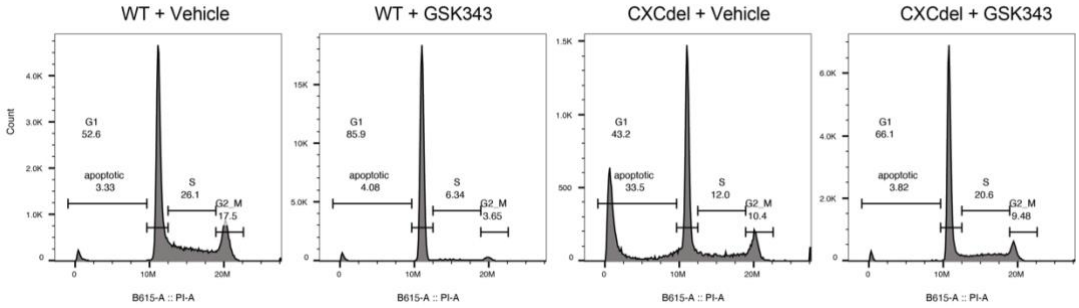
Supplementary Information for

Drug addiction unveils a repressive methylation ceiling in EZH2-mutant lymphoma

Hui Si Kwok,[#] Allyson M. Freedy,[#] Allison P. Siegenfeld, Julia W. Morriss, Amanda L. Waterbury, Stephen M. Kissler, & Brian B. Liau[*]

Supplementary Figure 1. Gating strategy for cell cycle analysis.
Representative gating strategy for cell cycle analysis. Wild-type and CXCdel$^{+/-}$ Karpas-422 cells were treated with 1 µM GSK343 or vehicle for 7 days. Cell cycle phases were defined by linear measurements of Propidium Iodide (PI) staining. Cells in G_1 phase were identified as having 1n DNA content, cells in $G_{2/M}$ phase as 2n, cells in S phase as cells between G_1 and $G_{2/M}$ phases, and apoptotic cells as < 1n.



Supplementary Table 1. (Related to Fig. 2f-g). Proportions, Z-values, and slopes for simulated sgRNAs relative to sgRNA 1.

sgRNA	Time	Proportion	Z	Slope (σ)
1	0	0.33	0	–
	8	0.36	0	0
	16	0.52	0	0
2	0	0.33	0	–
	8	0.33	-0.08	-0.01
	16	0.44	-0.16	-0.01
3	0	0.33	0	–
	8	0.3	-0.16	-0.02
	16	0.04	-2.64	-0.31

Supplementary Table 2. Oligonucleotides used in this study.

Oligonucleotide	Sequence (5' - 3')
sgRNA for luciferase non-targeting control	GGATCTACTGGGTTACCTAA
sgRNA for mutagenizing EZH2 I109/M110	GGGAGACCAAGAATACATTA
sgRNA for mutagenizing EZH2 A596/D597	TTTACTGTCCCAATGGTCAG
sgRNA for mutagenizing EZH2 D597	CTTGTGGAGCCGCTGACCAT
sgRNA for mutagenizing EZH2 D597/H598	TTGTGGAGCCGCTGACCATT
sgRNA for mutagenizing EZH2 Q575	CAAGCAGTGCCCGTGCTACC
ssODN for EZH2 Q575 knock-in	TGCCGCTGCAAAGCACAGTGCAACACCAAGCGGTGCCCG TGCTACCTGGCTGTCCGAGAGTGT
ssODN for EZH2 CXCdel knock-in	TCCGAGAGTGTGACCCTGACCTCTGTCTTAAAAATGTGTC CTGCAAGAACTGCAGTATTC
EZH2_Exon14_gDNA_Fw for NGS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAAA ACCGCTTTCCGGGATG
EZH2_Exon14_gDNA_Rv for NGS	TGGAGTTCAGACGTGTGCTCTTCCGATCTATCATCTAAGG CAATCCTGACATTTGC
EZH2_Exon14_Q575_gDNA_Fw for NGS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCC CTTGCCACGTATCTTTCTGCAAG
EZH2_Exon3_gDNA_Fw for NGS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTT GGATTTTCCAACACAAGTCAT
EZH2_Exon3_gDNA_Rv for NGS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTATGCTTTTTT ACTTCAAATAAGTTATTATC
EZH2_SETdomain_gDNA_Fw for NGS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCC GCAAGGGTAACAAAATTCG
EZH2_SETdomain_gDNA_Rv for NGS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGTCCATCATC ACAGGACTGAAAAGG
EED_Exon10_gDNA_Fw for NGS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGA AACCTGGCAAGATGGAAGAT
EED_Exon10_gDNA_Rv for NGS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTAGAATTAC AATGGTCTTTCATAAGTCAG
EZH2_Exon14_cDNA_Rv for NGS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTGATCACCGT TAACCATCATAACTTTTGC
SETD2 shRNA #1	ATGGTGTAACTTATGCATTAA
SETD2 shRNA #2	AAGCAGGACACTATATCTAAT
SETD2_qPCR_FW	TGCTTCTAGTCGATTTTTGCCC
SETD2_qPCR_REV	AGGGTTTGGAGTATCACTTTGC

Supplementary Methods

Calculation of addiction score

Motivation for the addiction score

In pooled ecological competition experiments, the hallmark of drug addiction is when a lineage's intrinsic fitness is greater when on drug than when off drug (*i.e.,* the intrinsic fitness declines when drug is removed). Thus, addiction may manifest for a given lineage as a faster growth rate when on drug than when off drug. However, other processes can also cause a lineage to grow faster on drug than off drug, such as out-competition by more fit strains when drug is removed (*i.e.,* the relative fitness declines when drug is removed, despite no change in intrinsic fitness). To identify lineages that may be addicted to drug, we therefore require a way to distinguish changes in intrinsic fitness from changes in relative fitness. We can accomplish this by modeling each lineage's proliferation over time using the laws of competitive logistic growth and comparing each lineage's growth against a reference lineage that is unaffected by the presence of drug (a "purely resistant" lineage). This approach yields an "addiction score", defined as the difference between a given lineage's intrinsic fitness when on drug *vs.* off drug, so that positive values indicate lineages that may be addicted to drug. The following sections outline the calculation of the addiction score.

Derivation of the addiction score

Dynamics of the overall population.

Consider a population of cells which consists of many sub-populations (sgRNAs), each of which may have a different intrinsic fitness. We can begin by describing the size N of the total population.

Let us assume that the population's growth is governed by logistic dynamics: when the population size is small, the population exhibits exponential growth, but when the population approaches the carrying capacity K, the growth plateaus. These dynamics are described by the standard logistic equation[19]:

$$\frac{dN}{dt} = \tilde{r}N\left(1 - \frac{N}{K}\right) \quad (1)$$

Here, $\bar{r}$ is the overall fitness of the population, *i.e.,* the exponential growth rate (offspring per unit time) of the population when growth is uninhibited.

To simplify, we can consider the dynamics of the population density, $\rho = N/K$, which can take values between 0 and 1. Substituting this expression for ρ into Equation 1 gives

$$\frac{d\rho}{dt} = \tilde{r}\rho(1 - \rho) \quad (2)$$

Dynamics of individual sgRNAs.

Now, let us consider the individual sgRNAs. Let y_i represent the density of sgRNA i, so that

$$\rho = \sum_{i=1}^{n} y_i \quad (3)$$

Substituting this into Equation 2 gives

$$\sum_{i=1}^{n} \dot{y}_i = \tilde{r}\left(\sum_k y_k\right)\left(1 - \sum_j y_j\right) \tag{4}$$

where $\dot{y}_i$ is the derivative of y_i with respect to time. Next, we assume that $\bar{r}$ is equal to the mean fitness across all sgRNAs:

$$\tilde{r} = \frac{y_1}{\sum_k y_k} r_1 + \frac{y_2}{\sum_k y_k} r_2 + \cdots + \frac{y_n}{\sum_k y_k} r_n \tag{5a}$$

$$= x_1 r_1 + x_2 r_2 + \cdots + x_n r_n \tag{5b}$$

Here, r_i is the intrinsic fitness (uninhibited exponential growth rate) of sgRNA *i* and x_i is the fraction of the population made up by sgRNA *i*.

Next, we seek an equation to describe the growth of an individual sgRNA *i*. We substitute Equation 5a into Equation 4 and express the right-hand side as a sum over *i*:

$$\sum_{i=1}^{n} \dot{y}_i = \sum_i \frac{y_i r_i}{\sum_k y_k}\left(\sum_k y_k\right)\left(1 - \sum_j y_j\right)$$

$$= \left(\sum_i y_i r_i\right)\left(1 - \sum_j y_j\right)$$

$$= \left(\sum_i y_i r_i\right) - \left(\sum_i y_i r_i\right)\left(\sum_j y_j\right)$$

$$= \left(\sum_i y_i r_i\right) - \left(\sum_{i,j} y_i r_i y_j\right)$$

$$= \left(\sum_i y_i r_i\right) - \left(\sum_{j,i} y_j r_j y_i\right)$$

$$= \left(\sum_i y_i r_i\right) - \left(\sum_i y_i\right)\left(\sum_j y_j r_j\right) \tag{6}$$

$$= \sum_i \left(y_i r_i - y_i \sum_j y_j r_j \right)$$

$$= \sum_i \left[y_i \left(r_i - \sum_j y_j r_j \right) \right]$$

This relation shows that we obtain the correct overall population dynamics if we assume that each sgRNA's growth is governed by another logistic equation:

$$\dot{y}_i = y_i \left(r_i - \sum_j y_j r_j \right) \tag{7}$$

Equation 7 defines a system of equations that is equivalent to the competitive Lotka-Volterra equations (*18*). An interpretation of this equation is as follows: when the overall population size (across all sgRNAs) is small, the growth rate of sgRNA *i* is equal to r_i. When the overall population size grows, competition is governed by the term $y_i\, y_j\, r_j$; that is, interactions between sgRNAs *i* and *j* are deleterious to sgRNA *i* by a factor that's proportional to the fitness of sgRNA *j*.

To illustrate the dynamics resulting from Equation 7, we can numerically simulate the densities of three hypothetical sgRNAs, each starting at a density of 1/500, where drug is applied on days 0 through 8 and removed for days 8 through 16 (**Extended Fig. 2e**).

The intrinsic fitnesses of the sgRNAs while on drug are r_1 = 1.51, r_2 = 1.50, and r_3 = 1.49. When drug is removed, the intrinsic fitnesses of sgRNAs 1 and 2 remain unchanged, but the intrinsic fitness of sgRNA 3 decreases to $\tilde{r}$ = 1.20 (a change of -0.29).

Dynamics of sgRNA proportions

Now, we derive a system of equations for the relative sgRNA proportions. Recall that the relative proportion of sgRNA *i*, x_i, is given by

$$x_i = \frac{y_i}{\sum_k y_k} \tag{8}$$

Bearing in mind the relations for ρ and $\bar{r}$ in Equations 3 and 5, the relative proportion of the pool made up by sgRNA *i* over time is described by:

$$\dot{x}_i = \frac{d}{dt}\left[\frac{y_i}{\sum_k y_k}\right]$$

$$= \frac{d}{dt}\left[\frac{y_i}{\rho}\right]$$

$$= \frac{\dot{y}_i\rho - y_i\dot{\rho}}{\rho^2}$$

$$= \frac{1}{\rho}y_i\left(r_i - \sum_j y_j r_j\right) - \frac{1}{\rho}y_i\tilde{r}(1-\rho)$$

$$= \frac{1}{\rho}x_i\rho\left(r_i - \rho\sum_j x_j r_j\right) - \frac{1}{\rho}x_i\rho\tilde{r}(1-\rho)$$

$$= x_i\left(r_i - \rho\sum_j x_j r_j\right) - x_i\tilde{r}(1-\rho)$$

$$= x_i(r_i - \rho\tilde{r}) - x_i\tilde{r} + x_i\tilde{r}\rho$$

$$= x_i r_i - x_i\tilde{r}$$

$$= x_i\left(r_i - \sum_j x_j r_j\right) \quad (9)$$

Note that this equation has the same form as Equation 7, which describes the relative densities of the sgRNAs. The only difference is in the initial conditions: in the previous example, we started with $y_1(0) = y_2(0) = y_3(0) = 1/500$, implying that $x_1(0) = x_2(0) = x_3(0) = 1/3$. To verify that this is true, we can plot each sgRNA's density y_i divided by the total population density using Equation 7 (**Extended Fig. 2f**) and compare with the relative frequencies x_i given by Equation 9 (**Extended Fig. 2g**), noting their equality.

Calculating change in intrinsic fitness

The next task is to determine the change in a sgRNA's intrinsic fitness when off drug *vs.* on drug given the relative frequencies of the sgRNAs at (a) time 0 ("start"), (b) when drug is removed ("switch"), and (c) at the end of the experiment ("end"). Let us consider the logarithm of the relative proportions of sgRNAs *i* and *j*:

$$z_{ij} = log\left[\frac{x_i}{x_j}\right] \quad (10)$$

The change in z_{ij} over time takes a simple mathematical form:

$$\dot{z}_{ij} = \left(\frac{x_j}{x_i}\right)\left(\frac{\dot{x}_i x_j - x_i\dot{x}_j}{x_j^2}\right) \quad (11a)$$

(11b)

$$= \frac{x_i(r_i - \sum_k r_k x_k)x_j - x_i x_j\left(r_j - \sum_k r_k x_k\right)}{x_i x_j}$$

$$= r_i - r_j \tag{11c}$$

This means that the solution for z_{ij} is a straight line with slope $r_i - r_j$. For the sgRNAs in our previous example, the dynamics of z_{21} and z_{31} are shown in **Extended Fig. 2h.**

Given the sgRNA proportions at the start, switch, and end timepoints, it is straightforward to calculate the slopes of these lines. **Supplementary Table 7** gives the proportions, *z*-values, and slopes (σ_{it}) for the simulated sgRNAs relative to sgRNA 1.

Next, we can derive an expression for the change in fitness off-drug *vs.* on-drug for each sgRNA. Let us denote the on-drug fitness of sgRNA *i* by r_i and the off-drug fitness of sgRNA *i* by $\widetilde{r_i}$.The difference in on-drug *vs.* off-drug fitness is thus $\widetilde{r_i} - r_i$. Let us furthermore assume that the fitness of some sgRNA ξ (the reference sgRNA) does not depend on drug, *i.e.,* $r_\xi = \widetilde{r_\xi}$. Then, we seek

$$\begin{aligned} \Delta r_i &= r_i - \widetilde{r_i} \\ &= \left(r_i - r_\xi\right) - \left(\widetilde{r_i} - \widetilde{r_\xi}\right) \\ &= \sigma_{i,t=8} - \sigma_{i,t=16} \end{aligned} \tag{12}$$

For simulated sgRNA 2 relative to sgRNA 1, the slope at time 8 ($\sigma_{2,t=8}$) is equal to the slope at time 16 ($\sigma_{2,t=16}$), so $r_2 - \widetilde{r_2} = 0$; that is, sgRNA 2 is equally fit off-drug as on-drug, which agrees with the construction of the simulation.

For simulated sgRNA 3 relative to sgRNA 1, the slope at time 8 ($\sigma_{3,t=8}$) differs from the slope at time 16 ($\sigma_{3,t=16}$) by 0.29 units, *i.e.,* $r_3 - \widetilde{r_3} = 0.29$. Recall that the on-drug fitness of sgRNA 3 was 1.49 and the off-drug fitness was 1.20, so this also agrees with the construction of the simulation and indicates that simulated sgRNA 3 is addicted to drug.

In general, when $\Delta r_i < 0$, the sgRNA is more fit off-drug than on-drug (wild-type behavior); when $\Delta r_i = 0$, the sgRNA is equally fit off-drug and on-drug (pure resistance); and when $\Delta r_i > 0$, the sgRNA is more fit on-drug than off-drug (addiction). This satisfies the characteristics of an addiction score, so we define the addiction score *A* as

$$A = \Delta r_i \tag{13}$$

Recipe for calculating the addiction score

Taking all of this together, we can calculate change in on-drug *vs.* off-drug fitness if:

- We know the relative frequencies *x* of the sgRNAs at time 0 (t_0), when drug is removed (t_{rem}), and when the experiment ends (t_{end}).
- We assume that the overall population is undergoing logistic growth and standard ecological competition.
- We assume that one sgRNA (sgRNA ξ) is agnostic to drug (*i.e.,* has a "pure" resistance mutation).

With these conditions in place, the intrinsic fitness of sgRNA *i* on-drug *vs.* off-drug (the addiction score) is

$$A = r_i - \tilde{r_i}$$

$$= \sigma_{i,t_{rem}} - \sigma_{i,t_{end}}$$

$$= \frac{z(t_{rem}) - z(t_0)}{t_{rem} - t_0} - \frac{z(t_{end}) - z(t_{rem})}{t_{end} - t_{rem}}$$

$$= \frac{log\left(\frac{x_i(t_{rem})}{x_\xi(t_{rem})}\right) - log\left(\frac{x_i(t_0)}{x_\xi(t_0)}\right)}{t_{rem} - t_0} - \frac{log\left(\frac{x_i(t_{end})}{x_\xi(t_{end})}\right) - log\left(\frac{x_i(t_{rem})}{x_\xi(t_{rem})}\right)}{t_{end} - t_{rem}} \quad (14)$$

This recipe for calculating the addiction score is implemented in the **Supplementary Code** [https://github.com/skissler/EZH2]**.**

Simulating the logistic dynamics of the real sgRNAs.

To interpolate the relative sgRNA proportions over time given the measured proportions at the start, switch, and end times, we first calculated r_i and $\tilde{r_i}$ for each sgRNA *i*. We assigned the intrinsic fitness of the reference sgRNA to an arbitrary value (0 for the figures in the main plot, with sensitivity analysis between -10 and 10). In general, the choice of intrinsic fitness for the reference sgRNA may change how quickly the sgRNAs reach a quasi-steady state but does not affect the sgRNA proportions at the start, switch, and end times. In practice, we found very little difference in the dynamics for different assumptions of the intrinsic fitness of the reference sgRNA. We then used the lsoda function from the deSolve package in R version 4.1.3 to numerically solve the system of differential equations given by Equation 9 (see **Supplementary Code** [https://github.com/skissler/EZH2]).