoptosis detection kit I (BD Pharmingen, 559763) and analyzed using BD FACSCalibur. Flow cytometry data was analyzed using flowJo software. To detect increase in caspase 3/7 activities, cells were plated in 96 wells at the density of 20,000 cells/well. Activities of caspase 3/7 were measured using capase-Glo 3/7 assay (Promega, G8090) and FLUOstar Omega (BMG Labtech). Cells treated with camptothecin (BioVision, K121-5) were used as a positive control for apoptosis anlaysis.

Culture of embryonic stem cells. Naïve and primed ESC were cultured as previously described (see main text ref# 23). Briefly, human ESC and primed mouse ESC were grown on a feeder layer of irradiated primary mouse embryonic fibroblasts in ESC medium: DMEM/F-12 media supplemented with 20% knock-out serum replacer, 0.1 mM nonessential amino acids (NEAA), 1 mM sodium pyruvate, and penicillin/streptomycin (all from Invitrogen, Carlsbad, CA) and 0.1 mM β-mercaptoethanol (Sigma-Aldrich, St. Louis, MO). For naïve hESCs [Elf-1(NIHhESC-12-0156), (64)], ESC medium was supplemented with 1μM GSK3 inhibitor (CHIR99021, Selleckchem), 1μM of MEK inhibitor (PD0325901, Selleckchem), 10ng/mL human LIF (Chemicon), 5ng/mL IFG1 (Peprotech) and 10ng/mL bFGF: Elf1 2i/L/I/F. For mouse primed ESC [EpiSC, (32)], ESC medium was supplemented with bFGF (10ng/mL) and Activin A (10ng/mL). Mouse naive ESC (R1) were culture in medium containing DMEM, 20% ES cell-qualified fetal bovine serum, 0.1mM NEAA, 1 mM sodium pyruvate, 0.1mM β-mercaptoethanol, penicillin/streptomycin, 2μM GSK3 inhibitor, 1μM of MEK inhibitor and 10ng/mL mouse LIF (Millipore). Elf1 and R1 cells were pushed toward a more primed stage by culturing them in mTeSR1 medium or ESC medium supplemented with Activin A (10ng/mL) and bFGF (10ng/mL) (R1 AF). One passage prior to the experiments, the cells were transferred onto Matrigel (Becton Dickinson, Moutainview, CA) in MEF conditioned media (CM). For primed hESC [WTC iPSC, see main text ref# 21; WIBR3(NIHhESC-0079)], cells were grown in mTeSR1 media on matrigel (Becton Dickinson, Moutainview, CA). Trypsin/EDTA, dispase or Versene were used to passage naïve ESC, naïve-primed transitioning cells and primed hESC, respectively. All cells were grown at 37 degrees Celcius and 5%CO2. Naïve WIBR3 hESC were cultured as previously described in (see main text ref# 24). Briefly, WIBR3 were grown on a feeder layer of irradiated primary mouse embryonic fibroblasts in hESC N2B27 base media, supplemented with 4i/L/A or 5i/L/A. Medium was generated by mixing the following : 120 ml DMEM/F12 (Invitrogen; 11320), 120 ml Neurobasal (Invitrogen; 21103), 2.5 ml N2 supplement (Invitrogen; 17502048), 5 ml B27 supplement (Invitrogen; 17504044), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM b-mercaptoethanol (Sigma), penicillin-streptomycin (Invitrogen), and 50 mg/ml BSA (Sigma). The media was freshly supplemented with 4i/L/A or 5i/L/A: BRAF (0.5μM SB590885, Selleckchem), SRC (1μM, WH-4-023, SB590885), MEKi(1μM, PD0325901, Selleckchem), ROCKi (10μM, Y-27632, Tocris), recombinant human LIF (20ng/mL, Chemicon), and Activin A (10ng/mL, PeproTech) , without (4i/L/A) or with (5i/L/A) GSK3i (IM-12, Selleckchem, 1μM). Elf1 cells were adapted to 4i/L/A or 5i/L/A conditions for at least 3 passages before analysis.

Vector construction. EEDbinder DNA sequences (EB22.2, EB15.2, EB22.2NC and EB15.2NC) were optimized based on the codon usage for human expression using Codon Optimization Tool (IDT), and synthesized as double-stranded gBlock fragments. Primers enclosing restriction cutting sites and decorations were utilized to amplify from gBlock fragments using PCR with Q5 High-Fidelity DNA Polymerase (New England Biolabs); and PCR products were Gibson assembled into mammalian vectors including AAVS1-TRE3G-EGFP (33), pcDNA3.1, pUS2 vectors. Purified products were confirmed by sequencing by Genewiz.

hESCs Electroporation. 0.7 to $1x10^6$ cells of hESC were transfected with 0.5µg AAVS1-TALEN-R plasmid (Addgene #59026), 0.5µg AAVS1-TALEN-L (Addgene #59025) and 4µg of either EEDbinder (EB22.2-NLS-GFP, EB22.2-3xFLAG-NLS, or EB15.2-3xFLAG-NLS) or EEDbinder negative control (EB22.2NC-NLS-GFP or EB22.2NC-3xFLAG-NLS) using Amaxa Lonza Human stem cell Kit #2. The cells were then plated with 5µM Rocki onto irradiated Drug Resistance 4 (DR4) Mouse Embryonic Fibroblasts. Two days following the nucleofection, the cells were selected for Puromycin 0.5µg/ml for 2 days.

RNA extraction and qPCR analysis. RNA was extracted from cells using Trizol and analysed with SYBRgreen qPCR using the 7300 real-time PCR system (Applied Biosystems) and TaqMan qPCR (Applied Biosystems). Primers used are listed in Table S10. Linear expression values for all qPCR experiments were calculated using the $2(-\Delta Ct)$ method. P values were calculated using Student's t-test.

Protein extraction and western blot analysis. Cellular extracts were prepared using a lysis buffer containing 20 mM Tris HCl (pH 7.5), 150 mM NaCl, 15% glycerol, 1% Triton, 25 mM β-glycerolphosphate, 50 mM NaF, 10 mM Na pyrophosphate, orthovanadate, phenylmethylsulphonyl fluoride (all chemicals are from Sigma-Aldrich), Protease inhibitor cocktail (Roche) and 2% SDS. Twenty-five units of benzonase nuclease (EMD Chemicals) and 20 mM of dithiothreitol (Sigma) were added to the lysis buffer right before use. Fifteen micrograms of protein (determined by Bradford) was loaded, separated by 4–20% SDS–PAGE, and transferred to polyvinylidene difluoride membranes, blocked with 5% non-fat dry milk for 60 min at room temperature, and incubated overnight at 4 °C with primary antibody. After incubation for one hour with horseradish peroxidase-conjugated secondary antibodies, they were visualized by enhanced chemiluminescence (Millipore Corp). Antibodies used in this study are: H3K27me3 (1:1,000, Abcam, ab6002), Flag (1:4000, Sigma), GFP (1:1000, Invitrogen, A-11122), EED (1:1,000, EMD millipore, 09-774), EZH2 (1:1000, Cell Signaling, 5246), SUZ12 (1:1000 Santa Cruz, sc46264), Jarid2 (1:1,000, Cell Signaling, 13594), Oct4 (1:1000, Novus Biologicals), Oct 4 (1:500, Santa Cruz) and b-tubulin (1:1000,Promega).

Immunostaining and fluorescence microscopy. Cells were fixed in 4% paraformaldehyde in PBS for 15 min, permeabilized for 10 min in 0.1% Triton X-100 and blocked for 1h in 2% BSA. The cells were then sequentially incubated in primary and secondary antibodies overnight at 4°C diluted in 2% BSA, followed by DAPI (1µ g/ml) staining for 10 minutes. Analysis was done on a Leica TCS-SPE Confocal microscope using a 40x objective and Leica Software. The following were used: rabbit anti-EED [EMD Millipore, 09-774]; rabbit anti-EED [Karol Bomsztyk, University of Washington]; mouse anti-Oct4 (Novus Biologicals); Alexa 488, 568 or 647-conjugated goat anti-mouse and anti-rabbit antibodies (Molecular Probes).

Co-immunoprecipitation (Co-IP). Cells were washed twice with PBS followed by the addition of 200µl GFP lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1.5 mM MgCl2, 1 mM DTT, 25 mM NaF, 5% glycerol, 0.5% NP-40, (1 tablet/10 ml) PMSF added freshly) or 1ml of Flag lysis buffer (50 mM Tris HCl, pH 7.4, with 150 mM NaCl, 1 mM EDTA, and 1% TRITON X-100, (1 tablet/10 ml) PMSF added freshly). Cells were left on ice with lysis buffer for 10 min and then scraped off using disposable scrapes. The lysate was transferred to fresh microtube and incubated on ice for 30 min with extensive vortexing every 10 min. Cell lysate was centrifuge at 20,000x g (14,000 rpm) for 15 min at 4ºC and the soluble fraction was transferred to fresh tube. GFP-Trap-A beads (gta-20, Chromotek) or Anti-Flag M2 Affinity Gel (A2220, Sigma) were treated exactly as described in the manufacturer's manual. To bind the protein to GFP beads or Flag agarose beads, 150ul of cell lysate or 900ul of cell lysate were resuspended with 20ul of beads slurry following by tumble end-over-end for 1 or 2h hour at 4ºC respectively. The bound protein was separated by centrifugation (1,500 rpm for 2 min at 4ºC) where the unbound flow-through was removed and the beads were washed extensively 4 times using lysis buffer and centrifugation at 1,500 rpm for 2 min at 4ºC. Finally, the beads were resuspended in 40ul of sample buffer X2, 20ul were loaded on gel.

Proteomics. Samples were analyzed as previously described (see main text ref# 23). Briefly, proteins were suspended in 1M urea, 50mM ammonium bicarbonate, pH 7.8, and heated to 50°C for 20 min. Denatured proteins were reduced with 2mM DTT, alkylated with 15mM iodoacetamide, and digested overnight with trypsin. The resulting peptides were desalted on Waters Sep-Pak C18 cartridges. Peptides were separated using a heated 50C 30cm C18 columns in a 180min gradient of 1% to 45% (vol/vol) acetonitrile with 0.1% (vol/vol) formic acid. Peptides were measured on a Thermo Scientific Q Exactive (QE) or Orbitrap Fusion Tribrid operated in data-dependent mode. Identification and label free quantification of proteins was done with MaxQuant 1.5 using a 1% false discovery rate
(FDR) against the human proteome dataset downloaded from Uniprot on July 27th, 2015. Peptides were searched using a 5ppm mass error and a match between run window of 2 min. Proteins that were significantly regulated between conditions were identified in Perseus 1.4.1.3.

FACS analysis. Cells were collected using Trypsin-EDTA (0.05%) phenol-Red, washed twice with PBS and resuspended with PBS-5% FBS. The primary antibody TRA-1-60 (MAB4360, EMD Millipore) or TG-30 (MAB4427, EMD Millipore) were added at 1:100 ratio and incubated on ice for 30 min. Cells were then washed once with PBS and incubated with secondary antibody IgM Alexa Fluor 647 at 1:100 ratio following incubation on ice at the absence of light for 20 min. Finally, cells were washed twice with PBS, resuspended in 500ul of PBS-5% FBS, analysed by FACS and FlowJo software.

EZH2 chemical inhibitory drugs. Cells were plated on 24 well plate or 35mm plate at the same density one day prior to drug treatment. Cells were either untreated (DMSO) or treated with either EPZ-6438 (Selleckchem, stock dilution 10mM) working solution of 1mM to final concentration of 2.5μM or 5μM, or Astemizole (Sigma, stock dilution 10mM) working solution of 0.1mM to final concentration of 0.25μM. Cells were grown for 5d in the presence of either drug and imaged on the last day of treatment. Finally, cells were harvested using lysis buffer and immunoblot with specific antibodies for further analysis. EPZ-6438, is a selective EZH2 inhibitor that is in clinical trials for B-cell and follicular Lymphomas (NCT01897571, Epizyme). EPZ-6438 (EPZ) competes with the substrate S-adenosylmethionine (SAM) (see main text ref# 10).

RNA-seq data analysis. RNA-seq were aligned to Ensembl GRCh37 using Tophat (version 2.0.13)(34). Gene-level read counts were quantified using htseq-count (35) using Ensembl GRCh 37 gene annotations. prcomp function from R was used to for Principal Component Analysis. DESeq (36) was used for differential gene expression analysis. Combat (37) was used to mitigate batch effects of RNA-seq samples. Human-mouse orthologous genes with one-to-one mapping was retrieved from GRCh 37 gene annotations using biomaRt package (38).

ChIP-seq. Cells were dissociated using trypsin, washed once in cold PBS and counted. Approximately 4.5 M cells were fixed and chromatin was sonicated as previously described (39). 11 ul of Dynabeads (Invitrogen) and 3 ug of H3K27me3 antibody (Active Motif) were conjugates and incubated with chromatin at 4C O/N. Chromatin-antibody-beads complexes were washed and sequencing libraries prepared as previously described (40). Input libraries were prepared from 5ng of purified DNA using Nextera Library Preparation Kit (Illumina). Libraries were sequenced on NextSeq500 in single end run and approximately 30 million raw reads were obtained for each sample.

ChIP-seq data analysis. ChIP-seq reads were aligned to Ensembl GRCh37 using Bowtie version 1.0.0. allowing 1 mismatch (-N 1). Reads within 2KB of transcription start site were counted using htseq-count. Differentially marked genomic regions were identified with diffReps version 1.55.4 (41) and annotated to the closest genes. Genes associated with at least one significant genomic region (FDR less than 0.01, fold change > 1, and read count >50 in either condition) were classified as differentially marked. When a gene is annotated with multiple significant genomic regions, the most significant one is assigned to that gene.

CRISPR-Cas9-based EZH2 and EED knockout. Two gRNAs targeting the exon 1 of the EED gene and one gRNA targeting the exon 10 of the EZH2 gene were designed using the CHOPCHOP web tool (42) and ordered as T7-gRNA primers. A dsDNA fragment was synthesized from these primers and a complementary scaffold primer following a Q5 High Fidelity-based PCR (New England Biolabs). This 120 bp strand served as template for IVT (MAXIscript T7 kit, applied Biosystems). The RNA was then purified using Pellet Paint® Co-Precipitant (Novagen). Elf1 iCas9 cells were treated with doxycycline (2 μg ml−1) for 2 to 3 days before and during transfection. For transfection, cells were dissociated with trypsin, replated onto matrigel-coated plates, and

transfected in suspension with gRNAs using Lipofectamine RNAiMAX (Life Technologies). gRNA was added at a 40 nM final concentration. A second transfection was performed after 24 h. Two days after the last gRNA transfection, iCas9 Elf1 cells were dissociated into single cells and replated onto MEF-coated plates. Single colonies were randomly selected and amplified. Genomic DNA was collected using DNAzol. Genomic regions flanking the CRISPR target sites were PCR amplified, purified and sent to Genewiz for sequencing. Alternatively, Samtools mpileup and BCF tools were used to identify variants on the basis of aligned RNA-seq BAM files. Sequences of the guides and PCR primers are presented in Table S10.

Supplemental References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235-242.
2. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popović Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 487:545-74.
3. Saikrishnan K, Kalapala SK, Varshney U, Vijayan M (2005) X-ray structural studies of Mycobacterium tuberculosis RRF and a comparative study of RRFs of known structure. Molecular plasticity and biological implications. *J Mol Biol* 345(1):29-38.
4. Nakano H, Uchiyama S, Yoshida T, Ohkubo T, Kato H, Yamagata Y, Kobayashi Y (2002) Crystallization and preliminary X-ray crystallographic studies of a mutant of ribosome recycling factor from Escherichia coli, Arg132Gly. *Acta Crystallogr D, Biol Crystallogr* 58(Pt 1):124-126.
5. Huang PS, Ban YE, Richter F, Andre I, Vernon R, Schief WR, Baker D (2011) RosettaRemodel: a generalized framework for flexible backbone protein design. *PloS One* 6(8):e24109.
6. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, Leaver-Fay A, Baker D, Popović Z, Players F (2010) Predicting protein structures with a multiplayer online game. *Nature* 466(7307)756-760.
7. King CA, Bradley P (2010) Structure-based prediction of protein-peptide specificity in Rosetta. *Proteins* 78(16):3437-3449.
8. Nivón LG, Bjelic S, King C, Baker D (2014) Automating human intuition for protein design. *Proteins* 82(5):858-866.
9. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. *Methods Enzymol* 383:66-93.
10. Benatuil L, Perez JM, Belk J, Hsieh CM (2010) An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* 23(4):155-159.
11. Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman SJ, Su M, Aramini J, Kornhaber G, Hunt JF, Tong L, Montelione GT, Baker D (2013) Computational design of a protein-based enzyme inhibitor. *J Mol Biol* 425(18):3563-3575.
12. Hulett HR, Bonner WA, Barrett J, Herzenberg LA (1969) Cell sorting: automated separation of mammalian cells as a function of intracellular fluorescence. *Science* 166(3906):747-749.
13. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 30(6):543-548.
14. Xu C, Bian C, Yang W, Galka M, Ouyang H, Chen C, Qiu W, Liu H, Jones AE, MacKenzie F, Pan P, Li SS, Wang H, Min J (2010) Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc Natl Acad Sci USA* 107(45):19266-19271.
15. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):133-144.
16. Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1204-14.
17. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
18. Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* 41:641-643.
19. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):355-367.
20. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
21. DiMaio F, Echols N, Headd JJ, Terwilliger TC, Adams PD, Baker D (2013) Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat Methods* 10(11):1102-1104.

22. G. Bricogne E. Blanc M. Brandl C. Flensburg P. Keller W. Paciorek P. Roversi A. Sharff O.S. Smart C. Vonrhein T.O. Womack, BUSTER version 2.10.2. Cambridge, United Kingdom: Global Phasing Ltd (2016).
23. Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):368-380.
24. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
25. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1)12-21.
26. Zucker F, Champ PC, Merritt EA (2010) Validation of crystallographic models containing TLS or other descriptions of anisotropy. *Acta Crystallogr D Biol Crystallogr* 66(Pt 8):889-900.
27. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 60(Pt 10):1833-1839.
28. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235-242.
29. Gildea RJ, Bourhis LJ, Dolomanov OV, Grosse-Kunstleve RW, Puschmann H, Adams PD, Howard JA (2011) iotbx.cif: a comprehensive CIF toolbox. *J Appl Crystallogr* 44(Pt 6):1259-1263.
30. Luthe DS (1983) A simple technique for the preparation and storage of sucrose gradients. *Anal Biochem* 135(1):230-232.
31. Ware CB, Nelson AM, Mecham B, Hesson J, Zhou W, Jonlin EC, Jimenez-Caliani AJ, Deng X, Cavanaugh C, Cook S, Tesar PJ, Okada J, Margaretha L, Sperber H, Choi M, Blau CA, Treuting PM, Hawkins RD, Cirulli V, Ruohola-Baker H (2014) Derivation of naive human embryonic stem cells. *Proc Natl Acad Sci USA* 111(12):4484-4489.
32. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448(7150):196-199
33. Qian K, Huang CT, Chen H, Blackbourn LW 4th, Chen Y, Cao J, Yao L, Sauvey C, Du Z, Zhang SC (2014) A simple and efficient system for regulating gene expression in human pluripotent stem cells and derivatives. *Stem Cells* 32(5):1230-1238.
34. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-1111.
35. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166-169.
36. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
37. Johnson WE, Li C, Rabinovic A. Rabinovic (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1):118-127.
38. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 4(8):1184-1191.
39. Valensisi C, Liao JL, Andrus C, Battle SL, Hawkins RD (2015) cChIP-seq: a robust small-scale method for investigation of histone modifications. *BMC Genomics* 16:1083.
40. Schmidl C, Rendeiro AF, Sheffield NC, Bock C (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* 12(10):963-965.
41. Shen L, Shao NY, Liu X, Maze I, Feng J, Nestler EJ (2013) diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PloS One* 8(6):e65598.
42. Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E (2014) CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 42(Web Server Issue):W401-407.
43. Collinson A, Collier AJ, Morgan NP, Sienerth AR, Chandra T, Andrews S, Rugg-Gunn PJ. (2016) Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. *Cell Rep.* 6;17(10):2700-2714.
44. Brooun A, Gajiwala KS, Deng YL, Liu W, Bolaños B, Bingham P, He YA, Diehl W, Grable N, Kung PP, Sutton S, Maegley KA, Yu X, Stewart AE (2016) Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. *Nat Commun* 7:11384.

45. Justin N, Zhang Y, Tarricone C, Martin SR, Chen S, Underwood E, De Marco V, Haire LF, Walker PA, Reinberg D, Wilson JR, Gamblin SJ (2016) Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun* 7:11316.
46. Sanulli S, Justin N, Teissandier A, Ancelin K, Portoso M, Caron M, Michaud A, Lombard B, da Rocha ST, Offer J, Loew D, Servant N, Wassef M, Burlina F, Gamblin SJ, Heard E, Margueron R (2015) Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol Cell* 57(5):769-783.
47. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ 3rd, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D, Gamblin SJ (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461(7265):762-767.
48. Gafni, O., Weinberger, L., Mansour, A. A., Manor, Y. S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., Rais Y, Shipony Z, Mukamel Z, Krupalnik V, Zerbib M, Geula S, Caspi I, Schneir D, Shwartz T, Gilad S, Amann-Zalcenstein D, Benjamin S, Amit I, Tanay A, Massarwa R, Novershtern N, Hanna JH. (2013) Derivation of novel human ground state naive pluripotent stem cells, *Nature* 504(7479): 282-6.

List of supplementary tables (found as individual tabs in separate Excel spreadsheet)

Table S1. Binding Affinities for EED Binding Proteins.
On rates were improved by two orders of magnitude for EB15.2 and EB22.2 compared to the native EZH2 protein (**11, *13) based on BLI measurements, consistent with on-yeast titration data showing designed proteins binding with sub-nanomolar affinity. Methods: Isothermal titration calorimetry (ITC), Fluorescence Polarization (FP), BioLayer Interferometry (BLI), yeast surface titration (YST).

Table S2. Yeast surface titration dissociation constants for EED binding proteins. ND = Not determined.

Table S3. Yeast surface titration controls. ND = Not determined.

Table S4. Crystallographic data collection and model statistics.

Table S5. Amino Acid sequences of designed binders and their evolved variants. The NdeI (His-Met, CATATG) and XhoI (Leu-Glu, CTCGAG) restriction sites on the N-terminus and C-terminus are given for reference, but other residues pertaining to the yeast display or bacterial expression vectors are omitted.

Table S6. Immunoprecipitation and Mass spectrometry analysis of EB22.2 binding proteins in K562 cells
K562 cells were treated with 0.5 μg/ml for two days and lysed with 0.5% NP40.
EB22.2-interacting proteins were co-immunoprecipitated with GFP antibody and analyzed using mass spectrometry.

Table S7. **Mass spectrometry analysis of EB22.2 binding proteins in Elf1 cells**
Proteins enriched in EB22.2 pulldowns were identified using a difference cut-off greater than two, EB22.2+-dox vs EB22.2 -dox (log2 LFQ intensity). Experiment #1: 22.2 vs GFP: dox 2.0 μg/mL, #2: EB22.2 vs EB22.2NC: dox 2.0 μg/mL, #3: EB22.2 vs EB22.2NC: dox 0.5 μg/mL. N=4-6 technical measurements.

Table S8. **Genes contributing to PCA in Figure 4A**

Table S9. **List of 223 genes represented in red in Figure 4B**
Those genes satisfy 3 criteria: 1. significantly higher H3K27me3 in Elf1 vs. Theunissen naive; 2. significantly lower expression in Elf1 vs. Theunissen naive; 3. significantly lower expression in Elf1 vs. Takashima naive.

Table S10. Sequences of guides and PCR and qPCR primers

Table S11. Mass spectrometry analysis of EB22.2 binding proteins in HeLa cells
Proteins enriched in EB22.2_FLAG pulldowns were identified using a cut-off of 0.01 (p-value), and difference greater than two, EB22.2_FLAG vs untransfected cells (log2 LFQ intensity) N=4 technical measurements.

Table S12. **List of genes up-regulated in Elf1 cells expressing EB22.2-FLAG from RNA-seq data**
(1290 genes that have significantly higher expression in EB22.2-FLAG+Dox (2ug/mL, 3 days) compared to –Dox in RNA-seq, and have H3K27me3 peaks in Elf1 cells. Among these 1290 genes, 286 bivalent genes are placed at the top. Genes are ranked by their combined signal in ChIP-seq and RNA-seq).

Table S13
Gene expression and H3K27me3 peak information for 374 genes that are expressed 2 fold higher in +Dox and have H3K27me3 peaks in –Dox (Fig.3D).

Table S14
List of 4632 genes marked within H3K27me3 peaks within 5KB of TSS in naïve 2iL-I-F EB22.2 treated with or without Dox (2 μg/mL, 3 days) (Fig. S21K-L).

List of additional data files: (Provided in compressed "Additional_Data_Files.zip")

Additional Data file 1 (separate file)
Atomic coordinates of phenix.rosetta_refine EB22:EED model in .pdb format

Additional Data file 2 (separate file)
Structure factor amplitudes of phenix.rosetta_refine EB22:EED model in .mtz format

Additional Data file 3 (separate file)
Atomic coordinates of Ezh1-EED homology model in .pdb format

Additional Data file 4 (separate file)
Atomic coordinates of EED-EB15 design model in .pdb format

Additional Data file 5 (separate file)
Atomic coordinates of EED-EB16 design model in .pdb format

Additional Data file 6 (separate file)
Atomic coordinates of EED-EB17 design model in .pdb format

Additional Data file 7 (separate file)
Atomic coordinates of EED-EB18 design model in .pdb format

Additional Data file 8 (separate file)
Atomic coordinates of EED-EB19 design model in .pdb format

Additional Data file 9 (separate file)
Atomic coordinates of EED-EB20 design model in .pdb format

Additional Data file 10 (separate file)
Atomic coordinates of EED-EB21 design model in .pdb format

Additional Data file 11 (separate file)
Atomic coordinates of EED-EB22 design model in .pdb format

Figure S1

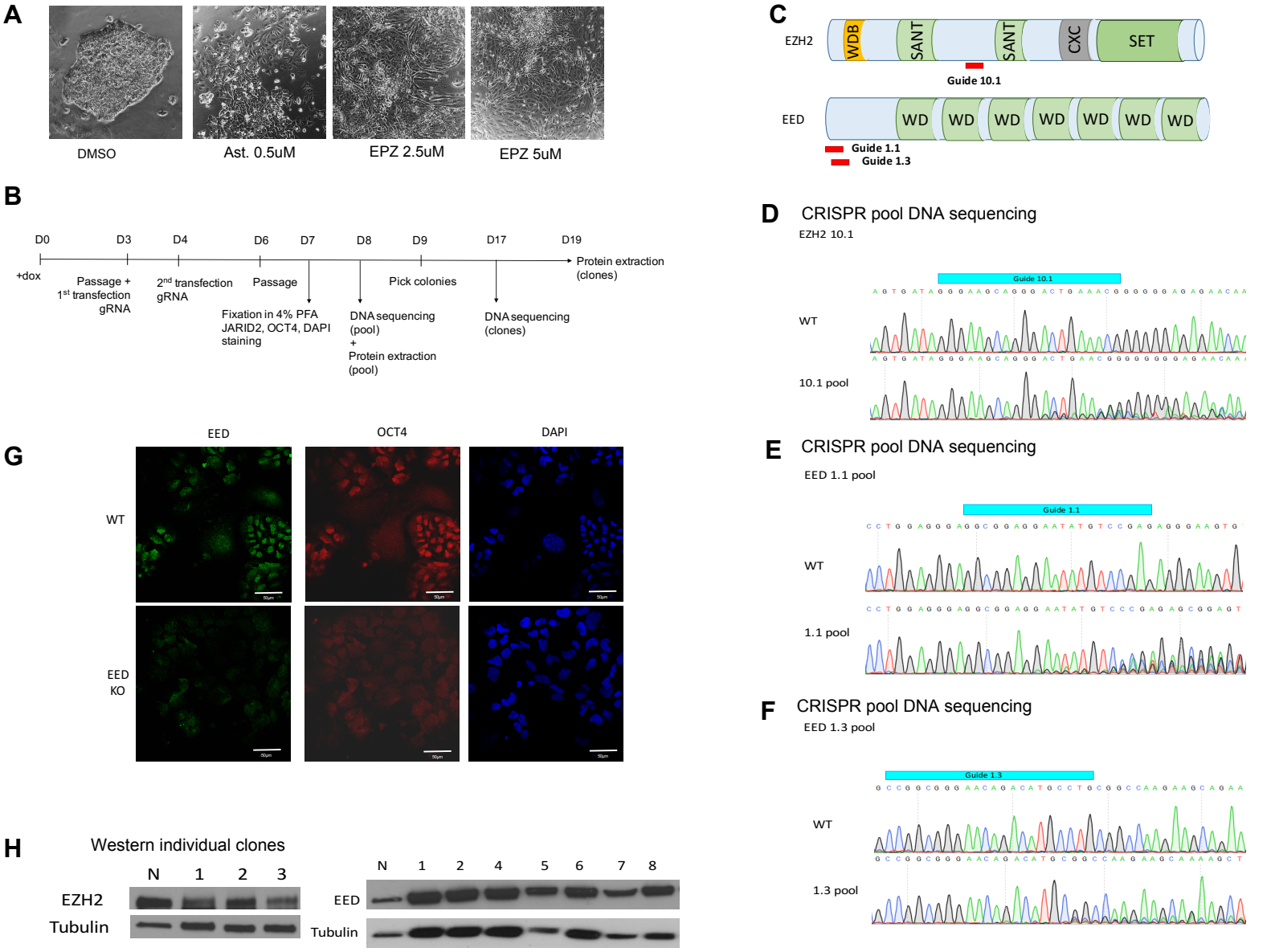

Figure S1. Elf1-iCas9 EED and EZH2 knock outs are unstable/unviable as hESC. A. Elf1 stem cell morphology under treatment of DMSO, Astemizole 0.5 μM, EPZ 2.5μM or 5μM (see main text ref# 9, 10) for 5 days resulting in dramatic morphological changes in Elf1 line. B-H. CRISPR mutations in EED and EZH2 were generated but expansion of clonal EED or EZH2 null lines was complicated, as seen before (43). Without using further sorting methods, all our single clonal lines generated from the pools expressed both proteins (H). B. Time line model of colony picking and selection. C. Schematic representation of EZH2 and EED proteins and relative localization of guides. D. CRISPR pool DNA sequencing of EZH2, guide 10.1. E-F. CRISPR pool DNA sequencing of EED, guide 1.1 (E) or 1.3 (F). G. Oct4 expression (red) is lost in EED knockout hESCs (loss of GFP). Confocal analysis of iCas9 hESC treated with EED g1.1. Scale bars represent 50 μm. H. Immunoblot analysis of knockout clones. N = naive hESC control. EED null cells were generated by the CRISPR-Cas9 system, since EED and Oct4 negative cells were observed by immunocytochemistry four days after mutant induction.

Figure S2

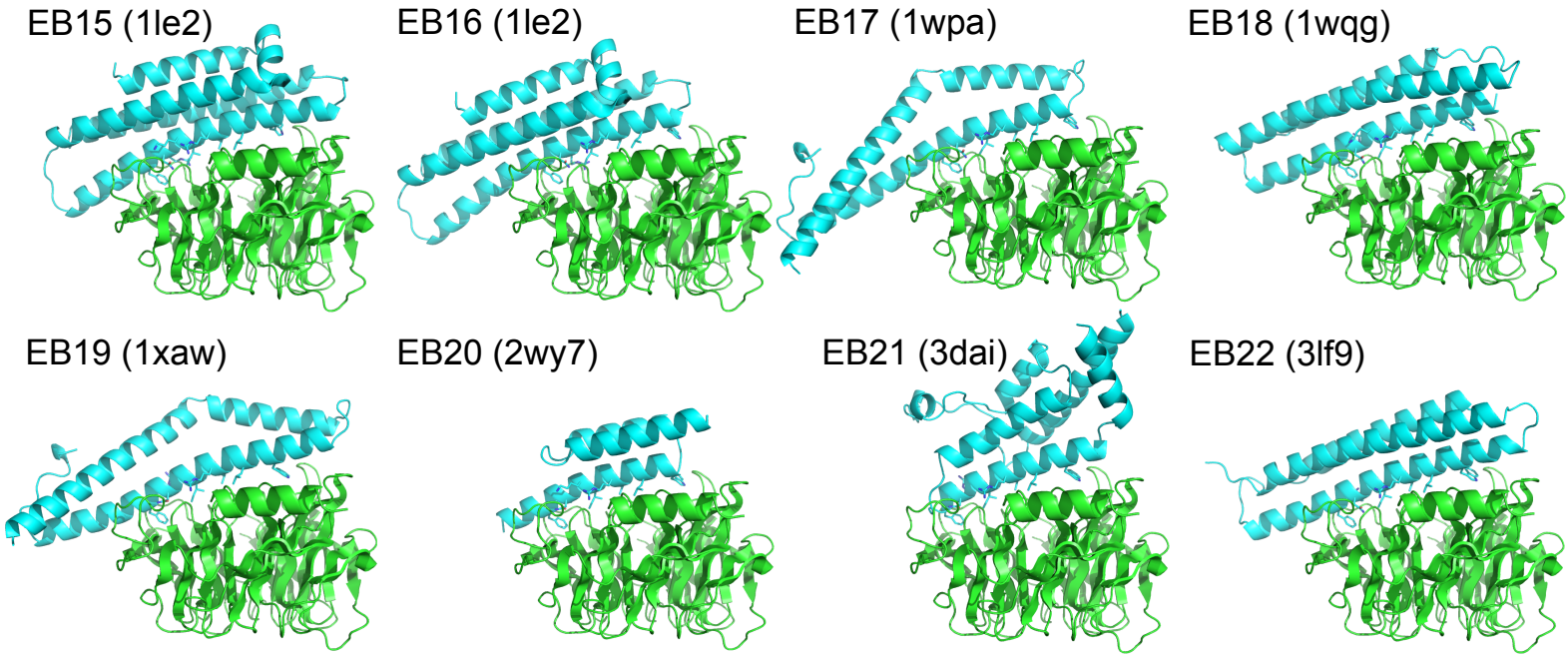

Figure S2. Models of designed EED binding proteins.
Designed EED binding proteins are shown as cyan cartoons while EED is shown as green cartoons.

Figure S3

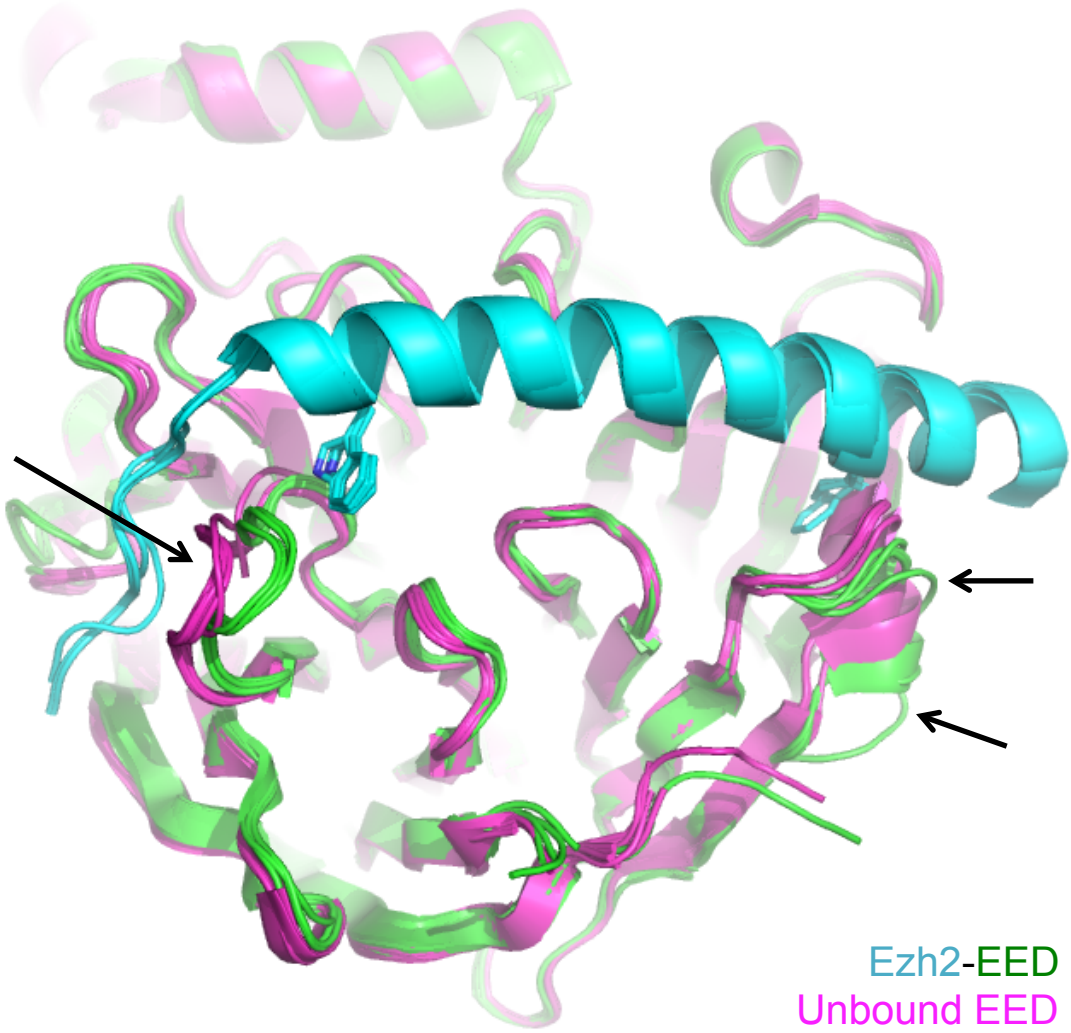

Figure S3. EED Loop Flexibility.
Four Crystal structures of EED (green cartoons) bound to Ezh2 (cyan cartoons), superimposed onto 12 crystal structures of EED not bound to Ezh2 (magenta cartoons). Regions with significant backbone conformational differences between bound and unbound states of EED are denoted with black arrows (44-47, main text ref# 13, 11).

Figure S4

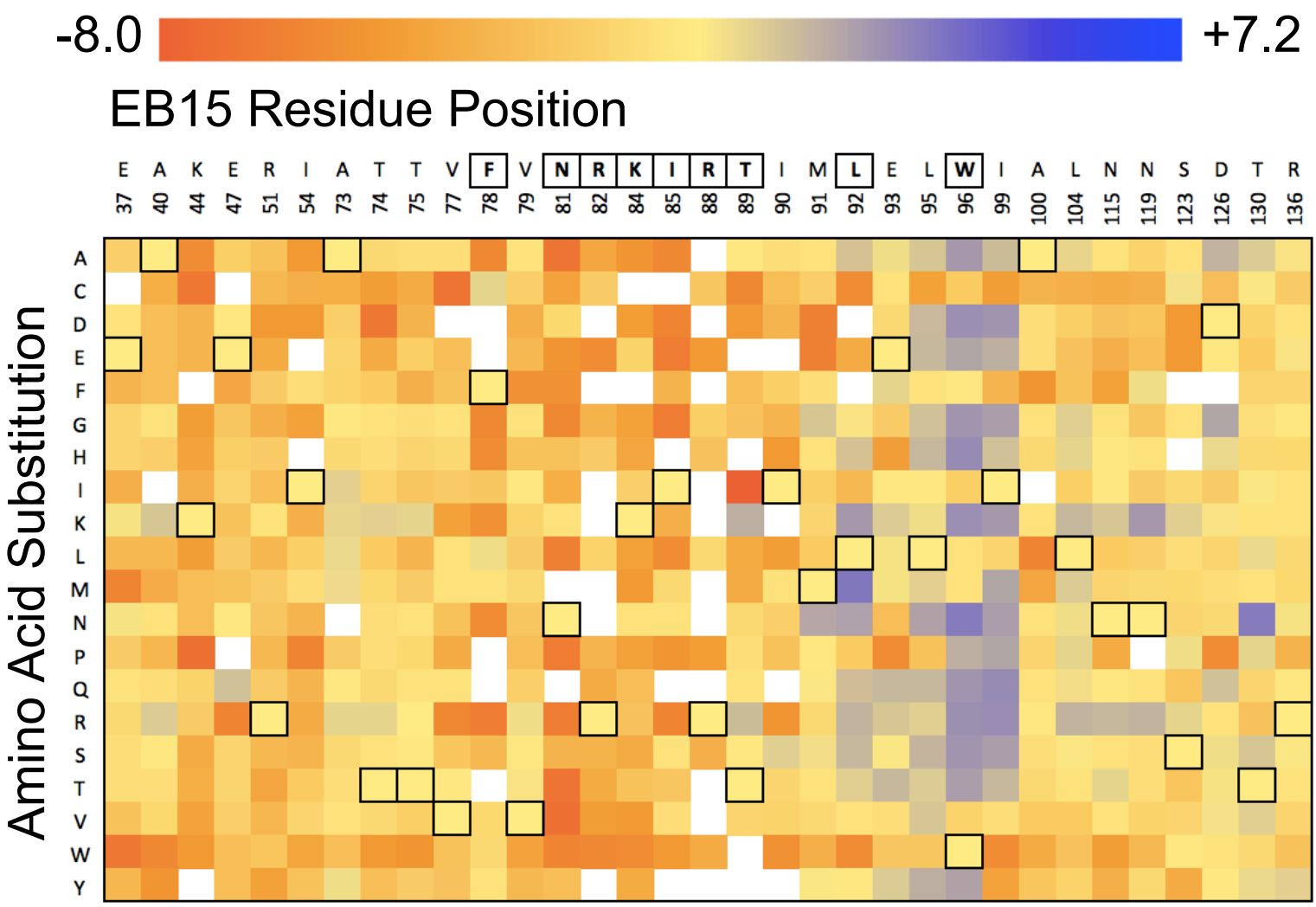

Figure S4. Interface residues at N-terminus of EB15 binding helix are conserved. Heatmap of log2 enrichment of variants at interface positions of EB15 after 1 round of sorting. The grafted residues are bolded and boxed at the top of the heat map while the wild type residue at each position is boxed within the heat map.

Figure S5

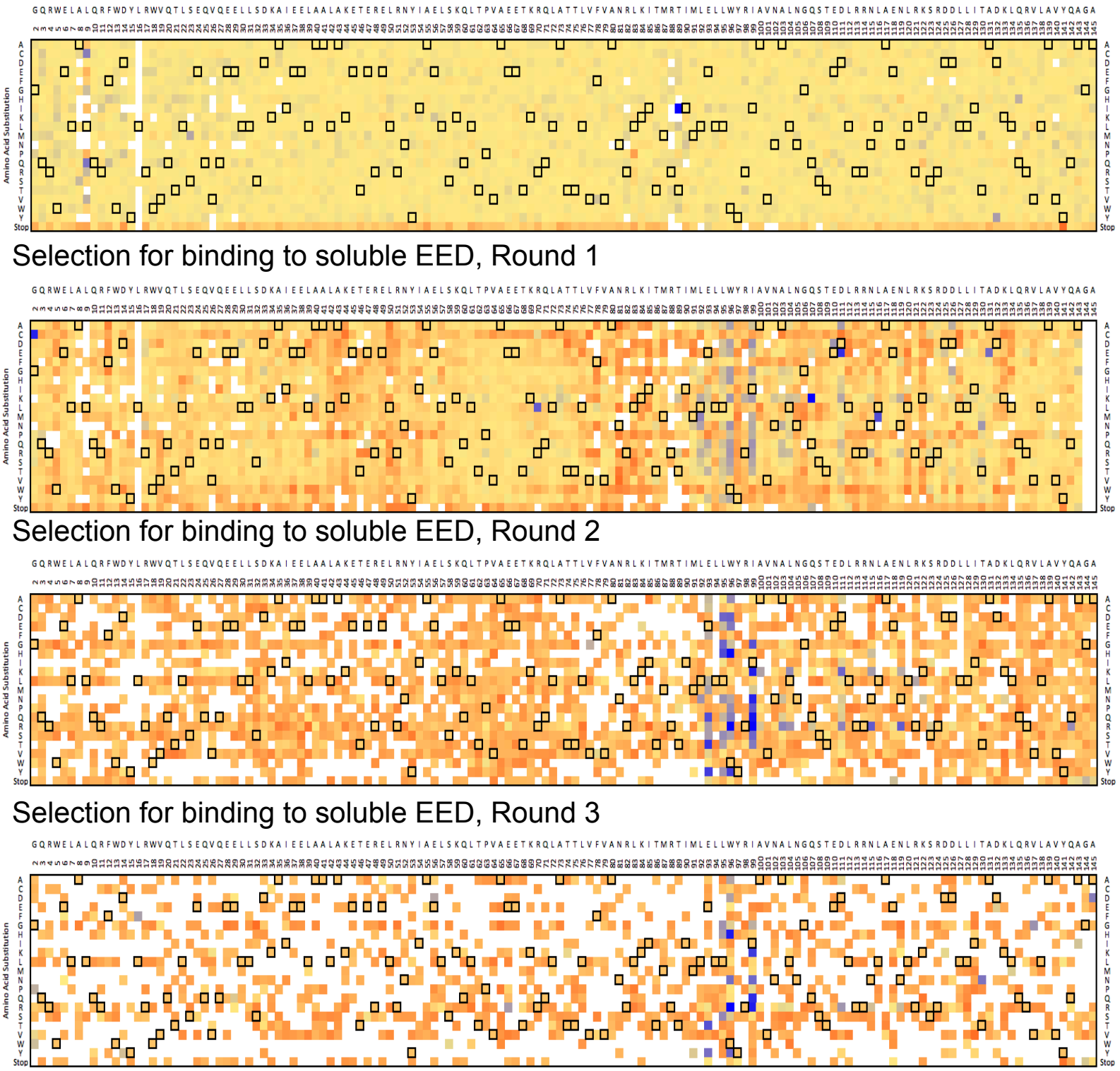

Figure S5. EB15 core residues are conserved.
Heatmaps giving the Log2 enrichment ratios for all tested point variants of EB15 through 4 rounds of FACS. Blue denotes enrichment, orange denotes depletion, and yellow denotes no enrichment or depletion. The wild type amino acid at each position is boxed.

Figure S6

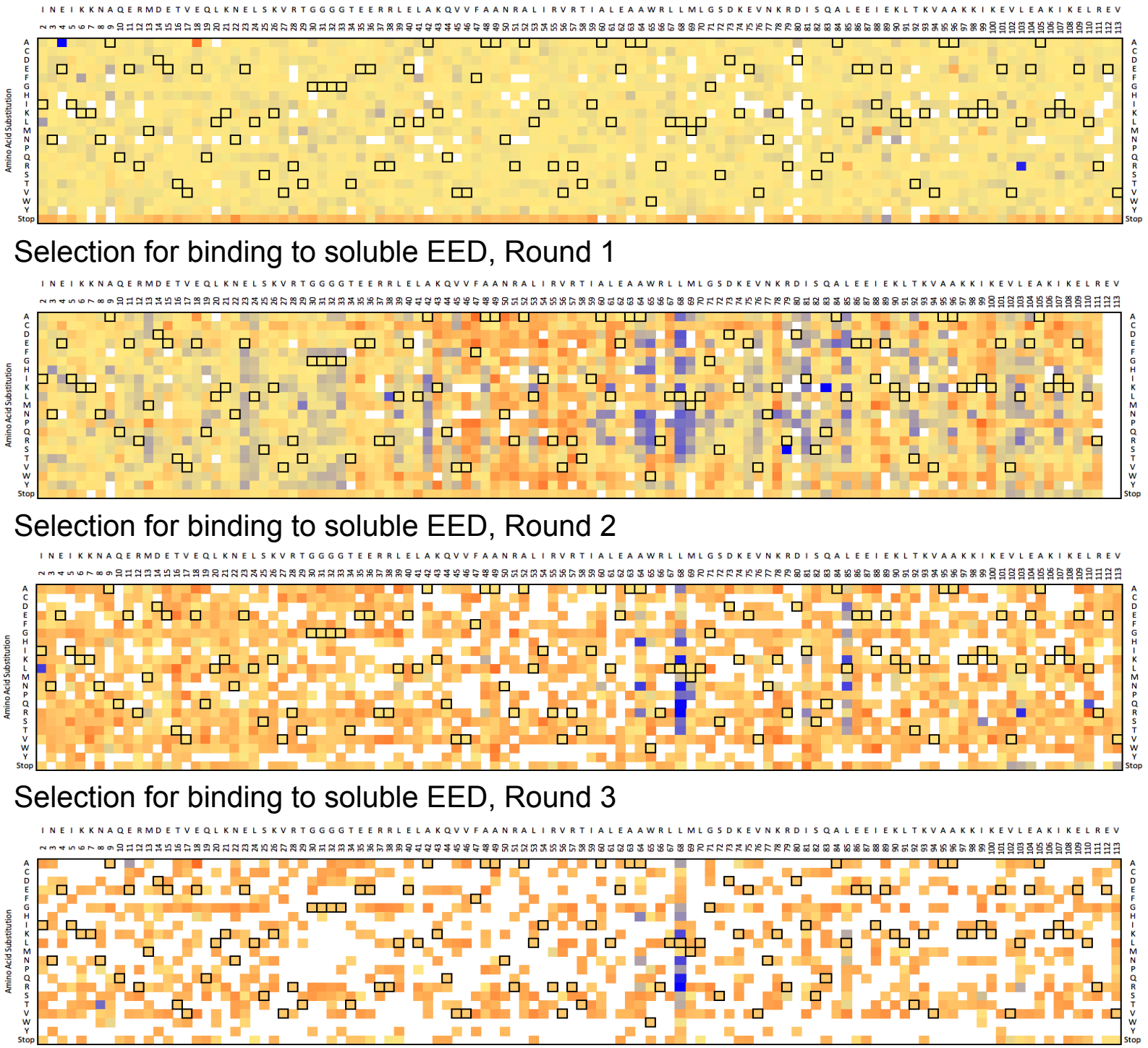

Figure S6. EB22 core residues are conserved.
Heatmaps giving the Log2 enrichment ratios for all tested point variants of EB22 through 4 rounds of FACS. Blue denotes enrichment, orange denotes depletion, and yellow denotes no enrichment or depletion. The wild type amino acid at each position is boxed.

Figure S7

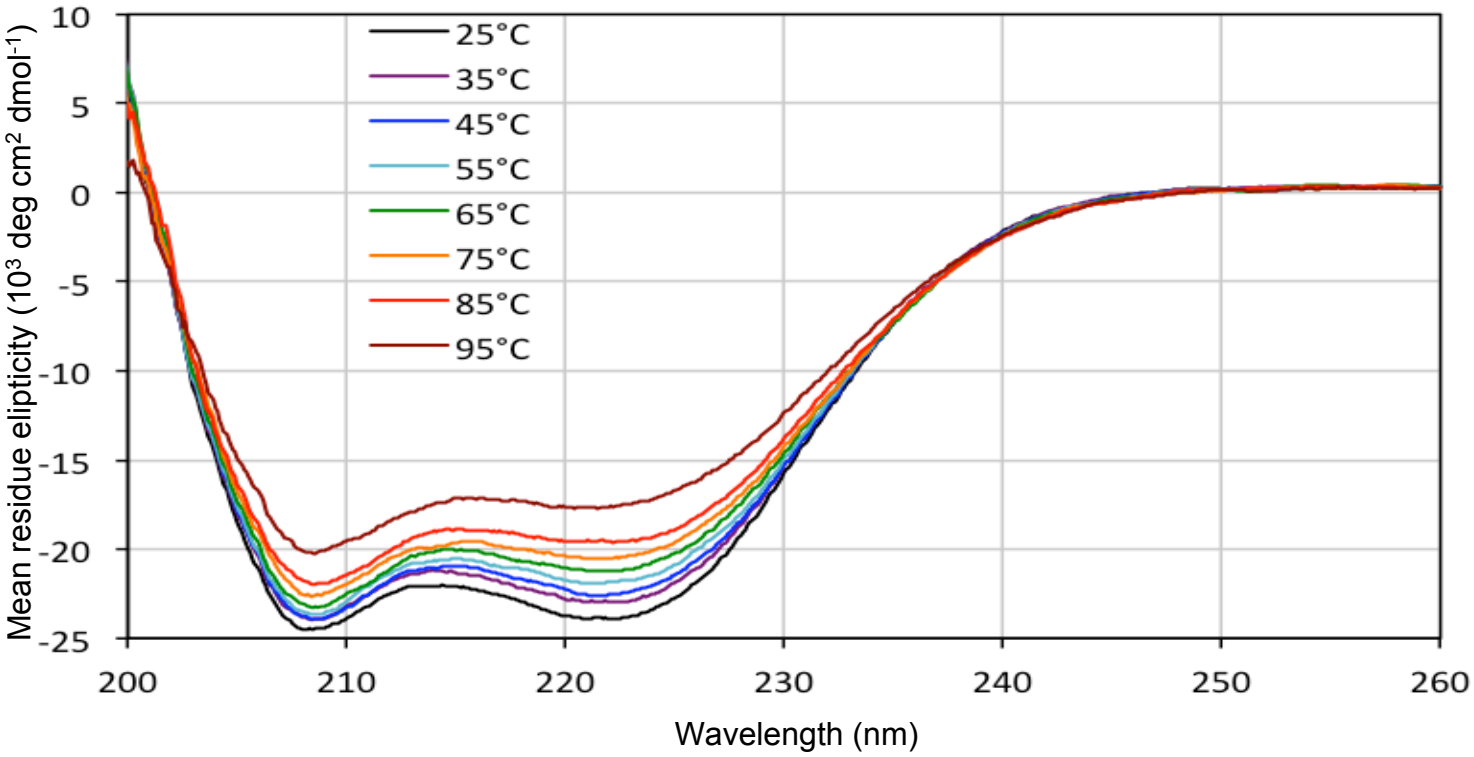

Figure S7. EB22 is thermally stable.
Circular dichroism thermal melt of 18 μM EB22, melting temperature > 95°C.

Figure S8

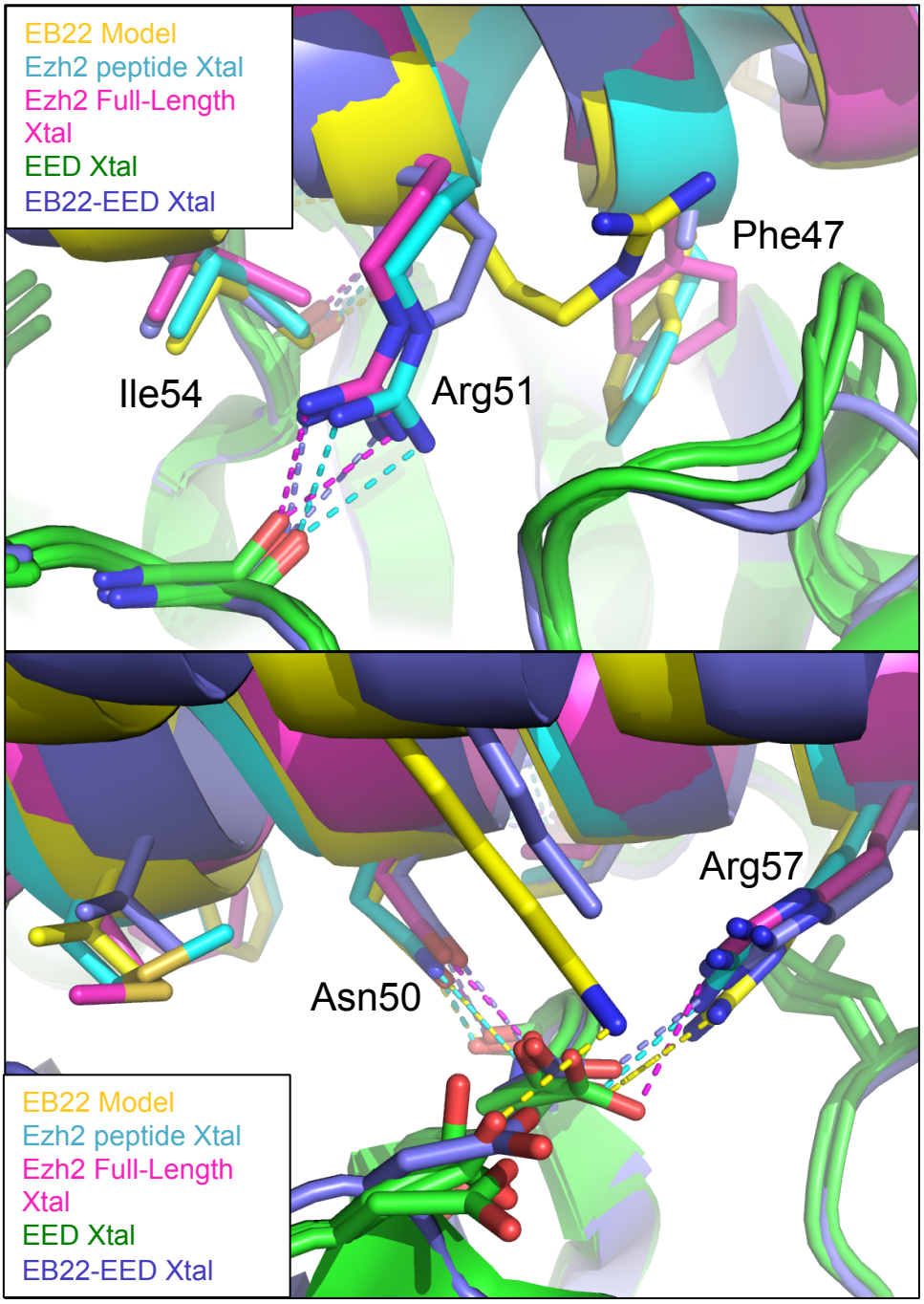

Figure S8. EB22 recapitulates contacts made to EED by Ezh2.
Superposition of the crystal structure of EB22 bound to EED (dark blue) onto the crystal structure of the Ezh2 peptide (cyan), the crystal structure of full-length Ezh2 (PDB ID 5ij8, magenta, or the original Rosetta design model of EB22 (yellow), each bound to EED (green) and shown as cartoons (44,45). Selected residues are shown as sticks and pertinent hydrogen bonds are shown as dashes. The view in the bottom panel is rotated approximately 180° from that in the top panel.

Figure S9

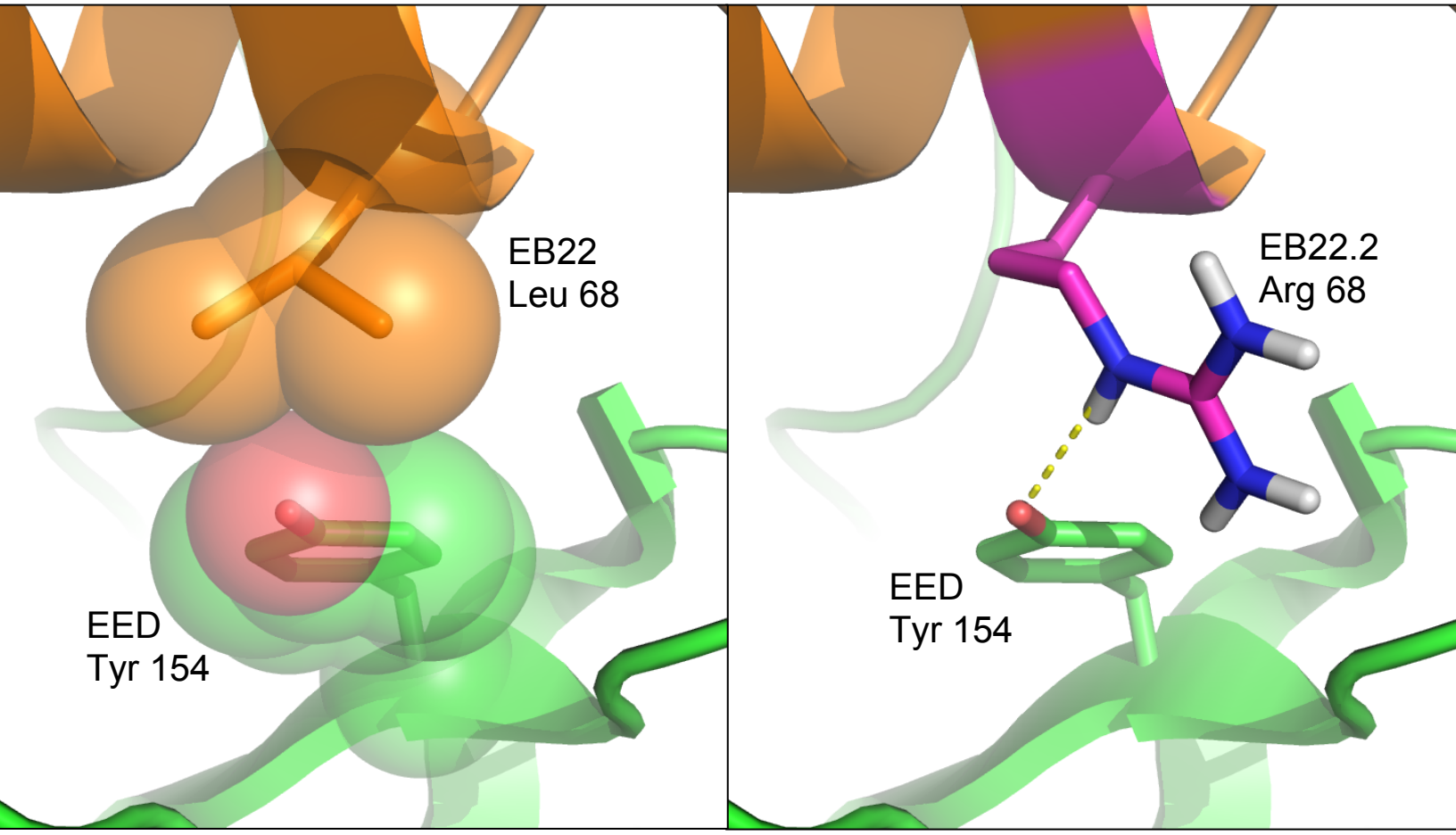

Figure S9. EB22.2 Arg 68 may solvate EED Tyr 154.
Left: A phenix.rosetta_refine model of the crystal structure of EED (green) in complex with EB22 (orange) includes side chains that were not resolved by electron density. phenix.rosetta_refine conformers of EB22-Leu 68 and EED-Tyr 154 are shown as van der Waals spheres. The main chain at and around EB22 Leucine 68 was modeled into weak and discontinuous electron density that suggested extension of the preceding residues' helical conformation, homologous to PDB entry 3LF9. Right: A simulated conformer of the EB22.2 L68R substitution is shown as maroon sticks, and a hypothetical hydrogen bond is shown as a yellow dashed line.

Figure S10

```
Mouse Ezh2:  39...TMFSSNRQKILERTETLNQEWKQRRIQPV...69
Human Ezh2:  39...SMFSSNRQKILERTEILNQEWKQRRIQPV...69
Human Ezh1:  45...ALYVANFAKVQEKTQILNEEWKKLRVQPV...75
```

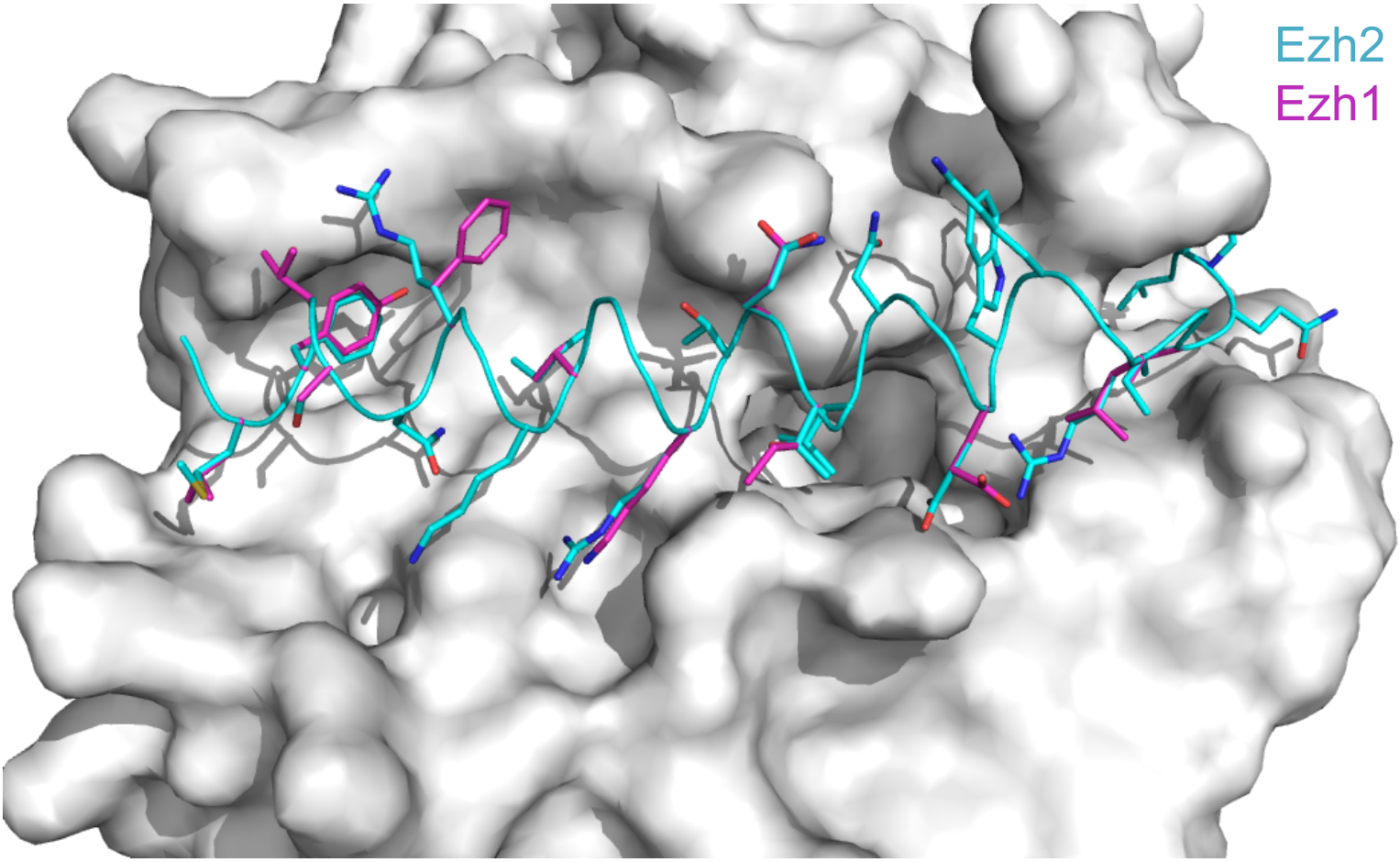

Figure S10. Ezh1 is predicted to bind EED at the same site as Ezh2.
Top panel: Amino acid sequence alignment of the EED binding domains of mouse Ezh2, human Ezh2, and human Ezh1. Blue letters indicate amino acids identical to human Ezh2, green letters indicate similar amino acides, and red letters indicate dissimilar amino acids. Bottom panel: EED is shown as a white molecular surface while the backbone selected sidechains of mouse Ezh2 are shown as a cyan ribbon and sticks, respectively. Selected sidechains of human Ezh1 that differ from the human sequence are shown as magenta sticks.

Figure S11

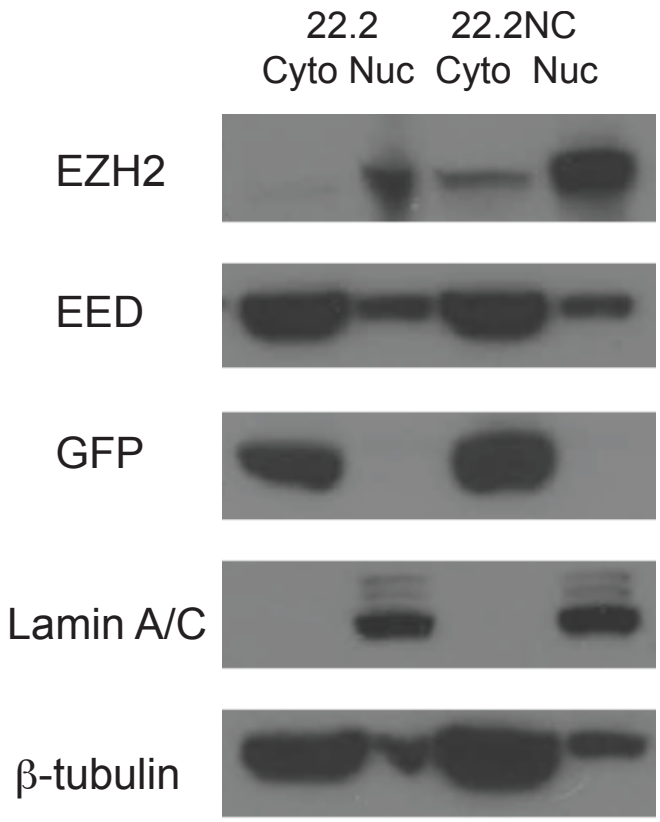

Figure S11. Subcellular localization of EZH2, EED and EB22.2/22.2NC (GFP fusion). Nuclear fractions and cytosol fractions were prepared from K562 cells expressing EB22.2 or EB22.2NC. Lamin A/C is a control for nuclear fractions. β-Tubulin is a control for cytosol fractions. EB22.2 decreases the amount of EZH2 in both nuclear and cytosol fractions.

Figure S12

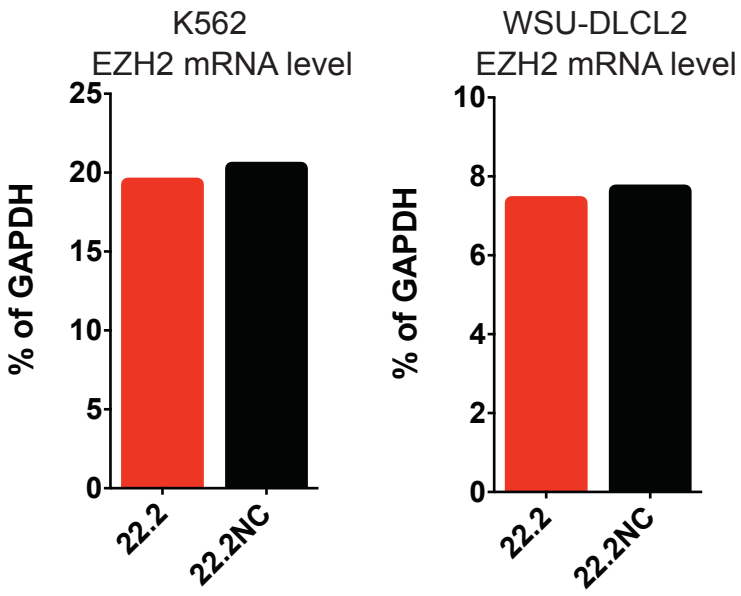

Figure S12. EZH2 mRNA level by RT-qPCR.
We treated both K562 and WSU-DLCL2 with 0.5 μg/ml doxycycline for 4 days and assessed mRNA level of EZH2. EB22.2 does not alter mRNA level of EZH2. mRNA levels of EZH2 in K562 cells are 19.2% (EB22.2) and 20.25% (EB22.2NC) of GAPDH. mRNA levels of EZH2 in WSU-DLCL2 cells are 7.3% (EB22.2) and 7.5% (EB22.2NC) of GAPDH.

Figure S13

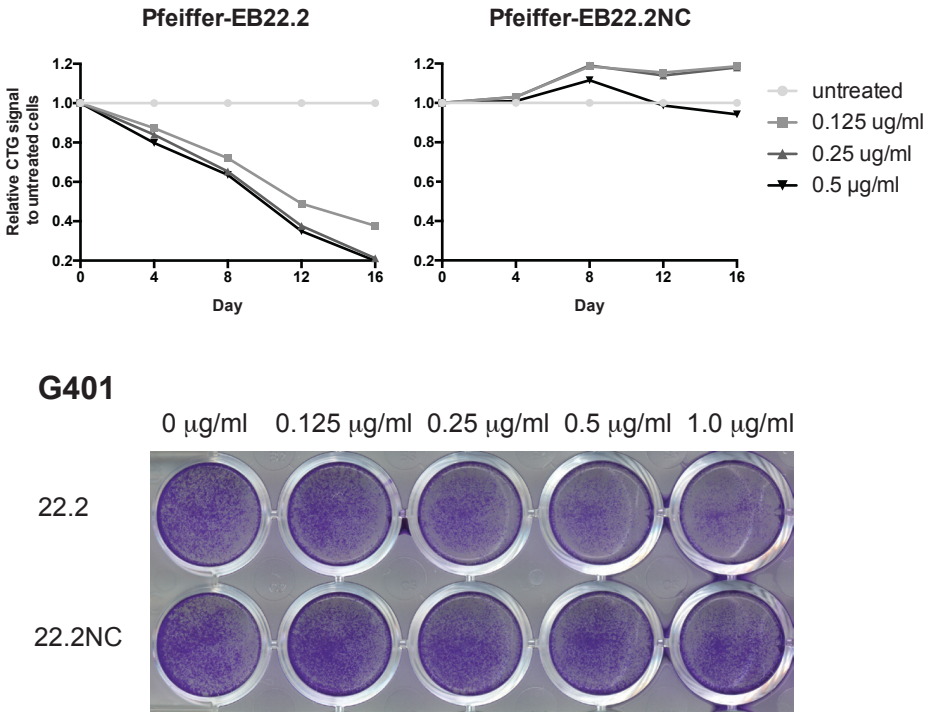

Figure S13. Anti-proliferative effects of EB22.2 on PRC2 dependent cancer cell lines. Both Pfeiffer and G401 cells manifest reduced proliferation upon expression of EB22.2. Proliferation analysis using Cell titer glo (CTG) exhibits 50% reduction with 0.125 µg/ml doxycycline induction for 12 days. Crystal violet staining of G401 cells after 1 week of doxycycline treatment displays anti-proliferative effects of EB22.2 from 0.25 µg/ml of doxycycline.

Figure S14

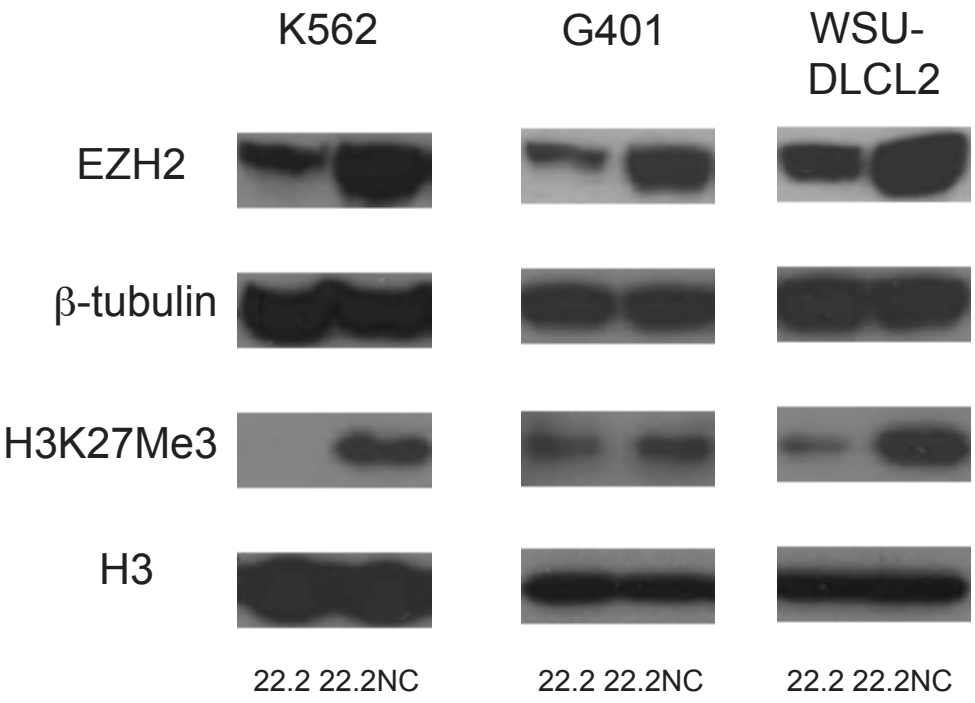

Figure S14. Western blot detection of H3K27me3 and EZH2.
K562, G401 and WSU-DLCL2 cells were treated with 0.5 μg/ml doxycycline for 4 days and assessed for the reduction of EZH2 protein and H3K27Me3 histone mark. EB22.2 decreases EZH2 protein and H3K27me3 histone marks in all of K562, G401 and WSU-DLCL2 cells.

Figure S15

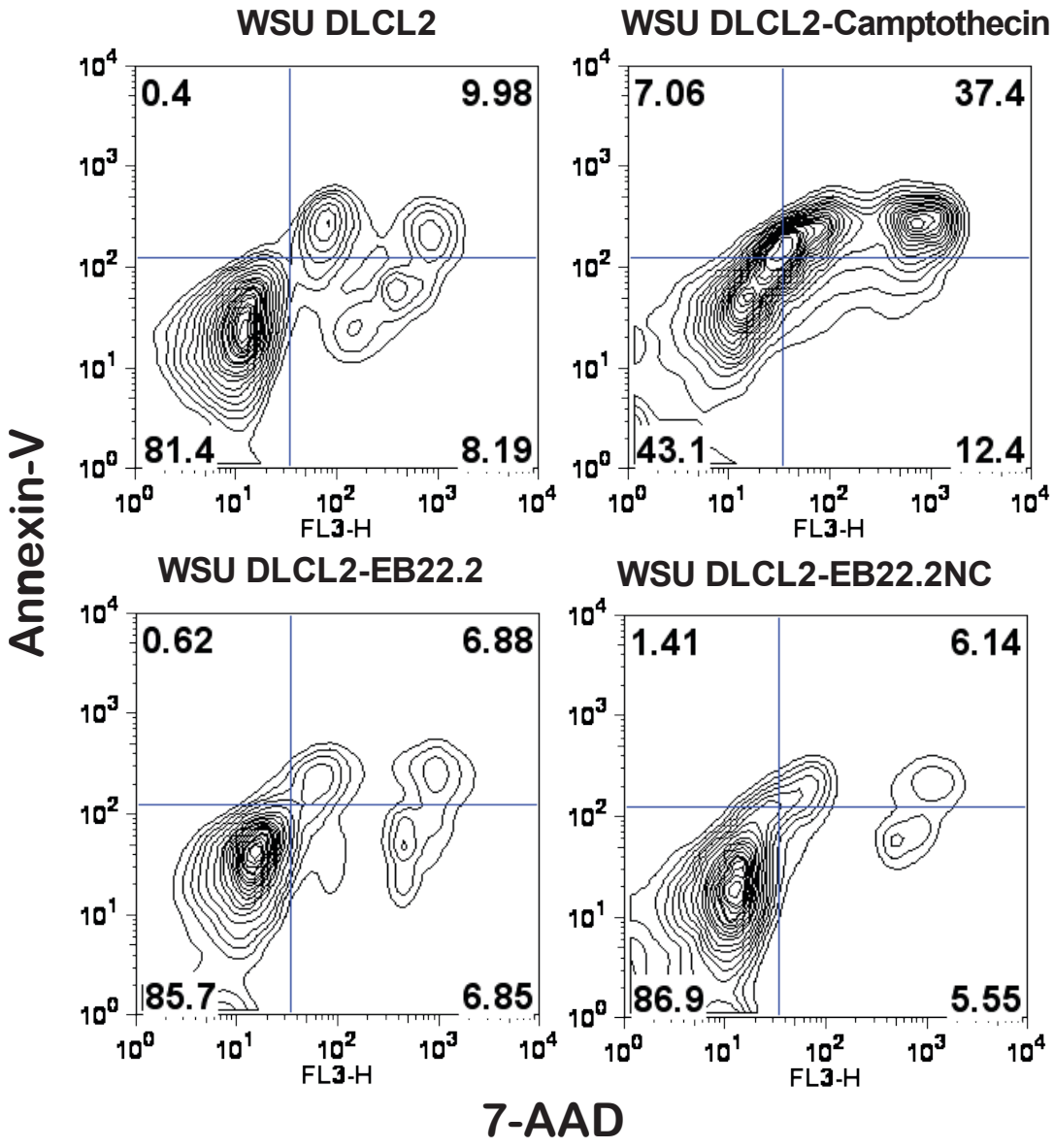

Figure S15. Apoptosis analysis using Anexin-V/7-AAD staining.
We induced expression of EB22.2/EB22.2NC at 0.5 μg/ml for 8 days and analyzed using flow cytometry. EB22.2 does not increase apoptotic cells. Camptothecin (2 μM for 1 day) treated cells are used as a positive control for apoptotic cells.

Figure S16

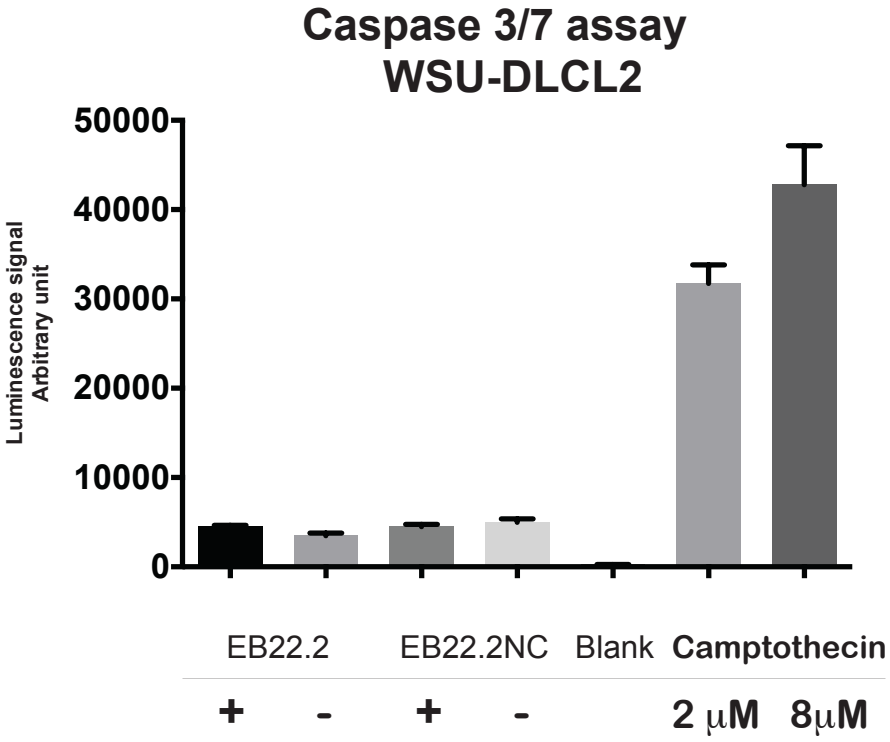

Figure S16. Apoptosis analysis by assessing caspase 3/7 activities.
WSU-DLCL2 cells were treated with doxycycline (0.5 μg/ml) for 8 days and analyzed for the activities of caspase 3/7. EB22.2 does not alter activities of caspase 3/7, indicating that EB22.2 does not induce apoptosis as shown by Figure S15. Camptothecin treated cells (2 μM, 8 μM for 1day) are used as a positive control for apoptotic cells. 'Blank' is a signal of media only.

Figure S17

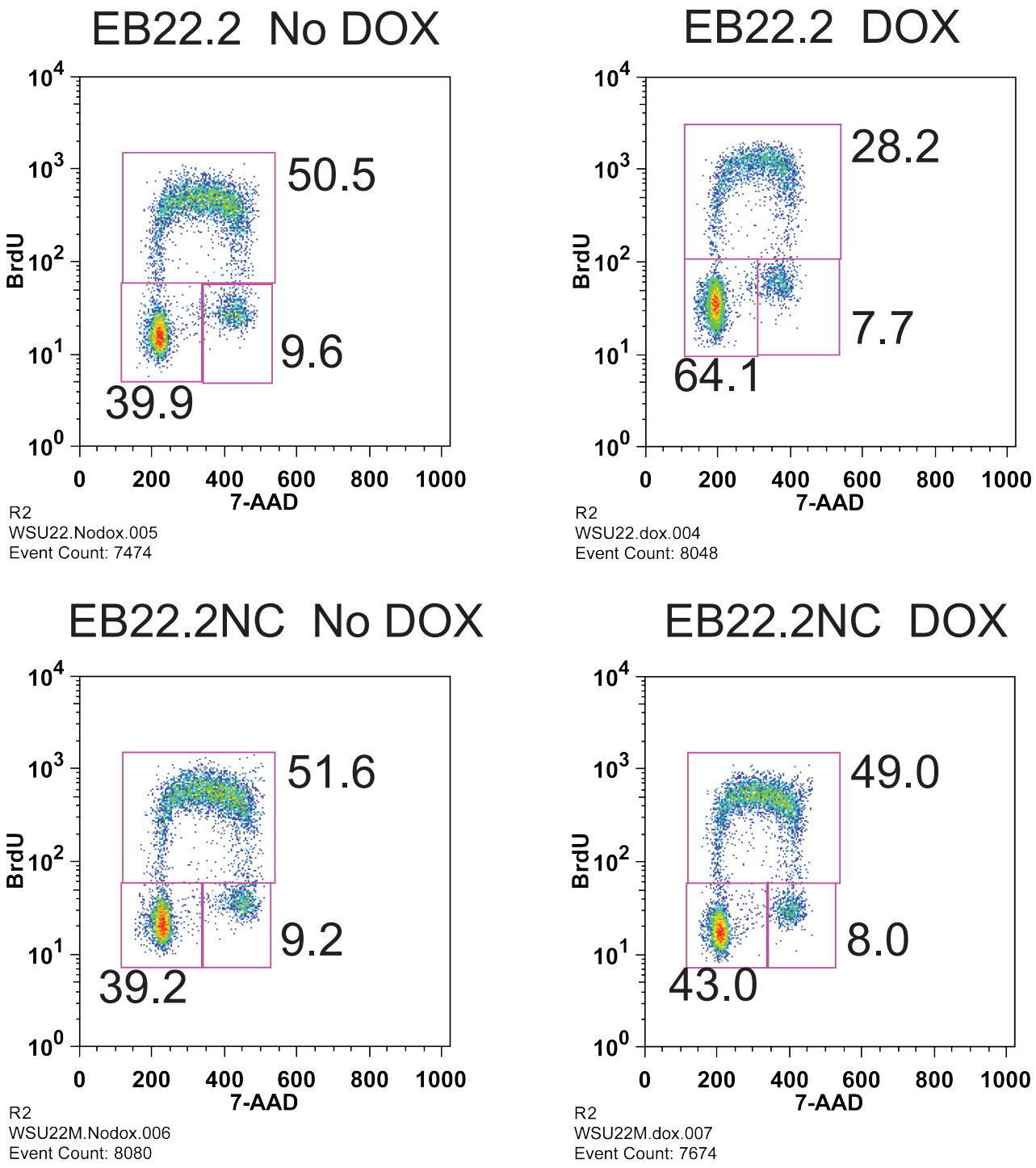

Figure S17. Cell cycle analysis using Brd-U/7-AAD staining.
We induced expression of EB22.2/EB22.2NC in WSU-DLCL2 cells (0.5 µg/ml, 4 days) and analyzed for cell cycle using flow cytometry. Cells were exposed to 10 µM Brd-U for 30 min. EB22.2 increase the population of cells in G0/G1 cycle.

Figure S18

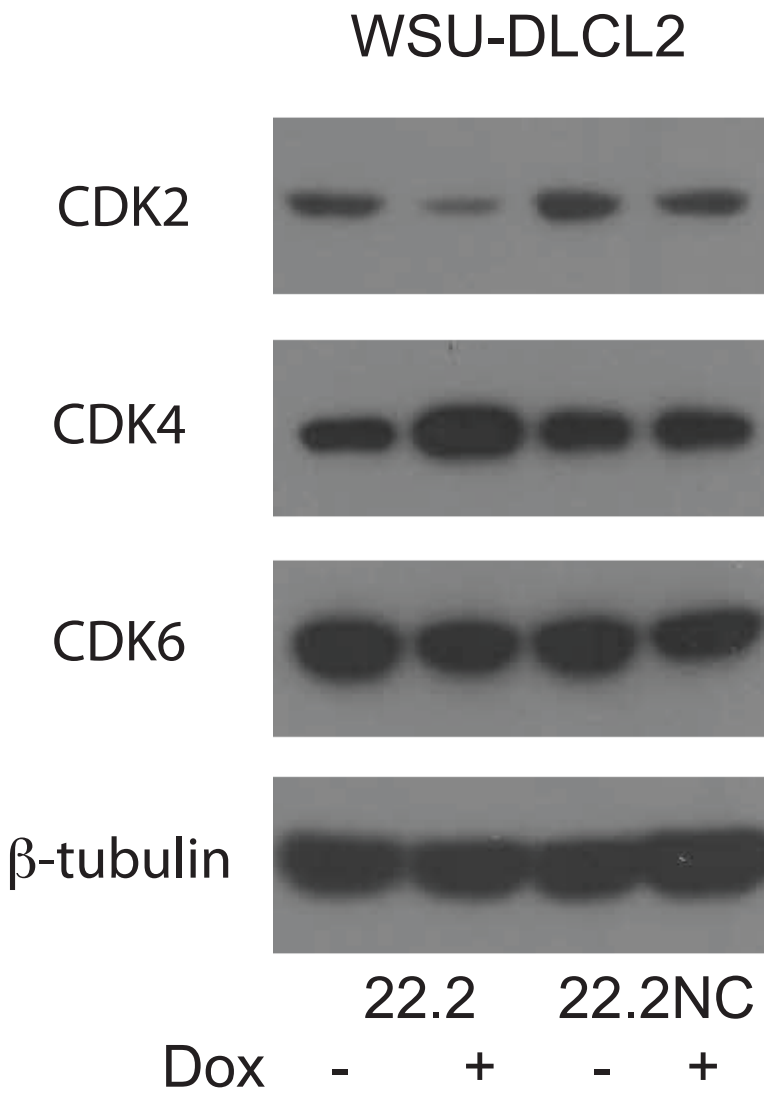

Figure S18. Western blot detection of cell cycle regulatory proteins.
We treated WSU-DLCL2 cells with doxycycline (0.5 μg/ml, 4 days) and assessed the protein level of cell cycle regulatory proteins. EB22.2 dysregulates cell cycle, decreases CDK2 expression and increases CDK4 expression.

Figure S19

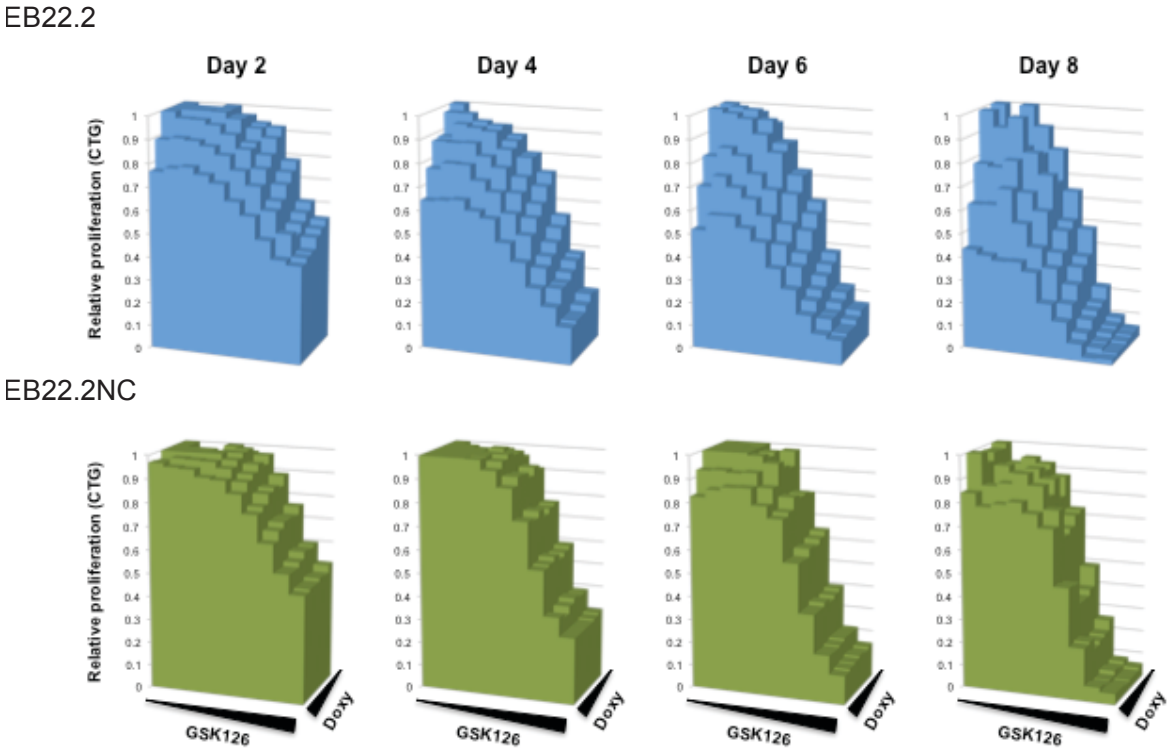

Figure S19. Collaborative effects of EB22.2 and GSK126.
WSU-DLCL2 cells were simultaneously treated with EB22.2 and GSK126 in varying concentrations for 8 days. We measured the proliferation of WSU-DLCL2 cells every other day. EB22.2 and GSK126 reduce proliferation of WSU-DLCL2 cells in a collaborative manner. ‘GSK126 0 μM’ means DMSO only (DMSO control). Each data point represent mean for independent experiments performed in duplicate.

Figure S20

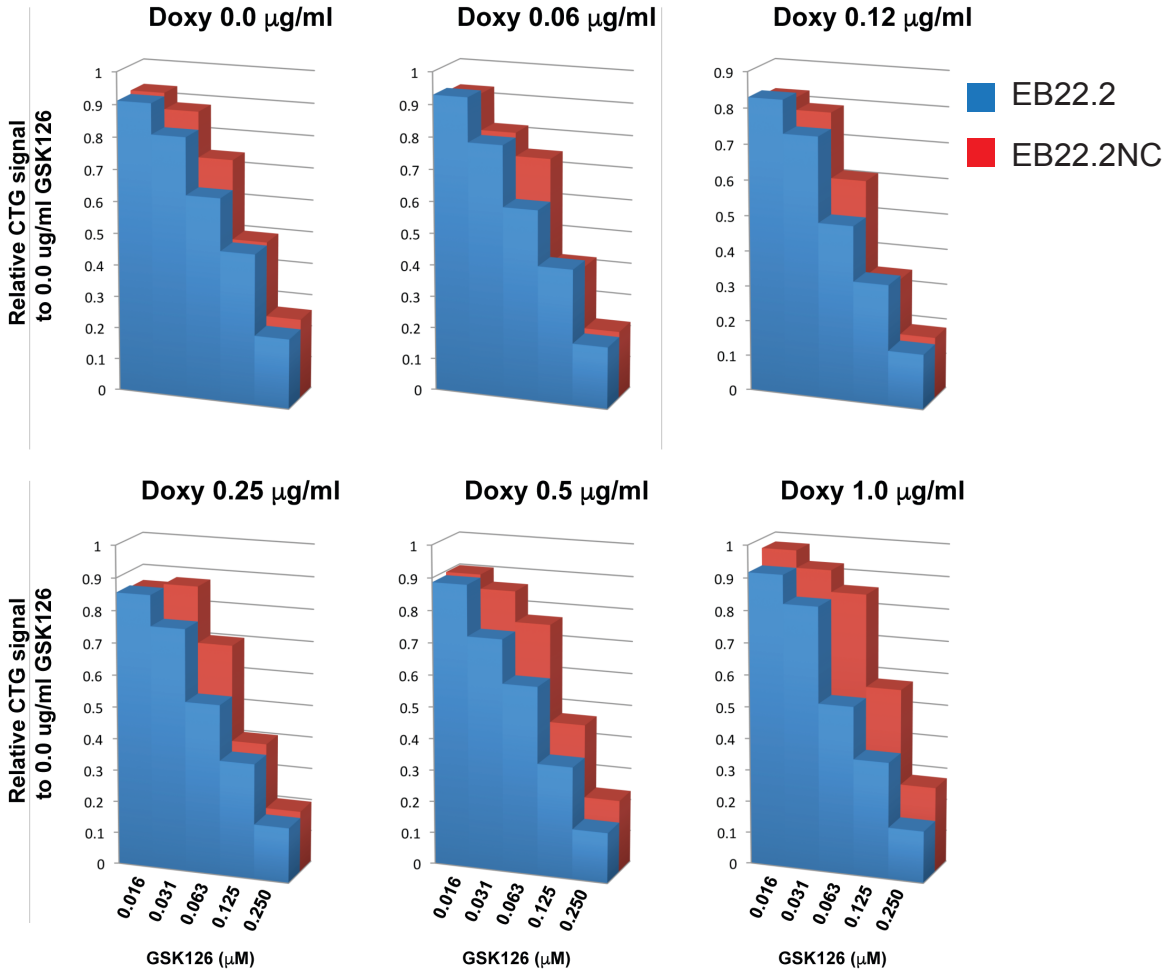

Figure S20. Sensitization of WSU-DLCL2 cells to GSK126.
WSU-DLCL2 cells were treated as in Figure S19. CTG signal of each points were normalized by the data point of GSK126 0 μM at given doxycycline concentration. Reduction of EZH2 sensitizes WSU-DLCL2 cells to low concentrations of GSK126 (31, 63 and 125 nM) in doxycycline dependent manner. Each data point represent mean for independent experiments performed in duplicate.

Figure S21

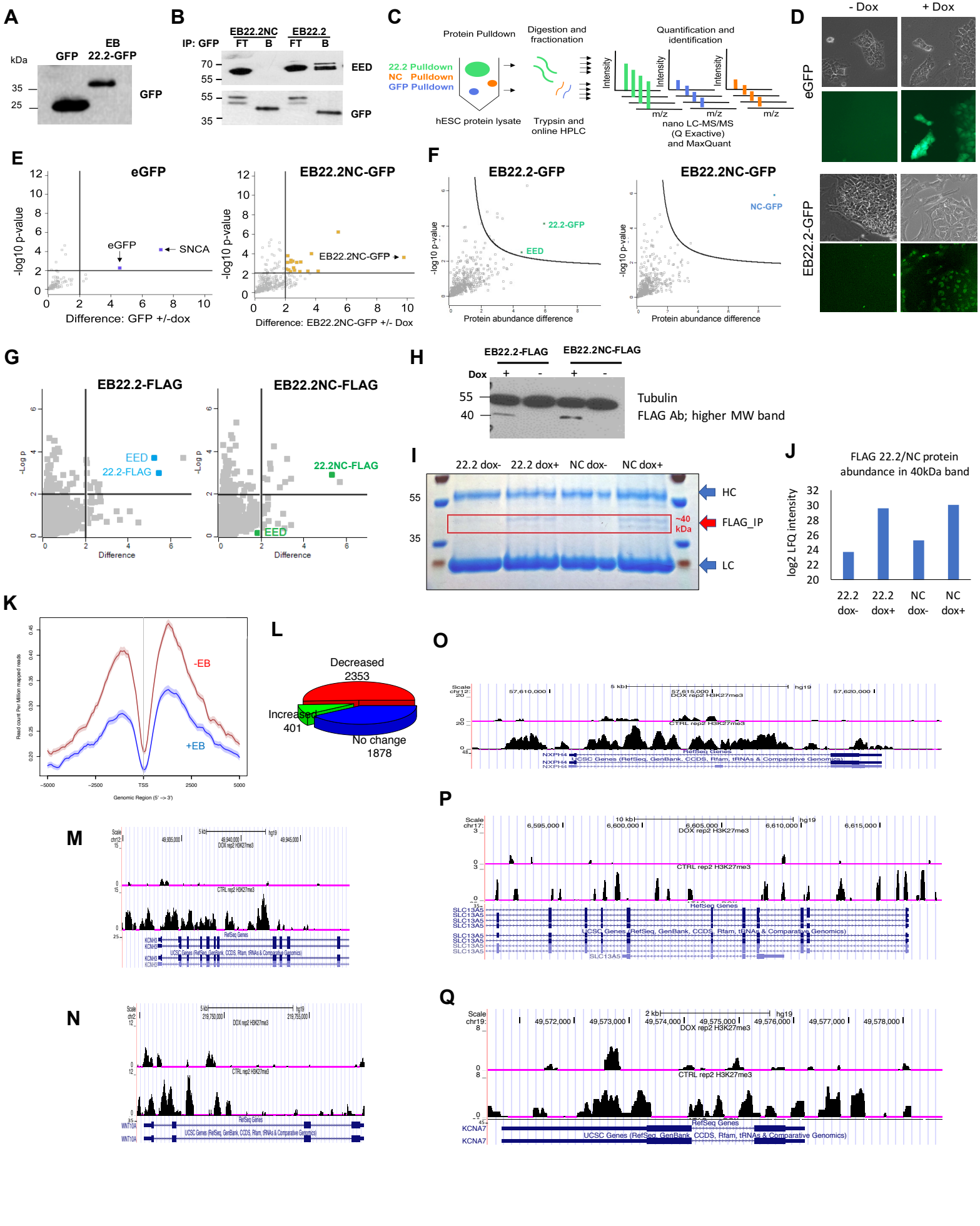
A
kDa
GFP
EB 22.2-GFP
35
25
GFP
B
IP: GFP
EB22.2NC
EB22.2
FT
B
FT
B
70
55
EED
55
35
GFP
C
Protein Pulldown
Digestion and fractionation
Quantification and identification
22.2 Pulldown
NC Pulldown
GFP Pulldown
hESC protein lysate
Trypsin and online HPLC
Intensity
m/z
nano LC-MS/MS (Q Exactive) and MaxQuant
D
- Dox
+ Dox
eGFP
EB22.2-GFP
E
eGFP
EB22.2NC-GFP
-log10 p-value
eGFP
SNCA
EB22.2NC-GFP
Difference: GFP +/-dox
Difference: EB22.2NC-GFP +/- Dox
F
EB22.2-GFP
EB22.2NC-GFP
22.2-GFP
EED
NC-GFP
-log10 p-value
Protein abundance difference
G
EB22.2-FLAG
EB22.2NC-FLAG
-Log p
EED
22.2-FLAG
22.2NC-FLAG
EED
Difference
H
EB22.2-FLAG
EB22.2NC-FLAG
Dox
+
-
+
-
55
40
Tubulin
FLAG Ab; higher MW band
I
22.2 dox-
22.2 dox+
NC dox-
NC dox+
55
35
~40 kDa
HC
FLAG_IP
LC
J
FLAG 22.2/NC protein abundance in 40kDa band
log2 LFQ intensity
32
30
28
26
24
22
20
22.2 dox-
22.2 dox+
NC dox-
NC dox+
K
Read count Per Million mapped reads
-EB
+EB
-5000
-2500
TSS
2500
5000
Genomic Region (5' -> 3')
L
Decreased 2353
Increased 401
No change 1878
O
DOX rep2 H3K27me3
CTRL rep2 H3K27me3
RefSeq Genes
NXPH4
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
P
DOX rep2 H3K27me3
CTRL rep2 H3K27me3
RefSeq Genes
SLC13A5
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
M
DOX rep2 H3K27me3
CTRL rep2 H3K27me3
RefSeq Genes
KCNH3
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
Q
DOX rep2 H3K27me3
CTRL rep2 H3K27me3
RefSeq Genes
KCNA7
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
N
DOX rep2 H3K27me3
CTRL rep2 H3K27me3
RefSeq Genes
WNT10A
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

Figure S21. EED binder binds EED in hESC.
A. Immunoblot of overexpressed EB22.2 after 48h (HeLa cells). **B**. EB22.2 but not EB22.2NC binds to EED (co-IP, HeLa cells, 48h +Dox). FT: Flow through, B: GFP-Trap beads. **C.** Proteomic workflow used to identify proteins co-immunoprecipated by EB22.2, EB22.2NC, or GFP. Peptides were identified by mass spectrometry, and the corresponding protein abundance was measured using label free quantification. **D.** Elf1-EB22.2 exhibits nuclear localization and differentiated morphology upon 5d dox induction. Imaged at X20 magnification. **E**. EED does not precipitate with Elf1-GFP or Elf1 EB22.2NC. Cells were treated with + 0.5µg/mL Dox, and proteins associated with GFP were co-immunoprecipated and quantified using mass spectrometry. Protein abundance (label free quantification (LFQ) intensity) differences were quantified between plus and minus dox treated cell lines. Black line represents the 5% false discovery rate calculated in Perseus 1.4.1.3 N=6 technical measurements. **F.** EB22.2 but not EB22.2NC binds to EED. Cells were treated with + 2.0µg/mL dox, and proteins enriched in EB or NC pulldowns were identified by mass spectrometry. Black lines represent a cut-off of 0.01 (p-value) and a difference (log2 transformed) greater than 2 +/- dox. **G**. EED binds to EB22.2_FLAG but not EB22.2NC_FLAG, co-IP from HeLa cells quantified by mass spectrometry. Proteins associated with black lines represent a cut-off of 0.01 (p-value) and a difference (log2 transformed) greater than 2. **H.** Elf1-EB22.2-FLAG and Elf1-EB22.2NC-FLAG express a higher molecular weight band sized at 40 kDa than the expected 17kDa. **I-J.** EB22.2-Flag and NC22.2-Flag identified at the 40kDa region. **I**. Coomassie stain of EB22.2 +/- Dox and NC22.2 +/-Dox. Red box indicates area of gel that was excised and analyzed by mass spectrometry. NC; negative control, HC; IgG heavy chain, LC; IgG light chain**. J.** FLAG protein abundance was quantified using label free quantification (LFQ). **K-L.** 3 days of treatment with Dox in Elf1 EB22.2-FLAG cells decreases H3K27me3 marks. (K) Averages of H3K27me3 profile of 4632 genes with peaks within 5KB of TSS in untreated Elf1 were plotted from ChIPseq data of Elf1 EB22.2-FLAG cells without Dox treatment (-EB) and with Dox treatment (+EB). (L) Pie chart representation of changes in H3K27me3 marks of 4632 genes in Elf1 2iLIF +EB compared to Elf1 2iLIF -EB. **M-Q**. Refseq gene positions and H3K27me3 profiles detected by ChIP in Elf1-EB22.2-FLAG treated or not with Dox for 3 days.

Figure S22

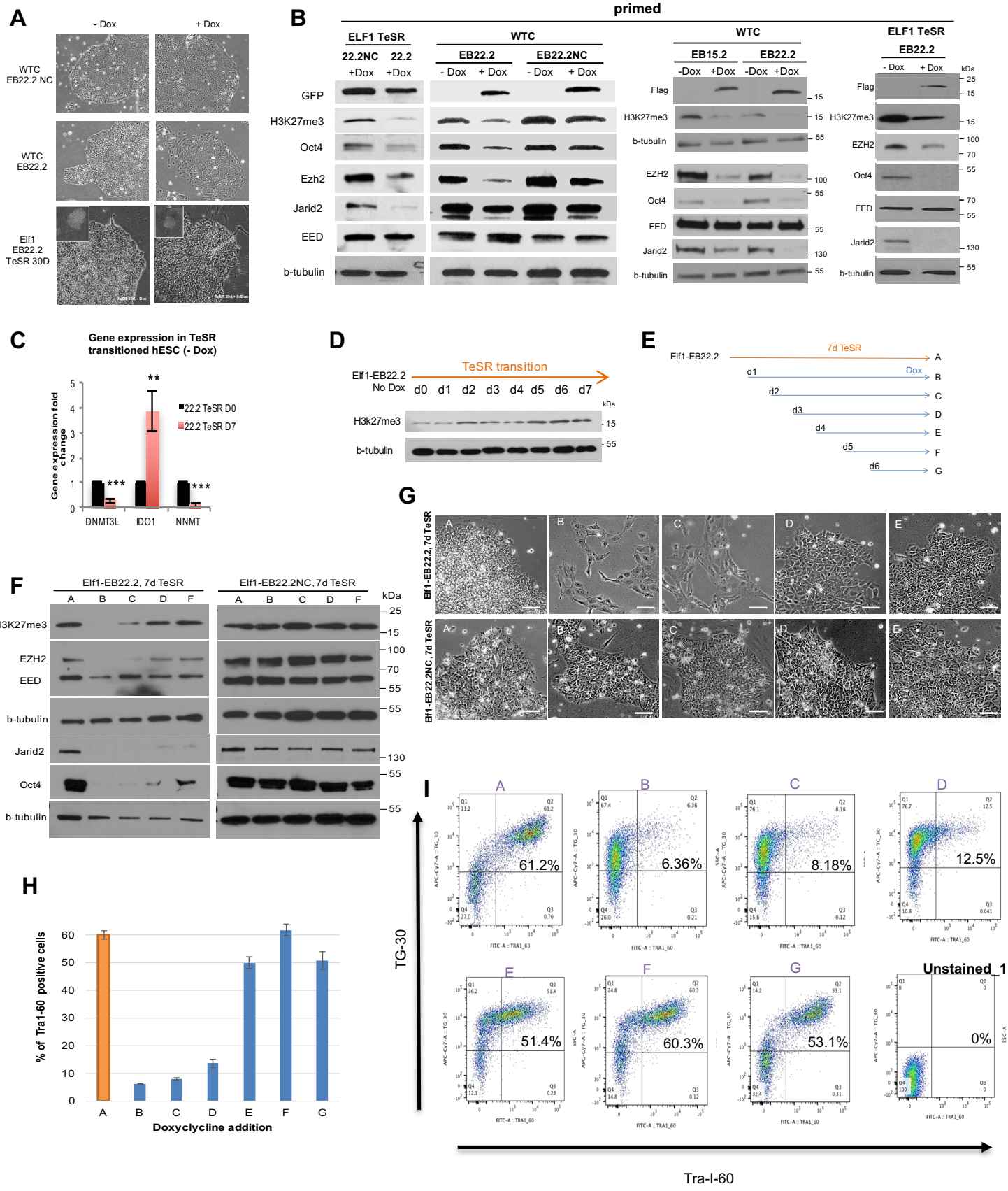
A
- Dox
+ Dox
WTC EB22.2 NC
WTC EB22.2
Elf1 EB22.2 TeSR 30D
B
primed
ELF1 TeSR
22.2NC
22.2
+Dox
WTC
EB22.2
EB22.2NC
- Dox
+ Dox
WTC
EB15.2
EB22.2
-Dox
+Dox
ELF1 TeSR
EB22.2
kDa
GFP
H3K27me3
Oct4
Ezh2
Jarid2
EED
b-tubulin
Flag
EZH2
C
Gene expression in TeSR transitioned hESC (- Dox)
Gene expression fold change
22.2 TeSR D0
22.2 TeSR D7
DNMT3L
IDO1
NNMT
D
Elf1-EB22.2 No Dox
TeSR transition
d0 d1 d2 d3 d4 d5 d6 d7
H3k27me3
b-tubulin
E
Elf1-EB22.2
7d TeSR
Dox
d1
d2
d3
d4
d5
d6
F
Elf1-EB22.2, 7d TeSR
Elf1-EB22.2NC, 7d TeSR
G
Elf1-EB22.2, 7d TeSR
Elf1-EB22.2NC, 7d TeSR
H
% of Tra1-60 positive cells
Doxycyclicine addition
I
TG-30
Tra-I-60
61.2%
6.36%
8.18%
12.5%
51.4%
60.3%
53.1%
0%
Unstained_1

Figure S22. EED binder disrupts primed pluripotency.
A. Peripheral differential morphology found in primed cell line upon Dox induction. Top two panels: WTC primed cells were transfected with EB22.2 NC-FLAG, EB22.2-FLAG, induced with Dox for 3 days and imaged (X20). Bottom panel: Elf1-EB22.2-FLAG cells were grown in the presence of TeSR for 30d (over multiple passages) followed by induction with Dox for 5d and imaging. B. PRC2 is disrupted, and H3K27me3 and Oct4 are significantly down-regulated in primed (Elf1 primed, WTC) EB22.2 but not NC22.2 following Dox induction (2μg/mL; 5 or 3 days, respectively), as analyzed by immunoblot. We observe variable extents of H3K27me3 marks in our EB22.2 experimental conditions, as common with PRC2 disruption. C. RT-qPCR analysis of naïve (DNMT3L, NNMT) and primed (IDO1) markers reveals that Elf1-EB22.2 (-Dox) have transitioned towards primed stage after 7 days of culture in mTeSR media (n=3, st err of mean, **p<0.01, ***p<0.001, 2-tailed t-test). D. Kinetics of H3K27me3 level during naïve to primed transition of Elf1-EB22.2 through 7-day culture in TeSR media. Immunoblot analysis of 7d transition of Elf1-EB22.2 (without dox induction) reveals elevating H3K27me3 marks. E. Transition experiment time point display; A, cells were grown in the presence of TeSR for 7d, no dox was added (orange), B, induction of doxycycline at day 1 (d1) (blue), C, induction of doxycycline at day 2 (d2), D, induction of doxycycline at day 3 (d3), E, induction of doxycycline at day 4 (d4), F, induction of doxycycline at day 5 (d5), G, induction of doxycycline at day 6 (d6).F. PRC2 is disrupted following doxycycline induction of transitioned Elf1-EB22.2-FLAG compared to Elf1-EB22.2NC-FLAG. Transitioned cells were grown as described in A according to time points A-G, and analyzed by immunoblot. G. Doxycycline induced Elf1-EB22.2-FLAG but not Elf1-EB22.2NC-FLAG display differentiated morphology at the first two days of transition compared to 7d primed morphology. Elf1-EB22.2-FLAG and Elf1-EB22.2NC-FLAG were grown on matrigel in the presence of TeSR according to the time points described in A and were imaged using confocal microscopy (X20). Scale bar=100um. H-I. Transitioned Elf1-EB22.2-FLAG B, C and D time points lose their stemness markers. Transitioned cells were grown as described in A according to time points A-G, immunostained with surface stem cell markers Tra-I-60 and TG-30 and analyzed using FACS. Quantitative representation of Tra-1-60 positive cells for each experiment is shown in H.

Figure S23

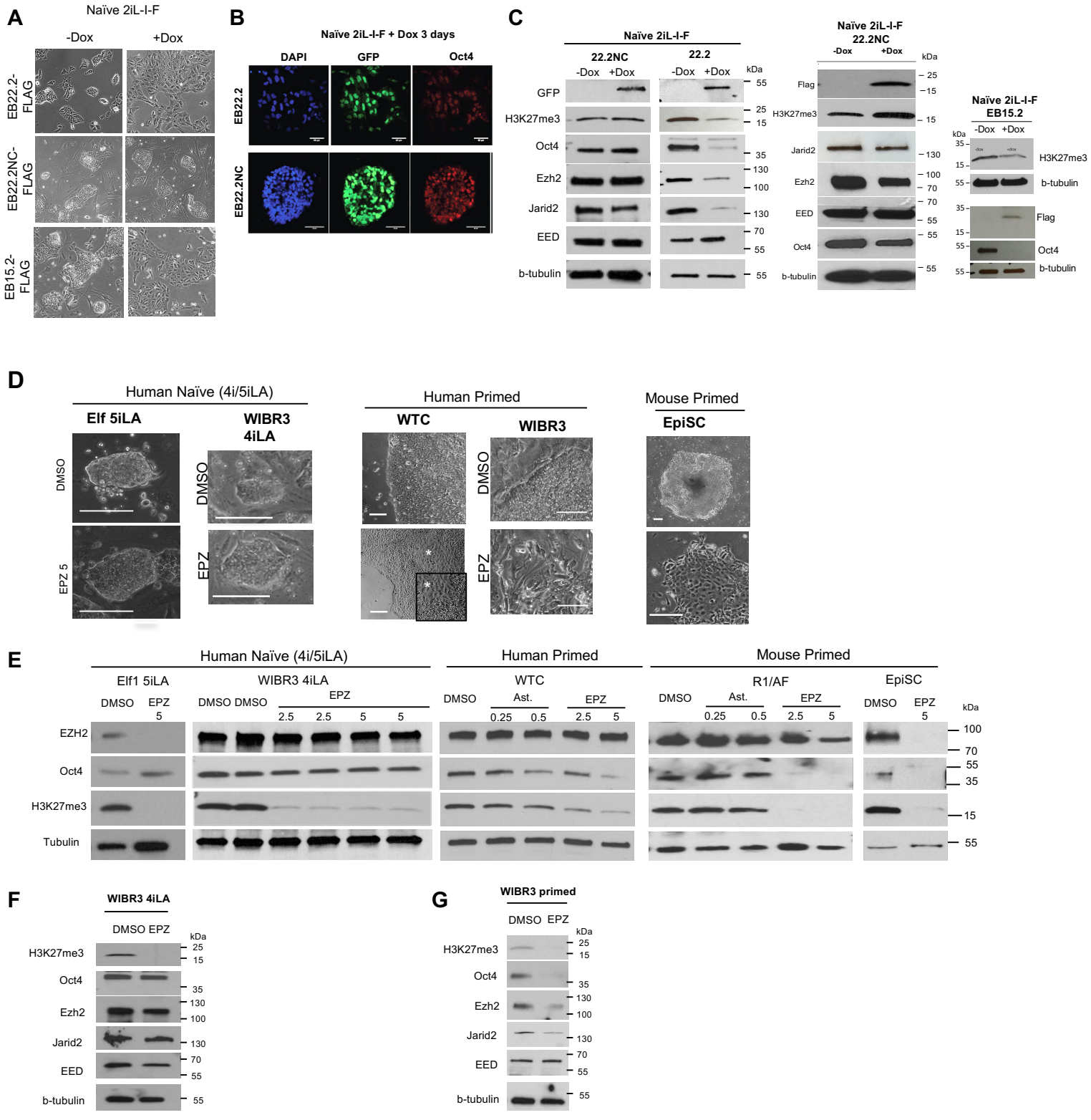
A
Naïve 2iL-I-F
-Dox
+Dox
EB22.2-FLAG
EB22.2NC-FLAG
EB15.2-FLAG
B
Naïve 2iL-I-F + Dox 3 days
DAPI
GFP
Oct4
EB22.2
EB22.2NC
C
Naïve 2iL-I-F
22.2NC
22.2
-Dox +Dox
kDa
GFP
H3K27me3
Oct4
Ezh2
Jarid2
EED
b-tubulin
Naïve 2iL-I-F
22.2NC
Flag
Naïve 2iL-I-F
EB15.2
D
Human Naïve (4i/5iLA)
Elf 5iLA
WIBR3 4iLA
Human Primed
WTC
WIBR3
Mouse Primed
EpiSC
DMSO
EPZ 5
EPZ
E
Human Naïve (4i/5iLA)
Elf1 5iLA
WIBR3 4iLA
Human Primed
WTC
Mouse Primed
R1/AF
EpiSC
DMSO
EPZ
Ast.
EZH2
Oct4
H3K27me3
Tubulin
F
WIBR3 4iLA
DMSO EPZ
H3K27me3
Oct4
Ezh2
Jarid2
EED
b-tubulin
G
WIBR3 primed
DMSO EPZ
H3K27me3
Oct4
Ezh2
Jarid2
EED
b-tubulin

Figure S23. EED binder disrupts Naive pluripotency.
A. Naïve Elf1EB22.2-flag and EB15.2-flag exhibit differentiated morphology upon 3d Dox induction in the presence of CM+2iL-I-F but not NC22.2 (magnification 10X). **B**. Stem cell marker Oct4 (red) is lost in Naïve Elf1 EB22.2-GFP 2iL-I-F but not in NC22.2-GFP. Cell were grown on matrigel in the presence of CM+2iL-I-F and induced for 3d in Dox followed by confocal imaging. Scale bars represent 50 μm. **C**. PRC2 is disrupted and Oct4 down-regulated in Naive 2iL-I-F EB22.2 however NC22.2 remained unchanged following 3d of Dox induction. EB15.2 flag shows similar phenotype at the same conditions. Analysis by immunoblot. **D**. Human ground naïve Elf1 5iLA, human primed WTC, and mouse primed EpiSC morphology under treatment of DMSO or EPZ-6438 (5μM). Scale bar=100μm. **E.** Immunoblot of Elf15iLA, WIBR3 4iLA, WTC, mouse R1AF and Epi ESC treated with DMSO, Astemizole 0.25 or 0.5 μM, EPZ-6438 2.5 or 5μM. **F-G-F.** Immunoblot of WIBR3 4iLA and WIBR3 primed treated with DMSO or 5uM EPZ-6438 for 5d.