Alternative drug sensitivity metrics improve preclinical cancer pharmacogenomics

To the Editor:

Measuring drug response in cancer cell lines is essential for studying mechanisms of drug action and identifying genetic variants associated with sensitivity and resistance (preclinical pharmacogenomics). The conventional approach involves computing relative viability based on the ratio of cells in control and drug-treated cultures at the end of a fixed time period (**Fig. 1a** and **Supplementary Fig. 1a**). We have recently shown that relative viability and the response metrics derived from it, such as IC_{50} (drug concentration resulting in 50% relative viability) and area under the dose–response curve (AUC), are confounded by variation in cell proliferation rates and assay duration[1], potentially explaining discrepancies among large-scale drug response data sets[2,3]. Here we show that use of relative viability as a measure of drug response results in false-positive and false-negative pharmacogenomic associations. An alternative approach, based on computing normalized growth rate (GR) inhibition minimizes these artifactual associations by scoring drug sensitivity on a per-division basis[1] (**Fig. 1a**). GR_{50}, a measure of potency, is the concentration of drug that reduces cell proliferation rate by one-half; GR_{max}, a measure of efficacy, is the maximum effect of a drug at the highest tested concentration; and GR_{AOC} combines these in an integrated 'area over the curve' value. The signs of GR values and GR_{max} relate directly to response phenotype: positive for partial growth inhibition, zero for complete cytostasis and negative for cell death. By contrast, E_{max} (relative viability at the highest tested concentration) values do not have this property as they are strongly confounded by proliferation rates[1,4]. Collecting GR values requires only modest changes in experimental approach[5,6] and calculations can be performed online (http://www.GRcalculator.org/).

A recently released data set from the Genentech Cell Line Screening Initiative (gCSI)[7] is unique among large-scale drug response data sets in reporting cell division times, which are required for GR calculations. Median cell line division times varied about twofold across tissue types and up to threefold within cell lines of a given type (note that division times also vary with culture conditions[1,2]). Lymphoid cancer lines had the shortest median division times at 29 h with an interquartile range (IQR) of 15 h, whereas breast cancer lines had the longest division times at 53 h (IQR = 16 h; **Fig. 1b**). All other things being equal, slower division rates increase IC_{50} and E_{max} values for simple arithmetic reasons (**Fig. 1a**)[1]. In the gCSI data set, both IC_{50} and E_{max} were in fact substantially higher for the slowest dividing cell types relative to the fastest (Spearman's $\rho > 0.61$, $P < 0.005$). For drugs such as paclitaxel (Abraxane, Taxol), docetaxel (Taxotere), and gemcitabine (Gemzar), drug sensitivity also correlated with cell division times ($\rho > 0.35$, $P < 5.6 \times 10^{-13}$ for IC_{50}; $\rho > 0.4$, $P < 1.4 \times 10^{-16}$ for E_{max}). On the basis of IC_{50} values, bone cancer cells were reportedly 60-fold more drug sensitive on average than breast cancer cells even though the drugs in the gCSI data set are primarily used to treat the latter disease. By contrast, we found that median GR_{50} and GR_{max} values are not significantly correlated across fast- and slow-growing tumor cell types ($P > 0.20$). Thus, the correlation between tumor type and median IC_{50} or E_{max} values arises from systematic variation in division time.

Conventional drug sensitivity metrics are similarly confounded by genetic mutations (natural and engineered) that alter proliferation rates[8,9] (**Supplementary Fig. 1b–d**), resulting in both bona fide and confounded pharmacogenomic associations. Overall, only 40% of 122 statistically significant associations (false-discovery rate (FDR) < 0.15) between IC_{50} values and genomic alterations (gene mutations, deletions, or amplifications) in gCSI data were significant when using GR_{50} values (**Fig. 1c**). Conversely, about half of associations scored as significant by GR_{50} were missed using IC_{50}. An example of a false-negative association is provided by *PTEN* loss-of-function (LOF) mutations, which are known to mediate lapatinib (Tykerb) resistance in breast cancer cells[10] but were not associated with differences in IC_{50} values ($P = 0.12$, rank-sum test) simply because *PTEN* mutant cell lines also had a median division time that was 22.5 h shorter than *PTEN* wild-type cell lines ($P = 0.01$), consistent with the known role of *PTEN* as a tumor suppressor. By contrast, the association between *PTEN* LOF mutations and lapatinib resistance was significant when GR_{50} values were used ($P = 0.003$; **Fig. 1d**). A false-positive example is provided by deletion of *CDC73* (a component of an RNA polymerase II complex[11]), which was associated in ovarian cancer lines with a 1,000-fold increase in median IC_{50} for docetaxel ($P = 0.002$; **Fig. 1e**). *CDC73*-deleted cells grew more slowly that *CDC73* expressers (57 h versus 35 h doubling time; $P = 0.03$), and when we corrected for this difference using GR_{50} values, the association between *CDC73* deletion and docetaxel sensitivity disappeared ($P = 0.13$). We conclude that differences in cell division rate underlie false-positive and false-negative associations between genomic features and conventional drug response metrics.

Efficacy is rarely explored as a metric of drug response, even though incomplete cytostasis and fractional cell killing result in persistence of viable cells[12] that likely contributes to incomplete tumor regression and residual disease[13]. Across the gCSI data set, both potency and efficacy exhibited substantial variation, and even highly potent drugs could be inefficient at killing (**Fig. 2a** and **Supplementary Fig. 2**). For example, the mTOR inhibitor rapamycin was potent ($GR_{50} < 0.1$ μM) in 34% of cell lines but ineffective in blocking growth in these lines ($GR_{max} = 0.34 \pm 0.23$, median and IQR). The converse was also true: the histone deacetylase (HDAC) inhibitor vorinostat (Zolinza) was highly cytotoxic in most lines ($GR_{max} = -0.52 \pm 0.48$)

but had poor potency (median GR_{50} = 1.53 μM). The microtubule inhibitor docetaxel was highly potent (median GR_{50} = 8 nM) but cytotoxic in only half of cell lines tested (GR_{max} = –0.04 ± 0.53). For doxorubicin (Adriamycin, Doxil), which is commonly used in hematopoietic cancers[14], GR_{max} values were significantly lower in blood cell lines than in cell lines from other tissues (median of –0.84 vs. –0.49, $P = 1.6 \times 10^{-6}$), but GR_{50} values were similar (P = 0.14).

We identified 99 statistically significant associations (FDR < 0.15) between genomic variants and GR_{max} values and the same number for GR_{50}, but only 18 of those variants were significantly associated with both GR_{max} and GR_{50} (**Fig. 2b**). For example, in most ovarian cell lines deleted for the anti-apoptotic *BCL2* gene, GR_{max} values for docetaxel were negative, which indicates cell death, whereas GR_{max} values in *BCL2* wild-type lines were positive, which indicates partial cytostasis (median of –0.18 versus 0.23; P = 0.007; **Fig. 2c**). In contrast, GR_{50} values were indistinguishable (P = 0.23). The sensitivity of *BCL2*-deleted ovarian cell lines to docetaxel, as measured by GR_{max}, is consistent with an observed increase in apoptosis when the BCL2 inhibitor ABT737 is combined with docetaxel[15].

To study the relationship between GR_{50} and GR_{max} in a systematic manner, we analyzed the CTRP (Cancer Therapeutic Response Portal) data set[16], which comprises 545 drugs but lacks data on cell-proliferation rates. We inferred division times for 121 cell lines using gemcitabine response as a fiducial (**Supplementary Methods**). In place of GR_{max}, we used an interpolated value $GR_{2\mu M}$ based on fitted dose–response curves. Shannon's entropy (H) was then used to measure the information content in distributions of GR_{50} and $GR_{2\mu M}$ values for each drug. We observed a wide range of entropies with some drugs varying primarily in GR_{50}, some in $GR_{2\mu M}$, and some in both (**Supplementary Fig. 3a**). The mutual information between GR_{50} and $GR_{2\mu M}$ was low, showing that the two metrics of drug response convey non-overlapping information. AUC is often used as a summary statistic for a drug, but we found that the entropy of GR_{AOC}, the GR equivalent of AUC, is lower than the joint entropy of GR_{50} and $GR_{2\mu M}$ (**Supplementary Fig. 3b,c**). This makes sense because the same GR_{AOC} value can be associated with a potent drug having poor efficacy or a less potent drug having good efficacy (**Supplementary Fig. 3d**).

We conclude that GR metrics have the potential to improve pre-clinical pharmacology in three ways. First, use of IC_{50} values based on relative viability in pharmacogenomic studies results in multiple false-positive and false-negative associations simply because cell division rates vary in a systematic manner with tumor type and genotype. Artifactual associations arise, regardless of experimental methodology or type of disease, whenever cells divide over the course of a drug-response assay and the phenotype of interest is directly or indirectly affected by growth. GR metrics correct not only for this but also for arbitrary differences in protocol that affect cell-division number (e.g., assay duration, media conditions, and plating density).

Second, because GR metrics reduce the impact of confounders, cross-study reproducibility is improved. For example, the

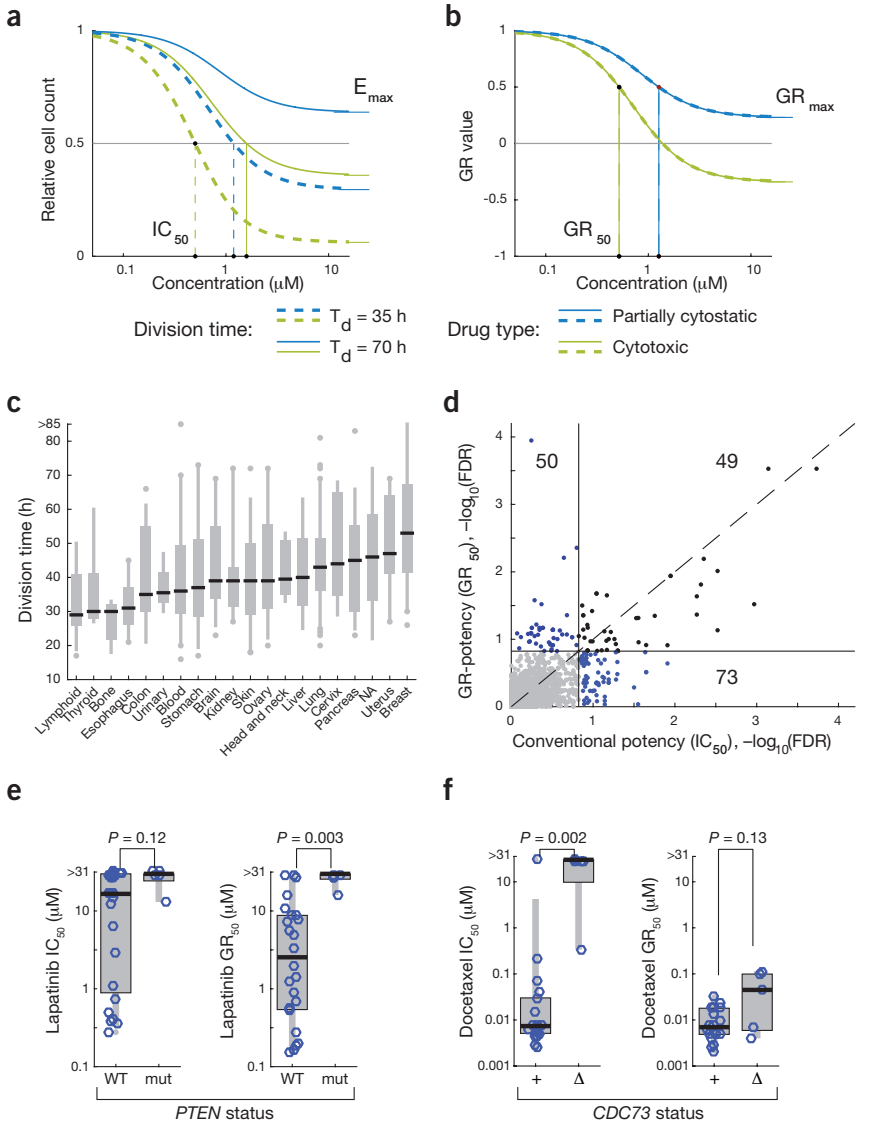

Figure 1 Division rate confounds IC_{50} values and pharmacogenomic associations. (**a**,**b**) Theoretical dose–response curves by relative cell count (**a**) and GR value (**b**) for cell lines with slow (T_d = 70 h; solid lines) or fast (T_d = 35 h; dashed lines) division times treated with a drug that partially inhibits division (blue) or causes cell death (green). Slow- and fast-growing lines are assumed to respond identically at a mechanistic level. IC_{50} and GR_{50} are marked on the *x*-axis, E_{max} and GR_{max} on the *y*-axis. Note that dashed and solid lines overlap in the GR value panel (**Supplementary Fig. 1a**). (**c**) Distribution of division time for cell lines by tissue of origin in the gCSI data set. (**d**) IC_{50} and GR_{50} enrichment by genomic alteration. Numbers represent significant (FDR < 0.15) associations for IC_{50} alone (blue dots), GR_{50} alone (blue), or IC_{50} plus GR_{50} (black). (**e**) Distribution of IC_{50} and GR_{50} values for lapatinib in breast lines based on *PTEN* status. (**f**) Distribution of IC_{50} and GR_{50} values for docetaxel in ovarian lines based on *CDC73* status. Rank-sum *P*-values are reported. GR_{50} and IC_{50} values are capped at 31 μM. n = 23 (for panel **f**).

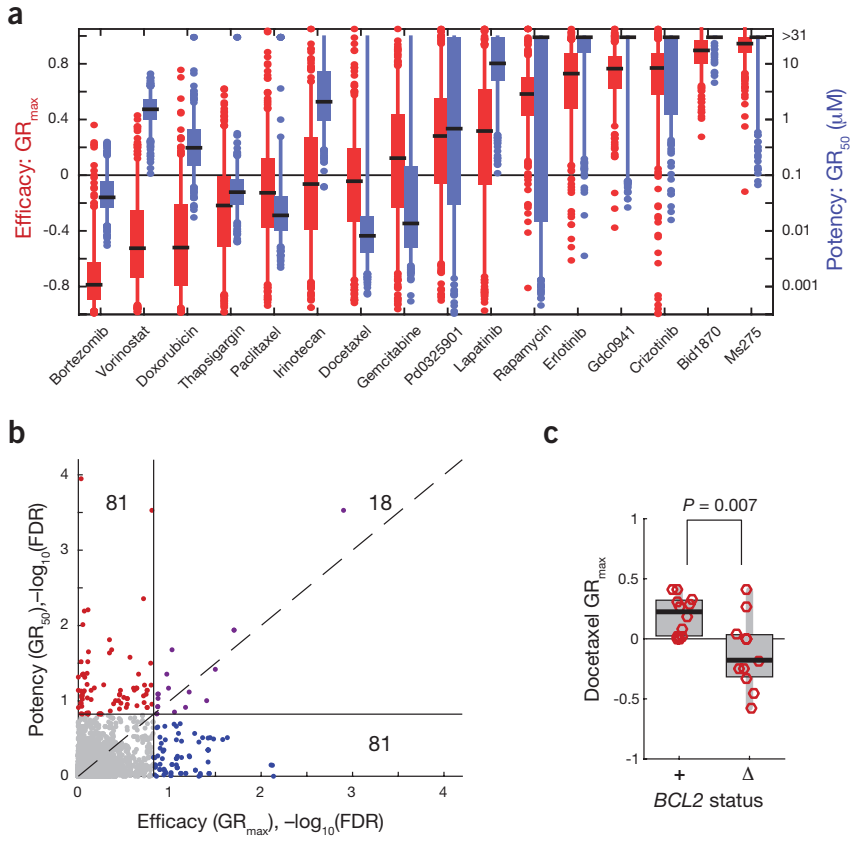

Figure 2 Variation in efficacy and genomic enrichment based on GR_{max} values in the gCSI data set. (**a**) Distribution of GR_{max} and GR_{50} values for each drug across all cell lines. GR_{50} values are capped at 31 μM. (**b**) GR_{50} and GR_{max} enrichment by genomic alteration for each drug and tissue. Numbers represent significant (FDR < 0.15) associations for GR_{max} alone (red dots), GR_{50} alone (blue), or GR_{max} plus GR_{50} (purple). (**c**) Distribution of GR_{max} values for docetaxel in ovarian cancer lines based on *BCL2* deletion status. Rank-sum *P*-value is reported. ($n = 23$).

correlation between gCSI and CTRP data sets is significantly increased using GR_{AOC} rather than AUC as a response metric ($P = 1.3 \times 10^{-3}$, Student's *t*-test; **Supplementary Fig. 4**).

And third, efficacy as measured by GR_{max} and potency as measured by GR_{50} differ at a biological level, carry complementary information (low mutual information), and are associated with largely non-overlapping genetic alterations. In principle, variation in potency and efficacy can be captured by integrating across dose–response curves (GR_{AOC}), but we find that information content is maximized if GR_{50} and GR_{max} are considered independently.

Because the ultimate purpose of antineoplastic drugs is to kill cancer cells[13], and high potency is no guarantee of good efficacy, we propose that the best drugs and most important pharmacogenomic associations are not those associated with low IC_{50} values, but rather those that result in the most negative GR value at clinically relevant drug concentrations (e.g., C_{max}). Relating *in vitro* measures of drug sensitivity to *in vivo* responses remains challenging[17], but for this to have any chance of success it is essential that *in vitro* data are as informative and reproducible as possible.

Editor's note: This article has been peer-reviewed.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
M.H., M.N., and P.K.S. conceived this study and wrote the paper. M.H. performed the analyses. This work was supported by NIH grant U54-HL127365. We thank J.W. Gray, R. Bourgon and colleagues at Genentech, and S. Schreiber for helpful discussions.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Marc Hafner, Mario Niepel & Peter K Sorger

HMS LINCS Center, Laboratory of Systems Pharmacology, Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA.
e-mail: peter_sorger@hms.harvard.edu

1. Hafner, M., Niepel, M., Chung, M. & Sorger, P.K. *Nat. Methods* **13**, 521–527 (2016).
2. Haibe-Kains, B. *et al. Nature* **504**, 389–393 (2013).
3. CCLE Consortium & GDSC Consortium. *Nature* **528**, 84–87 (2015).
4. Harris, L.A. *et al. Nat. Methods* **13**, 497–500 (2016).
5. Niepel, M., Hafner, M., Chung, M. & Sorger, P.K. *Curr. Protoc. Chem. Biol.* **9**, 55–74 (2017).
6. Hafner, M., Niepel, M., Subramanian, K. & Sorger, P.K. *Curr. Protoc. Chem. Biol.* **9**, 96–116 (2017).
7. Haverty, P.M. *et al. Nature* **533**, 333–337 (2016).
8. Im, H.K. *et al. PLoS Genet.* **8**, e1002525 (2012).
9. Choy, E. *et al. PLoS Genet.* **4**, e1000287 (2008).
10. Eichhorn, P.J. *et al. Cancer Res.* **68**, 9221–9230 (2008).
11. Rozenblatt-Rosen, O. *et al. Mol. Cell. Biol.* **25**, 612–620 (2005).
12. Fallahi-Sichani, M., Honarnejad, S., Heiser, L.M., Gray, J.W. & Sorger, P.K. *Nat. Chem. Biol.* **9**, 708–714 (2013).
13. Letai, A. *Clin. Cancer Res.* **21**, 5015–5020 (2015).
14. Fisher, R.I. *et al. N. Engl. J. Med.* **328**, 1002–1006 (1993).
15. Oakes, S.R. *et al. Proc. Natl. Acad. Sci. USA* **109**, 2766–2771 (2012).
16. Seashore-Ludlow, B. *et al. Cancer Discov.* **5**, 1210–1223 (2015).
17. Johnson, J.I. *et al. Br. J. Cancer* **84**, 1424–1431 (2001).

The illusion of control in germline-engineering policy

To the Editor:

The arrival and rapid adoption of the clustered, regularly interspaced, short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) system[1] has sparked ethical and societal controversy around genome editing of the human germline. Here, I point out the fallacy that such technologies and their applications can be globally prohibited on the basis of universal ethics and bans—the so-called 'illusion of control'. A look at previous technological developments suggests instead that differentiated and multi-faceted approaches that take into account the broadest range of possible ethical and social issues would be preferable for the oversight of CRISPR–Cas9 germline engineering. Such an approach would not only be more effective but also ensure that society has the greatest chance of capitalizing on potential opportunities of the technology.

To date, the CRISPR–Cas9 technique has been used to edit the genomes of (non-viable) one-celled human embryos[2,3]. The first of these genome-editing studies