Data variability in standardised cell culture experiments

Ian G. Reddin, Tim Fenton, Mark N. Wass*, Martin Michaelis*

School of Biosciences, University of Kent, Canterbury, UK

* Correspondence to: m.n.wass@kent.ac.uk (M.N.W.), m.michaelis@kent.ac.uk (M.M.)

Abstract

Despite much debate about a perceived 'reproducibility crisis' in the life sciences, it remains unclear what level of replicability is technically possible [1,2]. Here, we analysed the variation among drug response data of the NCI60 project, which for decades has tested anti-cancer agents in a 60-cell line panel following a standardised protocol [3]. In total, 2.8 million compound/cell line experiments are available in the NCI60 resource CellMiner [4]. The largest fold change between the lowest and highest GI50 (concentration that reduces cell viability by 50%) in a compound/cell line combination was 3.16×10^{10}. All compound/ cell line combinations with >100 experiments displayed maximum GI50 fold changes >5, 99.7% maximum fold changes >10, 87.3% maximum fold changes >100, and 70.5% maximum fold changes >1000. FDA-approved drugs and experimental agents displayed similar variation. The variability remained very high after removal of outliers and among experiments performed in the same month. Hence, our analysis shows that high variability is an intrinsic feature of experimentation in biological systems, even among highly standardised experiments in a world-leading research environment. Thus, a narrow focus on experiment standardisation does not ensure a high level of replicability on its own.

Introduction

In the life sciences, there is a crisis narrative and a perception of a lack of data reproducibility ("reproducibility crisis" or "replication crisis") [2,5-10]. However, the actual scale of the crisis remains unclear and evidence is largely anecdotal [2]. Much of the data are based on researcher views expressed in survey responses [11-13] or provided as Comments or Correspondence without providing detailed information [14,15].

One project that investigates the replicability of research data is the 'Reproducibility Project: Cancer Biology', which independently repeats influential preclinical studies (https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology). So far, 17 replication studies have been completed. Five studies reported the successful reproduction of the original studies [16-20], while eight reported a mixed outcome [21-28], and four failed to reproduce the original findings [29-32]. It is not clear whether these data are representative. It is a small dataset focused on small, early, and highly cited studies, which are more likely to overestimate effects [33].

There is also a lack of agreement on the expected level of data replicability [1]. In a dispute about the consistency of two large pharmacogenomic screens in cancer cell line panels, the Genomics of Drug Sensitivity in Cancer database and the Cancer Cell Line Encyclopedia [34-37], four analyses by different groups concluded a reasonable level of consistency [37-40], while six studies, all by the same group, disagreed [36,41-45].

To develop a realistic understanding of the replicability of standardised assays in a world-leading research environment, we investigated the variation in drug response data from the NCI60 screen [3]. Since 1985, the NCI60 screen has tested

the anti-cancer activity of thousands of compounds multiple times in a 60-cell line panel following strict standard operating procedures (Figure 1A) [3,4,47-51]. Thus, the NCI60 database provides an unprecedented wealth of data on the replicability of findings using highly standardised procedures by highly skilled experts.

Results

NCI60 drug response data are characterised by a high level of variability

All drug sensitivity data derived from NCI60 testing are made available via Cell Miner [4,47,48,50,51]. In total, 52,585 compounds were tested in the NCI60 resulting in 2.8 x 10^6 compound/cell line combinations. Two or more (up to 2,286) experiments were carried out for 11,841 compounds and 594,450 compound/cell line combinations (Figure 1B, Extended Data Table 1; Extended Data Table 2). More than 100 experiments in at least one cell line were performed for 18 compounds and more than 1,000 experiments for two compounds (Extended Data Table 3). Concentration ranges varied from $10^{1.2}$ to $10^{12.1}$. 612 compounds were screened with multiple concentration ranges, and the most common concentration range was 10^4 (11,213/ 94.7% of the compounds), representing the standard testing range using five 10-fold dilution steps (Extended Data Table 4).

The maximum fold change between the lowest and highest GI50 concentration (reduces cell viability by 50%) was detected for cyanomorpholinodoxorubicin in the colorectal cancer cell line COLO 205 (3.16 x 10^{10}) (Extended Data Figure 1A, Extended Data Table 5). 232,315 (39.1%) drug/cell line combinations displayed maximum fold changes >2, 108,247 (18.2%) drug/cell line combinations fold changes >5, 59,638 (10%) drug/cell line combinations fold changes >10, 19,089 (3.2%) drug/cell line combinations >100, and 8320 (1.4%) drug/cell line combinations >1,000 (Extended Data Table 5).

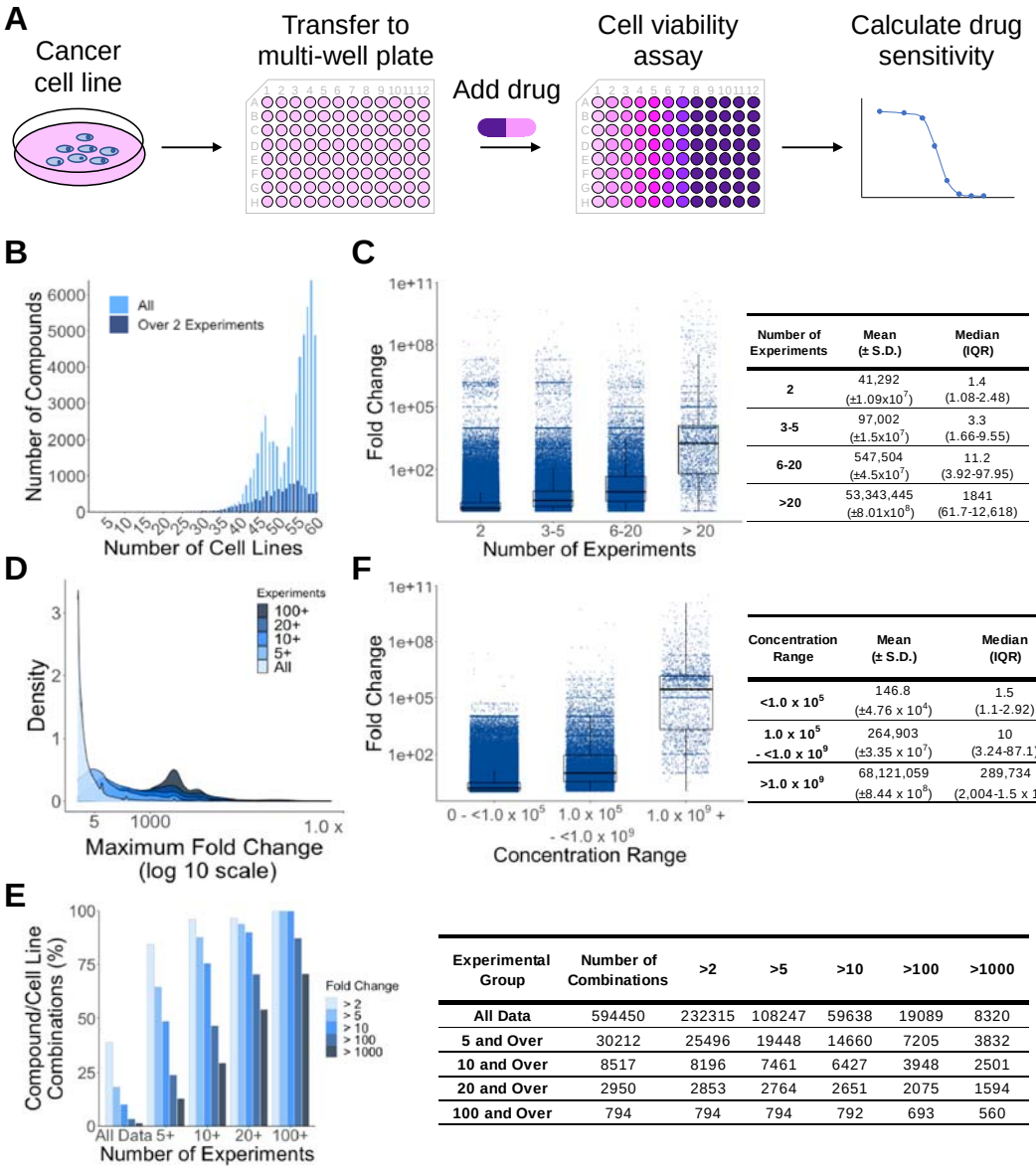
Variability increases with the number of experiments

The percentage of compound/cell line combinations with high maximum fold change strongly increased with the number of experiments (Figure 1C, Extended

Data Figure 2A, Extended Data Figure 2B, Extended Data Table 6). The mean and median GI50 fold changes increased from 41,292 and 1.4 for compound/cell line combinations with two experiments to 53,343,445 and 1841 for compound/cell line combinations with >20 experiments (Figure 1C, Figure 1D, Extended Data Table 6).

When we considered compound/cell line combinations with a minimum of five experiments, 25,496 (84.4%) of 30,212 compound/cell line combinations displayed maximum fold changes >2 and 3,832 (12.7%) compound/cell line combinations >1000. For compound/ cell line combinations with >100 experiments, 100% of 794 compound/ cell line combinations displayed a maximum fold change > 5 and 70.5% (560 out of 794) displayed a maximum fold change >1000 (Figure 1E, Extended Data Table 7).

Taken together, maximum GI50 fold changes increase with the number of experiments. In agreement, a significant correlation was detected between maximum GI50 fold changes and the number of experiments per compound/cell line combination (Spearman correlation coefficient = 0.34, $p < 2.2 \times 10^{-16}$) (Extended Data Figure 3A).



Number of Experiments	Mean (± S.D.)	Median (IQR)
2	41,292 ($\pm 1.09 \times 10^7$)	1.4 (1.08-2.48)
3-5	97,002 ($\pm 1.5 \times 10^7$)	3.3 (1.66-9.55)
6-20	547,504 ($\pm 4.5 \times 10^7$)	11.2 (3.92-97.95)
>20	53,343,445 ($\pm 8.01 \times 10^8$)	1841 (61.7-12,618)

Concentration Range	Mean (± S.D.)	Median (IQR)
$<1.0 \times 10^5$	146.8 ($\pm 4.76 \times 10^4$)	1.5 (1.1-2.92)
1.0×10^5 - $<1.0 \times 10^9$	264,903 ($\pm 3.35 \times 10^7$)	10 (3.24-87.1)
$>1.0 \times 10^9$	68,121,059 ($\pm 8.44 \times 10^8$)	289,734 ($2,004-1.5 \times 10^6$)

Experimental Group	Number of Combinations	>2	>5	>10	>100	>1000
All Data	594450	232315	108247	59638	19089	8320
5 and Over	30212	25496	19448	14660	7205	3832
10 and Over	8517	8196	7461	6427	3948	2501
20 and Over	2950	2853	2764	2651	2075	1594
100 and Over	794	794	794	792	693	560

Figure 1 Variability in NCI60 GI50 data. A) Overview of the principle of the NCI 60 screen. **B)** Compound/ cell line combinations with two or more experiments in the NCI60 database. **C)** GI50 fold changes in dependence on the number of experiments per compound/ cell line combination. Numerical data are presented in the adjacent table. **D)** Distribution of maximum GI50 fold changes illustrated by density plots for experimental compound/cell line combination groups with an increasing minimum number of experiments. **E)** Percentage of compound cell line combinations with maximum fold changes above the indicated thresholds in dependence of the number of experiments. Numerical data are presented in the adjacent table. **F)** Distribution of GI50 fold changes in dependence of the

concentration ranges in which compounds were tested. Numerical data are presented in the adjacent table.

Variability increases with the concentration range covered

The observed fold changes also reflected the tested concentration ranges per compound/ cell line combination in addition to the number of experiments, i.e. the broader the range of concentrations that were tested, the larger was the maximum fold change (Figure 1D, Extended Data Table 8). A positive correlation was observed between concentration range and maximum fold change for all compound data (Spearman correlation coefficient = 0.31, $p < 2.2 \times 10^{-16}$) (Extended Data Figure 3B).

The mean and median GI50 fold changes for compound/ cell line combinations for which a maximum concentration range $<1.0 \times 10^5$ was covered were 146.8 and 1.5, which increased to 68,121,059 and 289, 734 for those with a concentration range of $\geq 1.0 \times 10^9$ (Figure 1F, Extended Data Table 8).

Variability in FDA Approved Drugs

Since reliable clinical therapy outcomes depend on reproducible drug effects, it may be speculated that FDA-approved drugs are more robust in their drug response data than experimental agents. However, the drug response data observed for FDA-approved drugs displayed a similar variability like that observed across all tested compounds.

The NCI60 database contained data on 181 FDA-approved drugs, which had been tested at least twice, resulting in 399,686 experiments investigating 9,970 individual drug/cell line combinations (Extended Data Table 1). The number of experiments for drug/cell line combinations ranged from 2 to 2,286.

The maximum GI50 fold change was 1.25 x 10^{10} observed for mithramycin in four cell lines, the colorectal cancer cell line COLO-205 (26 experiments) (Extended Data Figure 1B), the CNS cell lines SF-295 (24 experiments) and U251 (27 experiments), and the ovarian cancer cell line IGROV1 (26 experiments). Mithramycin, a member of the aureolic acid family was approved in 1970 but only temporarily used for testicular carcinoma and other types of cancer due to serious side effects [52]. The second highest GI50 fold change (7.28 x 10^{7}) was detected for paclitaxel, a stabilising tubulin-binding agent and one of the most commonly used anti-cancer drugs [53], in MDA-MB-435 (Extended Data Figure 1C), which had originally been assumed to be a breast cancer cell line, but was later found to be derived from the melanoma cell line M14 [54].

The maximum GI50 fold changes were higher among the FDA approved drugs than for the non-FDA approved compounds (Figure 2A, Extended Data Table 9), probably because they were tested in more experiments and at bigger concentration ranges (Figure 2A).

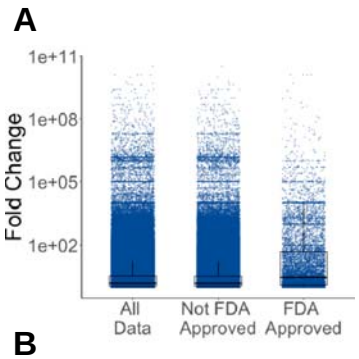
When we considered the percentage of FDA-approved drug/ cell line combinations with maximum GI50 fold changes >2, >5, >10, >100, and >1000 for combinations with >5, >10, >20, and >100 experiments (Figure 2B, Figure 2C, Extended Data Table 7, Extended Data Table 10), we obtained similar results to those across all compounds (Figure 1E).

In agreement with the findings across all compound/ cell line combinations, the maximum GI50 fold changes also increased with experiment number when the FDA approved drug/ cell line combinations were grouped into combinations with two experiments, 3 to 5 experiments, 6 to 20 experiments, and >20 experiments (Figure 2D, Suppl. Table 11), and the maximum GI50 fold change was also correlated with

the number of experiments performed (Spearman's correlation coefficient = 0.72, $p < 2.2 \times 10^{-16}$)(Extended Data Figure 3C).

Moreover, the maximum GI50 fold change increased with the concentration range covered (Figure 2E, Extended Data Table 12), and there was a significant correlation between the concentration range and the maximum GI50 fold change (Spearman's correlation coefficient = 0.62, $p < 2.2 \times 10^{-16}$) (Extended Data Figure 3D).

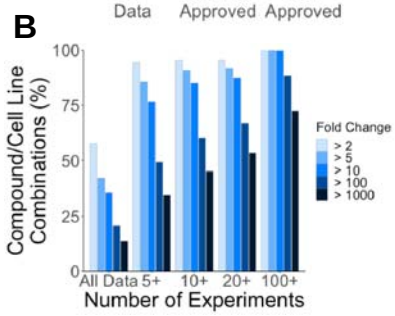
Taken together, there is no indication that FDA-approved drugs would display less variability than experimental compounds.



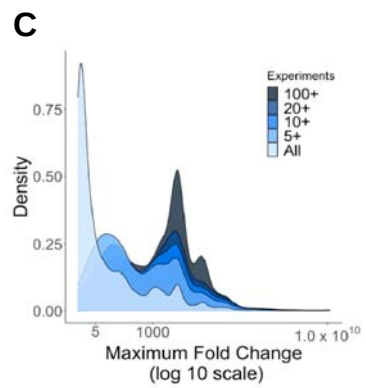
FDA Status	Mean (± S.D.)	Median (IQR)
All Data	318,410 (±5.71x10^7)	1.57 (1.12-3.36)
Not FDA Approved	218,154 (4.68±x10^7)	1.56 (1.12-3.3)
FDA Approved	6,195,807 (±2.57x10^8)	2.91 (1.27-50)

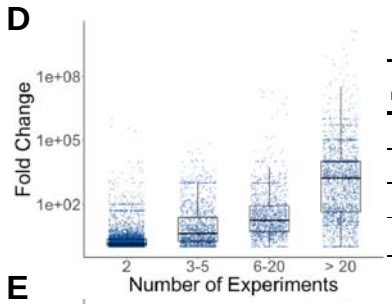
	Percentage of Total Combinations		
Number of Experiments	All Data	Not FDA	FDA
2	81.57	82.11	49.89
3-5	15.04	15.04	12.56
6-20	2.91	2.74	15.82
>20	0.48	0.11	21.74

	Percentage of Total Combinations		
Concentration Range	All Data	Not FDA	FDA
<1.0 x 10^5	94.35	94.56	81.62
1.0 x 10^5 ß- <1.0 x 10^9	5.21	5.03	15.55
>1.0 x 10^9	0.45	0.41	2.83

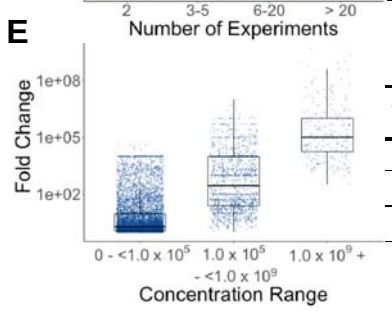


Experimental Group	Number of Combinations	>2	>5	>10	>100	>1000
All Data	9971	5752	4213	3547	2065	1361
5 and Over	3745	3533	3206	2868	1849	1295
10 and Over	2718	2594	2470	2312	1633	1225
20 and Over	2187	2089	2008	1911	1461	1168
100 and Over	751	750	750	749	663	544





Number of Experiments	Mean (± S.D.)	Median (IQR)
2	433.37 (±16,619)	1.39 (1.09-2.26)
3-5	170.52 (±1,184.1)	3.13 (1.6-10.7)
6-20	645 (±5,369.62)	12.65 (4.51-72.11)
>20	2.23x10^7 (±4.87x10^8)	465.59 (23.2-10,000)



Concentration Range	Mean (± S.D.)	Median (IQR)
<1.0 x 10^5	378.62 (±1922.17)	1.94 (1.18-9.12)
1.0 x 10^5 - <1.0 x 10^9	44,413 (±321,105)	284.45 (24.45-9,418)
>1.0 x 10^9	2.19 x 10^8 (±1.51 x 10^8)	100,000 (17,002-1.0 x 10^6)

Figure 2. GI50 variation for FDA-approved drugs. A) Compound/ cell line combinations with 2 or more experiments in the NCI60 database. Numerical data are presented in the adjacent tables. **B)** Percentage of FDA-approved drug/ cell line combinations with maximum fold changes above the indicated thresholds in dependence of the number of experiments. Numerical data are presented in the adjacent table. **C)** Distribution of maximum GI50 fold changes illustrated by density plots for experimental compound/cell line combination groups with increasing minimum numbers of experiments. **D)** GI50 fold changes in dependence on the number of experiments per compound/ cell line combination. Numerical data are presented in the adjacent table. (E) Distribution of GI50 fold

178 changes in dependence of the concentration ranges in which compounds were tested. Numerical data
179 are presented in the adjacent table.
180
181
182 **GI50 variability in experiments performed by month**
183 The reproducibility of results may be affected by parameters such as changes
184 in the reagents, e.g. use of different lots or batches, different experimenters, and
185 using cell lines at different passages [2,55-57]. Hence, experiments performed
186 closely together may be expected to display greater similarity than experiments
187 performed at more distant points in time during the decades of anti-cancer
188 compound testing by the NCI60.
189 To investigate the effects of the time of testing on data variability, we
190 compared experiments performed in the same month to control samples of the same
191 size that were randomly selected across the whole testing period. For this analysis,
192 we used the 18 FDA-approved drugs that were tested at least 100 times in at least
193 one cell line over periods of 95 to 275 months (Figure 3A, Extended Data Table 13),
194 resulting in 51,872 drug/cell line combinations and in total 321,709 experiments
195 (Figure 3B, Suppl. Table 14).
196 For every set of experiments performed on the same date, we generated
197 1,000 random control samples of the same size and compared the value distribution.
198 The variability of GI50 fold changes for same date experiments was indeed lower
199 than for random control samples, but remained very high reaching up to 1.74×10^8
200 (Figure 3C, Extended Data Table 15). Moreover, for 45% of the same date drug/ cell
201 line combinations the GI50 fold change was higher than the mean fold change of the
202 corresponding 1,000 random samples (Figure 3D, Extended Data Table 16).

When we looked at the data per cell line, the same date GI50 fold changes were higher than the mean random sample fold changes for the majority of drugs in ten cell lines, higher for half of the drugs in three cell lines, and lower for the majority drugs in the remaining 47 cell lines (Figure 3E, Extended Data Table 17). When we looked at the individual drugs, six displayed a majority of drug/cell line combinations with higher mean same date GI50 fold changes higher than in the random samples and twelve drugs displayed lower ones (Figure 3F, Extended Data Table 18).

Taken together, experiments performed in close timely proximity display lower variability than experiments performed over a longer time period, but the same date variability remains very high.

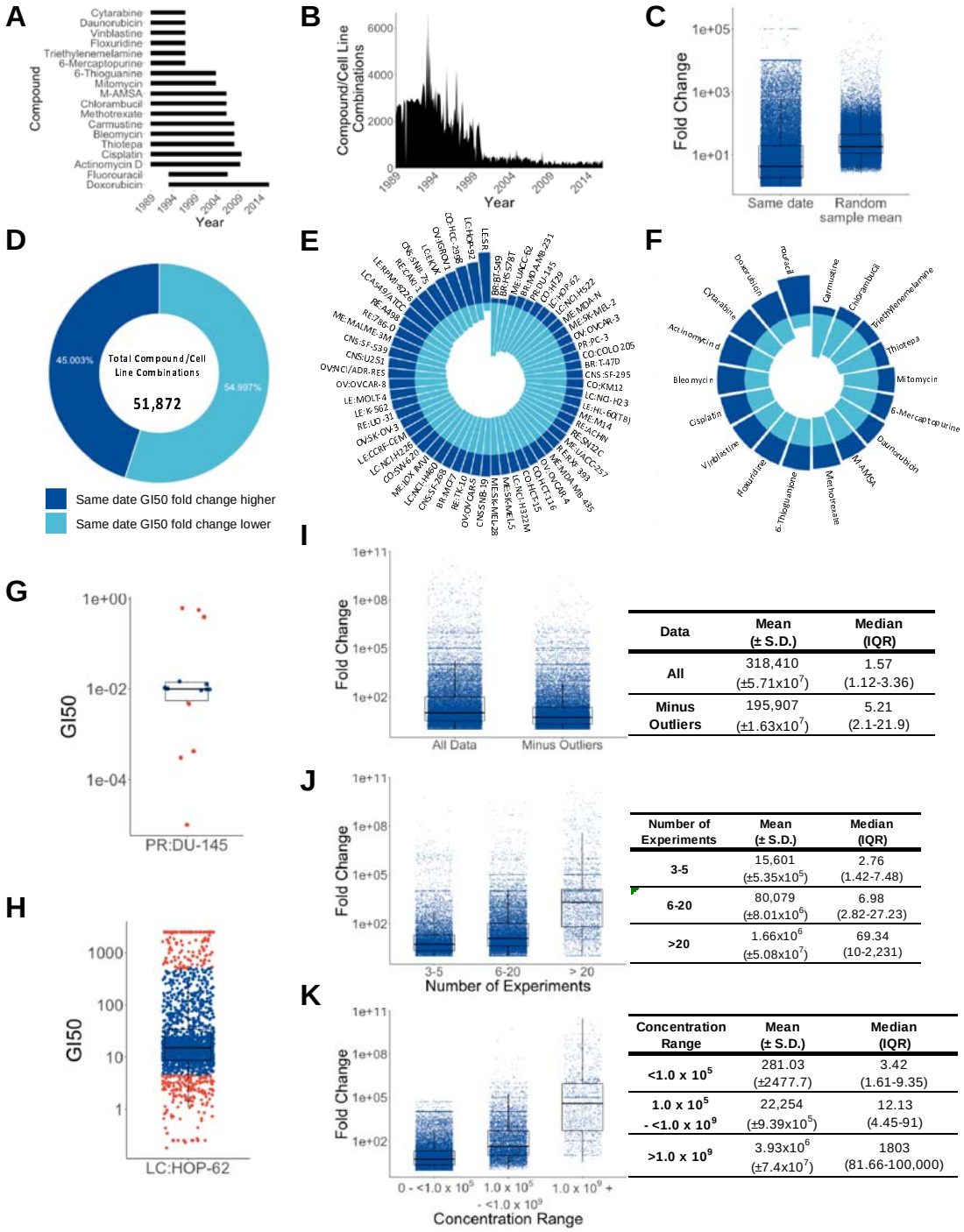


Figure 3 GI50 variability is high between compound/cell line combination experiments on the same date and is not caused by outliers. A) Time periods of drug testing for individual drugs. **B)** Testing of individual compound/cell line combinations by date. **C)** Maximum GI50 fold changes in experiments testing compound/ cell line combinations on the same date compared to maximum GI50

fold changes in 1000 random controls of the same sample size. **D)** Percentage of cases in which same date experiments had a higher fold change than control samples randomly picked across the timeline. **E)** Proportion of same date GI50 fold changes in compound/ cell line combinations that are higher or lower than random control samples per cell line. **F)** Proportion of same date GI50 fold changes in compound/ cell line combinations that are higher or lower than random control samples per drug. **G)** GI50 value distribution for maytansine in the prostate cancer cell line DU-145 (outliers indicated in red). **H)** GI50 value distribution for for **5-**fluorouracil in the lung cancer cell line HOP-62 (outliers indicated in red). **I)** Comparison of maximum GI50 fold changes before and after removal of outliers. Numerical values are presented in the adjacent table. **J)** Maximum GI50 fold changes increase with experiment number after removal of outliers. Numerical values are presented in the adjacent table. **K)** Maximum GI50 fold changes increase with the concentration range covered after removal of outliers. Numerical values are presented in the adjacent table.

High GI50 variability is not caused by outliers

Finally, we determined GI50 outliers for compound/cell line combinations with 5 or more experiments (738 compounds, 30,212 compound/cell line combinations, 598,243 GI50 values) using the adjusted boxplot method [58]. 5.7% (34,216) of GI50 values were outliers and 43.7% (13,208/30,212) of compound/cell line combinations had at least one GI50 outlier (Extended Data Table 19). This indicates that outliers are not responsible for the observed variability of GI50 in the majority (56.3%) of experiments.

The highest percentage of outliers was 50% (7/14 experiments for maytansine in DU-145 prostate cancer cells) (Figure 3G, Extended Data Table 19). The greatest number of outliers was 291 (16.8%) out of 1731 experiments for 5-fluorouracil in HOP-62 lung cancer cells (Figure 3H, Extended Data Table 19). Outlier number increased with the number of experiments for a compound/cell line combination with a Spearman correlation coefficient of 0.25 ($p < 2.2 \times 10^{-16}$) (Extended Data Figure 4).

The removal of outliers reduced data variability, but the overall variability remained very high with a maximum GI50 fold range of 2.5×10^9 detected for maytansine in the ovarian cancer cell line OVCAR-5 over 35 experiments (Figure 3I, Extended Data Table 19).

As detected in the analysis across all experiments, maximum GI50 fold changes increased with the number of experiments and the concentration ranges covered also after the removal of outliers (Figure 3J, Figure 3K, Extended Data Table 19, Extended Data Table 20). A significant correlation was observed between experiment number and maximum GI50 fold change with a Spearman correlation of 0.39 ($p < 2.2 \times 10^{-16}$) (Extended Data Figure 5A) and between concentration range and maximum GI50 fold change with a Spearman correlation of 0.47 ($p < 2.2 \times 10^{-16}$) (Extended Data Figure 5B).

No drift in drug sensitivity over time

Cancer cell lines may display substantial changes in genotype and phenotype over time [55,59]. Hence, part of the variability observed in drug sensitivity may be the consequence of a shift in drug response over time. To investigate this, we established timelines of the GI50 values for the 18 compounds, which had been tested at least 100 times in one or more cell lines. This resulted in time lines for 1080 compound/cell line combinations with time frames ranging from 95 months (vinblastine, floxuridine, cytarabine, daunorubicin, 6-mercaptopurine) to 275 months (Doxorubicin) (Figure 3A, Extended Data Table 13).

Drug/ cell line combinations, in which the fold change between the mean GI50 on the first experimental date and the mean GI50 on the last experimental date was 50% or greater than the maximum GI50 fold change for the data points in between

were considered as candidates for a drift in drug sensitivity. Only six (0.56%) out of 1080 drug/cell line combinations fulfilled these criteria (Figure 4, Extended Data Table 21).

The distribution of the individual GI50 values for three of the drug/ cell line combinations (floxuridine/ SK-OV-3, methotrexate/ BT-549, 6-mercaptopurine/ BT-549) did not indicate a GI50 shift over time (Figure 4A-C, Extended Data Table 21, Extended Data Table 22). For the other three drug/ cell line combinations (vinblastine/ T-47D, bleomycin/ K-562, M-AMSA/ MDA-MB-435) a drift in sensitivity appears unlikely but cannot be excluded based on the data (Figure 4D-E, Extended Data Table 21, Extended Data Table 22). However, such observations are very rare. Moreover, a phenotypic drift in a cell line would be expected to result in changes in sensitivity to more than one drug over time. Hence, the data provide no evidence suggesting that the drug sensitivity of individual cell lines may have changed over time. These findings may also reflect that the NCI60 uses cell lines within a window of 30 passages [60].

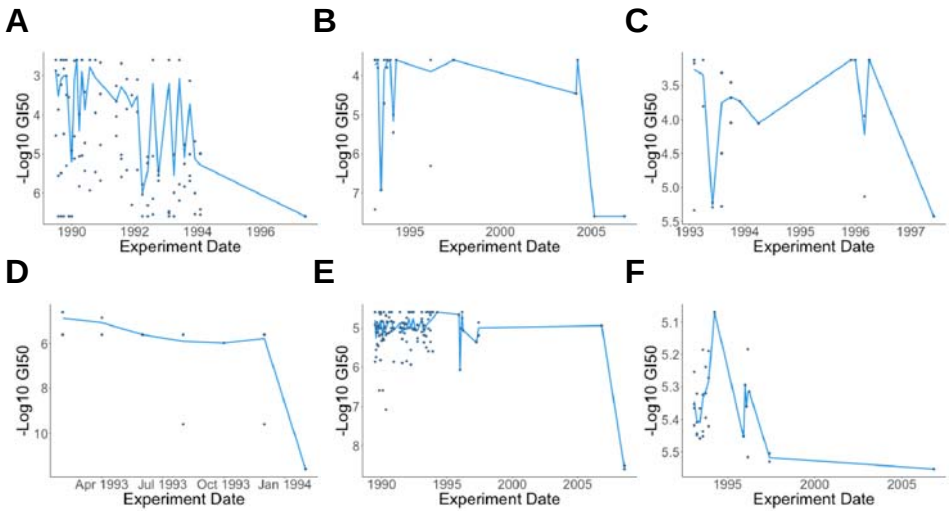


Figure 4 Experimental time lines for individual compound/cell line combinations. The six experimental timelines for compound/cell line combinations (with more than 100 experiments performed in at least one cell line) with a fold change between the first and last mean GI50 that is 50% greater than the maximum GI50 for the remaining experiments. **A)** floxuridine in the ovarian cancer cell line SK-OV-3, **B)** methotrexate in the breast cancer cell line BT-549, **C)** 6-mercaptopurine in the breast cancer cell line BT-549, **D)** bleomycin in the leukaemia cell line K-562, **E)** vinblastine in the breast cancer cell line T-47D, **F)** M-AMSA in the melanoma cell line MDA-MB-435. Plot lines represent mean GI50s while data points represent individual experiments at an experimental date.

Discussion

To gain real-life insights into the extent of data variability in a standardised research environment, we here investigated the variation of GI50 values indicating drug sensitivity for compounds that had been tested multiple times in the NCI60 panel. The variation was large with the highest fold change between the lowest and highest GI50 in a given compound/cell line combination being 3.16×10^{10}. As might have been expected, the fold change between the lowest and the highest GI50 in a specific compound cell line combination increased with the number of experiments and the concentration range tested.

CellMiner contains data on experimental compounds as well as on FDA-approved drugs that are in clinical use [4,46-51]. Although FDA-approved drugs might have been expected to result in more robust data, this was not the case and they displayed a similar data variability as that determined across all compounds. The variability also remained very high when we only considered experiments that were performed in the same months or removed outliers.

Given that the NCI60 uses highly standardised methods, a high variability of assay results appears to be an intrinsic feature of lab experiments. The large range of GI50 values observed for anti-cancer drugs is of potential relevance, given that cytotoxic anti-cancer drugs are typically used at maximum tolerated doses that cannot be further increased without unacceptable toxicity [61-63]. Moreover, the maximum effects of targeted drugs, e.g. antibodies or kinase inhibitors that interfere with cancer-specific structures or entities, do not further increase beyond the ‘optimal biological dose’, i.e. the dose at which the biological target is completely inhibited [61-64]. Hence, even a two-fold difference in the GI50, which occurred in 25,496 (84.4%) of 30,212 compound/cell line combinations with at least five experiments, is

of potential relevance, as a two-fold increase of the clinical dose of an anti-cancer drug is rarely feasible .

Notably, there is awareness of this variability within the NCI60 project as indicated by strict quality control procedures in the presentation of NCI60 GI50 values in CellMiner, which result in the exclusion of up to 96% of experiments (48 out of 50) for a given compound/ cell line combination (Extended Data Table 23) [4].

However, such knowledge has not penetrated into scientific discourse. Many authors promote strict standardisation of experimental procedures as a strategy to improve data quality and reliability [2,5,15,45,65-71].

Our analysis of NCI60 data, which is of unprecedented depth and provided by a world-leading institution with unprecedented transparency, indicates that data variation remains very high even under ideal conditions that the vast majority of research groups will not be able to afford. Hence, our data suggest that experiment heterogenization, the testing of a hypothesis in many different (experimental) systems and datasets and different laboratories [72-79], is a much better strategy to generate robust and meaningful data.

In conclusion, our analysis demonstrates that the variation of experimental data is extremely high even under optimal conditions in a world-leading environment applying the highest standards. This shows that increased standardisation is not a straightforward way to resolve issues associated with limited replicability. Hence, increased data robustness will have to include additional strategies such as independent replication and experiment heterogenization, i.e. multiple testing of the same hypothesis using different approaches and models. Awareness of the inherent variability of experimental results, will help researchers to develop a realistic

346 understanding of the meaning of their data and to design more diverse research

347 strategies that will result in higher data robustness and reliability.

348
349

Methods

Data Availability

All data were obtained from CellMiner [80] Version 2.2. Dose concentration range data (June 2018 release) were obtained from the National Cancer Institute DTP NCI bulk data for download pages (https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-60+Growth+Inhibition+Data). Of the 52,585 NCI codes given to compounds tested on the NCI-60 cell line panel, 42,794 were given to unnamed compounds and 9,791 were given to 9,027 named compounds. 262 named compounds received two or more individual codes. Some GI50 values represent minimum or maximum drug concentrations where the actual GI50 was not reached [4]. Since such values understate the actual data variation, we did not remove these data. All data generated during this study are included in this published article and its supplementary information files.

Maximum GI50 fold change calculation

The drug sensitivity data was converted from –log10 GI50, to the GI50 (μM) for all compound, all cell lines and all experiments. Maximum fold changes were calculated for each compound/cell line combination with more than one experiment (594,450) by dividing the maximum GI50 for a cell line by the minimum GI50.

Number of experiments and experimental groups

The number of experiments for each individual compound/cell line combination was calculated by counting all experiments performed on the same experimental date as well as experiments on different dates. The relationship

between number of experiments and maximum fold change was investigated by using Spearman's correlation coefficient as the distribution of maximum GI50 fold change was not normal.

The compound/cell line combinations were then assigned experimental groupings base on the number of experiments performed: all data, 5 or more experiments, 10 or more experiments, 20 or more experiments, and 100 or more experiments. This allowed comparison of "high" maximum fold changes (>2, >5, >10, >100, and >1000) for combinations with varied number of experiments. Additionally, compound/cell line combinations were assigned to experimental groups: 2 experiments, 3 to 5 experiments, 6 to 20 experiments and over 20 experiments. These experimental groupings enabled comparison of GI50 fold change statistics (mean, median, minimum, maximum, variance) for compound/cell line combinations with number of experiments ranging from lower to higher.

Concentration range and experimental groups

Maximum dose concentration range for a compound/cell line combination was determined by using the minimum and maximum dose concentration used in an experiment for an individual compound on an individual cell line. The minimum concentration range was $1.0 \times 10^{1.2}$ and the maximum concentration was $1.0 \times 10^{12.1}$. The relationship between dose concentration range and maximum fold change was investigated by using Spearman's correlation coefficient as the distribution of maximum GI50 fold change was not normal.

Compound/cell line combinations were assigned to groups based on the dose concentration range for that combinations: maximum concentration range less than 1.0×10^{5}, maximum concentration range 1.0×10^{5} to 1.0×10^{9} exclusive and

maximum concentration range 1.0 x 10^9 and above. These experimental groupings enabled comparison of GI50 fold change statistics (mean, median, minimum, maximum, variance) for compound/cell line combinations between lower and higher concentration ranges.

FDA-approved compound analysis

All compounds that were classed as FDA-approved drugs by the NCI-60 in CellMiner Database Version 2.2 and where two or more experiments had been performed were extracted from the complete dataset. This created an FDA-approved dataset of 181 drugs for which 399,686 experiments for 9,970 individual drug/cell line combinations were performed. Analysis of relationship between the number of experiments/concentration ranges and maximum GI50 fold change for drug/cell line combinations was performed as for the complete dataset, described above.

Experiments on the same date

Month and year of each experiment was available so experimental timelines were established for compounds by calculating the time between the first and last experiment date. Multiple experiments were carried out on the same date for many of the compound/cell line combinations, particularly the 18 compounds with at least one cell line with 100 total experiments. The data for these 18 compounds, 17 of which were FDA-approved, was extracted from the complete dataset to create a subset of data deemed suitable to compare GI50 variability on the same date with GI50 variability over an experimental timeline.

The maximum GI50 fold change on each date where there were multiple experiments for a compound/cell line combination were calculated by dividing

maximum GI50 by minimum GI50 value. The number of experiments on a specific date for a compound/cell line combination was used to determine the maximum GI50 fold change over the same number of experiments picked randomly from that combination's experimental timeline. This was performed 1000 times so that for every compound/cell line combination and experimental date with a maximum GI50 fold change over multiple experiments there were 1000 corresponding maximum GI50 fold changes calculated from random samples of the same number of experiments on that compound/cell line combination's timeline. The mean maximum GI50 fold change was calculated for the 1000 random samples for each compound/cell line combination and the number of maximum GI50 fold changes for experiments on the same date higher and lower than the random sample mean were counted. For each compound, significance of the difference between same date maximum GI50 fold change and sample mean GI50 fold change was calculated using Wilcoxon Rank Sum Test. This was performed using all cell line data combined for each compound and for each cell line individually for each compound. Where a significant difference between same date and random sample mean maximum GI50 fold changes were observed, the number of times the same date GI50 fold change was higher or lower than the random sample mean maximum GI50 fold change was counted.

Drift in drug sensitivity

The mean GI50 fold change was calculated for each experimental date (month) for the 18 compounds with 100 or more experiments for at least one cell line. The GI50 fold change between the first experimental date and the last experimental date was calculated using the mean GI50 on those dates. The first/last

GI50 fold change was then compared to the maximum GI50 fold change for each compound/cell line combination and considered a candidate for a drift in sensitivity if it was 50% or more of the maximum fold change.

Removal of outliers

The adjusted boxplot method was used to identify outlier thresholds. This method was chosen as the data set was highly skewed. To use this method the medcouple (MC), a robust measure of skewness, had to be calculated (where $X_n = \{ x_1, x_2, \dots , x_n \}$ represents data for every compound/cell line combination):

$$MC(x_1, \dots , x_n) = med \frac{(x_j - med_k) - (med_k - x_i)}{x_j - x_i}$$

Where med_k is the median of X_n, and i and j have to satisfy $x_i \leq med_k \leq x_j$, and $x_i \neq x_j$.

Using the MC the upper (*U*) and lower (*L*) thresholds could be determined. If $MC \geq 0$:

$$L = Q_1 - 1.5 \times \exp(-3.5MC) \times IQR$$

$$U = Q_3 + 1.5 \times \exp(4MC) \times IQR$$

If $MC \leq 0$:

$$L = Q_1 - 1.5 \times \exp(-4MC)\, IQR$$

$$U = Q_3 + 1.5 \times \exp(3.5MC) \times IQR$$

If $MC = 0$ the adjusted boxplot method was not used but instead the Tukey method was used:

$$L = Q_1 - 1.5IQR$$

$$U = Q_3 + 1.5IQR$$

Where Q_1 is the lower quartile, Q_3 is the upper quartile and IQR is the interquartile range.

For each compound/cell line combination any GI50 value below L or above U were removed from the dataset. Analyses were performed on this dataset as previously described for the complete dataset.

Data processing

Data was carried out using perl version 5.26.0, Microsoft Excel (2011) and R statistical packages version 3.4.4. Perl modules Statistics:Descriptive and Statistics::R were used. Packages used in R were robustbase, dplyr, webr, moonBook, tidyverse, reshape2, scales, gplots, ggpubr, ggExtra, RColorBrewer, corrplot, ggplot2, and tidyr.

Author contributions

Data acquisition and analysis (I.R.), data interpretation (all authors), study conception (M.N.W., M.M.), study design (I.R., M.N.W., M.M.), manuscript drafting (I.R., M.M.), manuscript revision (all authors).

All authors have approved the submitted version and agreed both to be personally accountable for their own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which they were not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Conflicts of interest

The authors declare no competing interests.

Data availability

All raw data are available from CellMiner [80] Version 2.2. All other data are provided in the manuscript and its supplements.

References

1. França TF, Monserrat JM. Reproducibility crisis in science or unrealistic expectations? EMBO Rep. 2018 Jun;19(6). pii: e46008.

2. Wass MN, Ray L, Michaelis M. Understanding of researcher behavior is required to improve data reliability. Gigascience. 2019 May 1;8(5). pii: giz017.

3. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer. 2006 Oct;6(10):813-23.

4. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn K, Morris J, Doroshow J and Pommier Y. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. Cancer Res. 2012 Jul 15; 72(14): 3499-3511.

5. Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. Circ Res. 2015 Jan 2;116(1):116-26.

6. Frye SV, Arkin MR, Arrowsmith CH, Conn PJ, Glicksman MA, Hull-Ryde EA, Slusher BS. Tackling reproducibility in academic preclinical drug discovery. Nat Rev Drug Discov. 2015 Nov;14(11):733-4.

7. Drucker DJ. Never Waste a Good Crisis: Confronting Reproducibility in Translational Research. Cell Metab. 2016 Sep 13;24(3):348-360.

8. Fanelli D. Opinion: Is science really facing a reproducibility crisis, and do we need it to? Proc Natl Acad Sci U S A. 2018 Mar 13;115(11):2628-2631.

9. Samsa G, Samsa L. A Guide to Reproducibility in Preclinical Research. Acad Med. 2019 Jan;94(1):47-52.

10. Munafò MR, Chambers CD, Collins AM, Fortunato L, Macleod MR. Research Culture and Reproducibility. Trends Cogn Sci. 2020 Feb;24(2):91-93.

528 11. Mobley A, Linder SK, Braeuer R, Ellis LM, Zwelling L. A survey on data
529 reproducibility in cancer research provides insights into our limited ability to translate
530 findings from the laboratory to the clinic. PLoS One 2013;8(5):e63221.
531 12. Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016;533:452-4.
532 13. Nature Editorial. Checklists work to improve science. Nature 2018;556:273-4.
533 14. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on
534 published data on potential drug targets? Nat Rev Drug Discov. 2011;10:712.
535 15. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer
536 research. Nature 2012;483:531-3.
537 16. Aird F, Kandela I, Mantis C; Reproducibility Project: Cancer Biology. Replication
538 Study: BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Elife
539 2017;6. pii: e21253.
540 17. Kandela I, Aird F; Reproducibility Project: Cancer Biology. Replication Study:
541 Discovery and preclinical validation of drug indications using compendia of public
542 gene expression data. Elife 2017;6. pii: e17044.
543 18. Shan X, Fung JJ, Kosaka A, Danet-Desnoyers G; Reproducibility Project: Cancer
544 Biology. Replication Study: Inhibition of BET recruitment to chromatin as an effective
545 treatment for MLL-fusion leukaemia. Elife 2017;6. pii: e25306.
546 19. Showalter MR, Hatakeyama J, Cajka T, VanderVorst K, Carraway KL, Fiehn O;
547 Reproducibility Project: Cancer Biology. Replication Study: The common feature of
548 leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity
549 converting alpha-ketoglutarate to 2-hydroxyglutarate. Elife 2017;6. pii: e26030.
550 20. Essex A, Pineda J, Acharya G, Xin H, Evans J; Reproducibility Project: Cancer
551 Biology, Iorns E, Tsui R, Denis A, Perfito N, Errington TM, Iorns E, Tsui R, Denis A,
552 Perfito N, Errington TM. Replication Study: Wnt activity defines colon cancer stem

553 cells and is regulated by the microenvironment. Elife. 2019 Jun 19;8. pii: e45426.
554 doi: 10.7554/eLife.45426.
555 21. Horrigan SK; Reproducibility Project: Cancer Biology. Replication Study: The
556 CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for
557 human solid tumors. Elife 2017;6. pii: e18173.
558 22. Horrigan SK, Courville P, Sampey D, Zhou F, Cai S; Reproducibility Project:
559 Cancer Biology. Replication Study: Melanoma genome sequencing reveals frequent
560 PREX2 mutations. Elife 2017;6. pii: e21634. doi: 10.7554/eLife.21634.
561 23. Lewis LM, Edwards MC, Meyers ZR, Talbot CC, Hao H, Blum D, Reproducibility
562 Project: Cancer Biology, Iorns E, Tsui R, Denis A, Perfito N, Errington TM.
563 Replication Study: Transcriptional amplification in tumor cells with elevated c-Myc.
564 Elife 2018;7. pii: e30274.
565 24. Eaton K, Pirani A, Snitkin ES; Reproducibility Project: Cancer Biology, Iorns E,
566 Tsui R, Denis A, Perfito N, Errington TM. Replication Study: Intestinal inflammation
567 targets cancer-inducing activity of the microbiota. Elife 2018;7. pii: e34364.
568 25. Kim J, Afshari A, Sengupta R, Sebastiano V, Gupta A, Kim YH; Reproducibility
569 Project: Cancer Biology, Iorns E, Tsui R, Denis A, Perfito N, Errington TM.
570 Replication study: Melanoma exosomes educate bone marrow progenitor cells
571 toward a pro-metastatic phenotype through MET. Elife. 2018 Dec 11;7. pii: e39944.
572 26. Vanden Heuvel JP, Maddox E, Maalouf SW; Reproducibility Project: Cancer
573 Biology, Iorns E, Tsui R, Denis A, Perfito N, Errington TM. Replication Study:
574 Systematic identification of genomic markers of drug sensitivity in cancer cells. Elife
575 2018;7. pii: e29747.
576 27. Sheen MR, Fields JL, Northan B, Lacoste J, Ang LH, Fiering S; Reproducibility
577 Project: Cancer Biology, Iorns E, Tsui R, Denis A, Perfito N, Errington TM.

578 Replication Study: Biomechanical remodeling of the microenvironment by stromal
579 caveolin-1 favors tumor invasion and metastasis. Elife. 2019 Dec 17;8. pii: e45120.
580 doi: 10.7554/eLife.45120.
581 28. Kerwin J, Khan I; Reproducibility Project: Cancer Biology, Iorns E, Tsui R, Denis
582 A, Perfito N, Errington TM. Replication Study: A coding-independent function of gene
583 and pseudogene mRNAs regulates tumour biology. Elife. 2020 Apr 21;9:e51019.
584 29. Mantis C, Kandela I, Aird F; Reproducibility Project: Cancer Biology. Replication
585 Study: Coadministration of a tumor-penetrating peptide enhances the efficacy of
586 cancer drugs. Elife 2017;6. pii: e17584. doi: 10.7554/eLife.17584.
587 30. Repass J; Reproducibility Project: Cancer Biology, Iorns E, Denis A, Williams
588 SR, Perfito N, Errington TM. Replication Study: Fusobacterium nucleatum infection is
589 prevalent in human colorectal carcinoma. Elife 2018;7. pii: e25801.
590 31. Yan X, Tang B, Chen B, Shan Y, Yang H; Reproducibility Project: Cancer
591 Biology, Iorns E, Tsui R, Denis A, Perfito N, Errington TM. Replication Study: The
592 microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly
593 repressing CD44. Elife. 2019 Mar 12;8. pii: e43511.
594 32. Wang H, Radomska HS, Phelps MA; Reproducibility Project: Cancer Biology,
595 Iorns E, Tsui R, Denis A, Perfito N, Errington TM. Replication Study: Coding-
596 independent regulation of the tumor suppressor PTEN by competing endogenous
597 mRNAs. Elife. 2020 Oct 19;9:e56651.
598 33. Fanelli D, Costas R, Ioannidis JP. Meta-assessment of bias in science. Proc Natl
599 Acad Sci U S A. 2017;114:3714-9.
600 34. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S,
601 Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF,
602 Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Thibault

603 J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P
604 Jr, de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S,
605 Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li
606 N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J,
607 Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR,
608 Schlegel R, Garraway LA. The Cancer Cell Line Encyclopedia enables predictive
609 modelling of anticancer drug sensitivity. Nature. 2012 Mar 28;483(7391):603-7.
610 35. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW,
611 Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L,
612 Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, Lutz SR,
613 Kogera F, Lawrence K, McLaren-Douglas A, Mitropoulos X, Mironenko T, Thi H,
614 Richardson L, Zhou W, Jewitt F, Zhang T, O'Brien P, Boisvert JL, Price S, Hur W,
615 Yang W, Deng X, Butler A, Choi HG, Chang JW, Baselga J, Stamenkovic I,
616 Engelman JA, Sharma SV, Delattre O, Saez-Rodriguez J, Gray NS, Settleman J,
617 Futreal PA, Haber DA, Stratton MR, Ramaswamy S, McDermott U, Benes CH.
618 Systematic identification of genomic markers of drug sensitivity in cancer cells.
619 Nature. 2012 Mar 28;483(7391):570-5.
620 36. Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJ,
621 Quackenbush J. Inconsistency in large pharmacogenomic studies. Nature. 2013 Dec
622 19;504(7480):389-93.
623 37. Cancer Cell Line Encyclopedia Consortium; Genomics of Drug Sensitivity in
624 Cancer Consortium. Pharmacogenomic agreement between two cancer cell line data
625 sets. Nature. 2015 Dec 3;528(7580):84-7.

626 38. Bouhaddou M, DiStefano MS, Riesel EA, Carrasco E, Holzapfel HY, Jones DC,
627 Smith GR, Stern AD, Somani SS, Thompson TV, Birtwistle MR. Drug response
628 consistency in CCLE and CGP. Nature 2016;540:E9-E10.
629 39. Geeleher P, Gamazon ER, Seoighe C, Cox NJ, Huang RS. Consistency in large
630 pharmacogenomic studies. Nature 2016;540:E1-E2.
631 40. Mpindi JP, Yadav B, Östling P, Gautam P, Malani D, Murumägi A, Hirasawa A,
632 Kangaspeska S, Wennerberg K, Kallioniemi O, Aittokallio T. Consistency in drug
633 response profiling. Nature 2016;540:E5-E6.
634 41. Safikhani Z, El-Hachem N, Smirnov P, Freeman M, Goldenberg A, Birkbak NJ,
635 Beck AH, Aerts HJ, Quackenbush J, Haibe-Kains B. Safikhani et al. reply. Nature.
636 2016 Nov 30;540(7631):E2-E4.
637 42. Safikhani Z, El-Hachem N, Smirnov P, Freeman M, Goldenberg A, Birkbak NJ,
638 Beck AH, Aerts HJ, Quackenbush J, Haibe-Kains B. Safikhani et al. reply. Nature.
639 2016 Nov 30;540(7631):E6-E8.
640 43. Safikhani Z, El-Hachem N, Smirnov P, Freeman M, Goldenberg A, Birkbak NJ,
641 Beck AH, Aerts HJ, Quackenbush J, Haibe-Kains B. Safikhani et al. reply. Nature.
642 2016 Nov 30;540(7631):E11-E12.
643 44. Safikhani Z, El-Hachem N, Quevedo R, Smirnov P, Goldenberg A, Juul Birkbak
644 N, Mason C, Hatzis C, Shi L, Aerts HJ, Quackenbush J, Haibe-Kains B. F1000Res.
645 2016 May 9;5:825.
646 45. Safikhani Z, Smirnov P, Freeman M, El-Hachem N, She A, Rene Q, Goldenberg
647 A, Birkbak NJ, Hatzis C, Shi L, Beck AH, Aerts HJWL, Quackenbush J, Haibe-Kains
648 B. Revisiting inconsistency in large pharmacogenomic studies. Version 3. F1000Res.
649 2016 Sep 16 [revised 2017 Jan 1];5:2333.

650 46. Monks A, Scudiero DA, Johnson GS, Paull KD, Sausville EA. The NCI anti-
651 cancer drug screen: a smart screen to identify effectors of novel targets. Anticancer
652 Drug Des. 1997 Oct;12(7):533-41.
653 47. Reinhold WC, Varma S, Sousa F, Sunshine M, Abaan OD, Davis SR, Reinhold
654 SW, Kohn KW, Morris J, Meltzer PS, Doroshow JH, Pommier Y. NCI-60 whole
655 exome sequencing and pharmacological CellMiner analyses. PLoS One. 2014 Jul
656 17;9(7):e101670.
657 48. Reinhold WC, Sunshine M, Varma S, Doroshow JH, Pommier Y. Using CellMiner
658 1.6 for Systems Pharmacology and Genomic Analysis of the NCI-60. Clin Cancer
659 Res. 2015 Sep 1;21(17):3841-52.
660 49. Chabner BA. NCI-60 Cell Line Screening: A Radical Departure in its Time. J Natl
661 Cancer Inst. 2016 Jan 11;108(5). pii: djv388.
662 50. Reinhold WC, Varma S, Sunshine M, Rajapakse V, Luna A, Kohn KW,
663 Stevenson H, Wang Y, Heyn H, Nogales V, Moran S, Goldstein DJ, Doroshow JH,
664 Meltzer PS, Esteller M, Pommier Y. The NCI-60 Methylome and Its Integration into
665 CellMiner. Cancer Res. 2017 Feb 1;77(3):601-612.
666 51. Reinhold WC, Varma S, Sunshine M, Elloumi F, Ofori-Atta K, Lee S, Trepel JB,
667 Meltzer PS, Doroshow JH, Pommier Y. RNA sequencing of the NCI-60: Integration
668 into CellMiner and CellMiner CDB. Cancer Res. 2019 May 21. pii: canres.2047.2018.
669 52. Kormanec J, Novakova R, Csolleiova D, Feckova L, Rezuchova B, Sevcikova B,
670 Homerova D. The antitumor antibiotic mithramycin: new advanced approaches in
671 modification and production. Appl Microbiol Biotechnol. 2020 Sep;104(18):7701-
672 7721.

673 53. Gallego-Jara J, Lozano-Terol G, Sola-Martínez RA, Cánovas-Díaz M, de Diego
674 Puente T. A Compressive Review about Taxol®: History and Future Challenges.
675 Molecules. 2020 Dec 17;25(24):5986.
676 54. Rae JM, Creighton CJ, Meck JM, Haddad BR, Johnson MD. MDA-MB-435 cells
677 are derived from M14 melanoma cells--a loss for breast cancer, but a boon for
678 melanoma research. Breast Cancer Res Treat. 2007 Jul;104(1):13-9.
679 55. Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, Strathdee CA,
680 Dempster J, Lyons NJ, Burns R, Nag A, Kugener G, Cimini B, Tsvetkov P, Maruvka
681 YE, O'Rourke R, Garrity A, Tubelli AA, Bandopadhayay P, Tsherniak A, Vazquez F,
682 Wong B, Birger C, Ghandi M, Thorner AR, Bittker JA, Meyerson M, Getz G,
683 Beroukhim R, Golub TR. Genetic and transcriptional evolution alters cancer cell line
684 drug response. Nature. 2018 Aug;560(7718):325-330.
685 56. Noronha N, Ehx G, Meunier MC, Laverdure JP, Thériault C, Perreault C. Major
686 multilevel molecular divergence between THP-1 cells from different biorepositories.
687 Int J Cancer. 2020 Mar 12. doi: 10.1002/ijc.32967.
688 57. Voskuil JLA, Bandrowski A, Begley CG, Bradbury ARM, Chalmers AD, Gomes
689 AV, Hardcastle T, Lund-Johansen F, Plückthun A, Roncador G, Solache A, Taussig
690 MJ, Trimmer JS, Williams C, Goodman SL. The Antibody Society's antibody
691 validation webinar series. MAbs. 2020 Jan-Dec;12(1):1794421.
692 58. Hubert M, Vandervieren E. An adjusted boxplot for skewed distributions.
693 Computational Statistics and Data Analysis. 2008; 52(12): 5186-5201
694 59. Liu Y, Mi Y, Mueller T, Kreibich S, Williams EG, Van Drogen A, Borel C, Frank M,
695 Germain PL, Bludau I, Mehnert M, Seifert M, Emmenlauer M, Sorg I, Bezrukov F,
696 Bena FS, Zhou H, Dehio C, Testa G, Saez-Rodriguez J, Antonarakis SE, Hardt WD,

697 Aebersold R. Multi-omic measurements of heterogeneity in HeLa cells across
698 laboratories. Nat Biotechnol. 2019 Mar;37(3):314-322.
699 60. Lorenzi PL, Reinhold WC, Varma S, Hutchinson AA, Pommier Y, Chanock SJ,
700 Weinstein JN. DNA fingerprinting of the NCI-60 cell line panel. Mol Cancer Ther.
701 2009 Apr;8(4):713-24.
702 61. Eisenhauer EA, O'Dwyer PJ, Christian M, Humphrey JS. Phase I clinical trial
703 design in cancer drug development. J Clin Oncol. 2000 Feb;18(3):684-92.
704 62. Sachs JR, Mayawala K, Gadamsetty S, Kang SP, de Alwis DP. Optimal Dosing
705 for Targeted Therapies in Oncology: Drug Development Cases Leading by Example.
706 Clin Cancer Res. 2016 Mar 15;22(6):1318-24.
707 63. Mansinho A, Boni V, Miguel M, Calvo E. New designs in early clinical drug
708 development. Ann Oncol. 2019 Sep 1;30(9):1460-1465.
709 64. Corbaux P, El-Madani M, Tod M, Péron J, Maillet D, Lopez J, Freyer G, You B.
710 Clinical efficacy of the optimal biological dose in early-phase trials of anti-cancer
711 targeted therapies. Eur J Cancer. 2019 Oct;120:40-46.
712 65. Begley CG. Six red flags for suspect work. Nature 2013;497:433-4.
713 66. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D,
714 Schulz KF, Tibshirani R. Increasing value and reducing waste in research design,
715 conduct, and analysis. Lancet 2014;383:166-75.
716 67. Hatzis C, Bedard PL, Birkbak NJ, Beck AH, Aerts HJ, Stem DF, Shi L, Clarke R,
717 Quackenbush J, Haibe-Kains B. Enhancing reproducibility in cancer drug screening:
718 how do we move forward? Cancer Res. 2014 Aug 1;74(15):4016-23.
719 68. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in
720 Preclinical Research. PLoS Biol. 2015;13:e1002165.

721 69. Jarvis MF, Williams M. Irreproducibility in Preclinical Biomedical Research:
722 Perceptions, Uncertainties, and Knowledge Gaps. Trends Pharmacol Sci.
723 2016;37:290-302.
724 70. Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: Progress and
725 priorities. F1000Res. 2017;6:604.
726 71. Kaelin WG Jr. Publish houses of brick, not mansions of straw. Nature
727 2017;545:387.
728 72. Karp NA, Speak AO, White JK, Adams DJ, Hrabé de Angelis M, Hérault Y, Mott
729 RF. Impact of temporal variation on design and analysis of mouse knockout
730 phenotyping studies. PLoS One. 2014 Oct 24;9(10):e111239.
731 73. Ding KF, Finlay D, Yin H, Hendricks WPD, Sereduk C, Kiefer J, Sekulic A,
732 LoRusso PM, Vuori K, Trent JM, Schork NJ. Analysis of variability in high throughput
733 screening data: applications to melanoma cell lines and drug responses. Oncotarget.
734 2017 Apr 25;8(17):27786-27799.
735 74. Kafkafi N, Golani I, Jaljuli I, Morgan H, Sarig T, Würbel H, Yaacoby S, Benjamini
736 Y. Addressing reproducibility in single-laboratory phenotyping experiments. Nat
737 Methods. 2017 Apr 27;14(5):462-464.
738 75. Karp NA. Reproducible preclinical research-Is embracing variability the answer?
739 PLoS Biol. 2018 Mar 5;16(3):e2005413.
740 76. Milcu A, Puga-Freitas R, Ellison AM, Blouin M, Scheu S, Freschet GT, Rose L,
741 Barot S, Cesarz S, Eisenhauer N, Girin T, Assandri D, Bonkowski M, Buchmann N,
742 Butenschoen O, Devidal S, Gleixner G, Gessler A, Gigon A, Greiner A, Grignani C,
743 Hansart A, Kayler Z, Lange M, Lata JC, Le Galliard JF, Lukac M, Mannerheim N,
744 Müller MEH, Pando A, Rotter P, Scherer-Lorenzen M, Seyhun R, Urban-Mead K,

745 Weigelt A, Zavattaro L, Roy J. Genotypic variability enhances the reproducibility of
746 an ecological study. Nat Ecol Evol. 2018 Feb;2(2):279-287.
747 77. Voelkl B, Vogt L, Sena ES, Würbel H. Reproducibility of preclinical animal
748 research improves with heterogeneity of study samples. PLoS Biol. 2018 Feb
749 22;16(2):e2003693.
750 78. Bodden C, von Kortzfleisch VT, Karwinkel F, Kaiser S, Sachser N, Richter SH.
751 Heterogenising study samples across testing time improves reproducibility of
752 behavioural data. Sci Rep. 2019 Jun 3;9(1):8247.
753 79. von Kortzfleisch VT, Karp NA, Palme R, Kaiser S, Sachser N, Richter SH.
754 Improving reproducibility in animal research by splitting the study population into
755 several 'mini-experiments'. Sci Rep. 2020 Oct 6;10(1):16579.
756 80. Shankavaram UT, Varma S, Kane D, Sunshine M, Chary KK, Reinhold WC,
757 Pommier Y, Weinstein JN. CellMiner: a relational database and query tool for the
758 NCI-60 cancer cell lines. BMC Genomics. 2009 Jun 23;10:277.
759
760