

Ranking small molecules by how much they preferentially inhibit the growth of cancer cell lines with either BRAF or KRAS oncogene mutations

James J. Langham*
SciSolve Software, LLC
Seattle, WA
USA

Abstract

We were interested in the question of whether it might be possible to use knowledge of cancer-related mutations in the cell lines of the NCI60 screening data set to identify small molecules that preferentially inhibit the growth of cell lines containing either BRAF or KRAS oncogene mutations. Our hypothesis was that this cell line mutation knowledge could help to identify small molecules that were more likely to preferentially inhibit growth of cell lines with a particular mutation. It seems that any such molecules might be further investigated to try to better understand the molecular mechanisms of growth inhibition.

We defined a quantity, Diff_{mut} , that estimates how much more a given small molecule inhibits cell lines with a mutation of interest than cell lines without that mutation. We ranked the small molecules in descending order of Diff_{mut} and then tried to explain whether the ranking of the highest ranked molecules made sense in terms of independent facts about these molecules.

This method showed the BRAF inhibitor vemurafenib to be highly ranked in the BRAF ranking. The cytidine analog cytarabine was found to be highly ranked in the KRAS ranking. Other cytidine analogs were also found to be highly ranked with respect to KRAS.

*Email:james@scisolve.com

23 **Introduction**

24 Tens of thousands of small molecules have been tested against the US National Cancer Insti-
25 tute (NCI) 60 tumor cell line anticancer drug screen (NCI60) since the late 1980s (Shoemaker,
26 2006). The 60 tumor cell lines were all taken from patients with cancers of nine different
27 tissue types (Table 1). The tens of thousands of small molecules have been screened by the
28 NCI for their ability to inhibit in vitro growth of the 60 cell lines. In 2006 a study was pub-
29 lished detailing whether 24 common cancer-related mutations were found in the NCI60 cell
30 lines (Ikediobi et al., 2006). These mutations included some in the p53 tumor suppressor gene
31 as well as the BRAF (B-Raf proto-oncogene, serine/threonine kinase) and KRAS (Kirsten
32 rat sarcoma viral oncogene homolog) oncogenes (Gray et al., 2013).

NCI60 Tissues
Leukemia
Non-small-cell lung
Colon
CNS
Melanoma
Ovarian
Renal
Prostate
Breast

Table 1: The cancer tissue types in the NCI60 cell lines (Shoemaker, 2006).

33 We were interested in the question of whether it might be possible to use the NCI60
34 screening data and knowledge of the mutations in the cell lines to identify molecules that
35 preferentially inhibited the growth of cell lines with particular mutations. Two approaches
36 to find such molecules suggested themselves to us.

37 The first approach that suggested itself was to rank molecules that had been tested
38 against the NCI60 cell line panel in numerical order of some quantity that estimates how
39 much each molecule preferentially inhibits the growth of cell lines with a mutation of interest
40 versus cell lines that lacked that mutation. A simple way of calculating such a quantity
41 is described in the Materials and Methods Section. Such a ranking might help identify
42 molecules that inhibited cell lines with a particular mutation. This might be helpful in
43 highlighting particular molecules for further study. Such further study might be to investigate
44 the molecular mechanism by which the observed growth inhibition is achieved. Further study
45 might also include chemically modifying highly ranked molecules and testing such analogs in
46 order to explore structure activity relationships.

47 The second approach, closely related to the first, would be to make statistical or machine
48 learning models of the preferential inhibition as a function of the molecular structures of the
49 screened molecules and perhaps other relevant information. Assuming that predictive models
50 could be built then they might be used to screen large data sets of molecules to suggest novel
51 molecules that have preferential inhibition for the mutation of interest. Any such suggestions
52 would need to be confirmed by experimental testing but such computational screening might
53 be helpful by helping to identify interesting molecules faster and with less cost.

As a specific example consider the BRAF oncogene mutations listed for the NCI60 cell lines shown in Table 2. The data in this table show that 10 of the cell lines have the V600E mutation in BRAF while one has the G464V mutation. In the NCI60 screening data screened molecules are tested against the cell lines and the negative \log_{10} of the concentration at which 50% growth inhibition (pGI50) occurs is reported. We expected that a molecule that preferentially inhibited the growth of cell lines with a particular mutation, like the BRAF V600E mutation, would have a higher average pGI50 for those cell lines with V600E mutations versus those that lacked this mutation. One could then rank the molecules in terms of the difference between the two types of pGI50 average. The main hypothesis of this paper is that there might be molecules in the NCI60 data set which have higher average pGI50 values for cell lines with a particular mutation than those that lack the mutation. The experimental plan was to look for such molecules in the data set and where possible look for independent literature evidence to help explain why such molecules might selectively inhibit the cell lines with the particular mutation.

Cell Line	Tissue	BRAF Mutation
MDA-MB-231/ATCC	Breast	G464V
MDA-MB-435	Melanoma	V600E
MALME-3M	Melanoma	V600E
HT29	Colon	V600E
COLO 205	Colon	V600E
LOX IMVI	Melanoma	V600E
SK-MEL-28	Melanoma	V600E
UACC-257	Melanoma	V600E
M14	Melanoma	V600E
SK-MEL-5	Melanoma	V600E
UACC-62	Melanoma	V600E

Table 2: Cells with BRAF mutations in the NCI60 cell lines. The particular amino acid mutation is shown (Ikediobi et al., 2006).

The second approach, to look for novel molecules that are selective for cell lines with a particular mutation using statistical modeling, presupposes that a valid method exists for ranking such selectivity. This current paper explores whether a statistic computed from the publicly available NCI60 screening data may be used to identify such selectivity. In future work we plan to use statistical modeling to build models of selectivity of cell lines with a particular mutation. The focus of this paper is to try to first show that such a statistic may be meaningful.

We looked at two oncogenes separately for this paper, mutated BRAF and mutated KRAS. The specific oncogene-causing mutations were those studied in NCI60 cell lines by Ikediobi et al. (Ikediobi et al., 2006). The BRAF and KRAS proteins are component kinases in the $RAS \rightarrow RAF \rightarrow MAP \rightarrow ERK$ signaling pathway. RAS is a GTP kinase which is central to several important signaling cascades in addition to the $RAS \rightarrow RAF \rightarrow MEK \rightarrow ERK$ signaling pathway (Weinberg, 2014a). KRAS, a variant of RAS, has been implicated in several cancers, including in 90% of pancreatic cancers and has been a major target of drug discovery efforts for three decades continuing to the present (Stephen et al., 2014; Weinberg,

2014b; Wang et al., 2013; Thompson, 2013). RAF is an effector of RAS in the RAS → RAF → MEK → ERK pathway. Mutated BRAF, a variant of RAF, is implicated in about 80% of incidents of melanoma (Namba et al., 2003). Since 2011 two therapeutics, vemurafenib and dabrafenib, have been approved to treat melanoma, both of which target BRAF with the V600E mutation (Nguyen, 2013).

These two oncogenes were selected because they are prominent drug targets. As mentioned, there are approved drugs for V600E mutated BRAF making it possible to see if known inhibitors were ranked highly. There are no widely acknowledged inhibitors of mutated KRAS to serve as a test of the ranking but because of the interest in KRAS as a target we thought it would be interesting to see which molecules would be highly ranked.

Expected Results

To recap, our hypothesis was that there might be molecules in the NCI60 data set that had higher average pGI50 values for those cell lines which contained an oncogene mutation of interest than those cell lines that lacked that mutation. Our expectation was that those molecules that were tested as having a higher average pGI50 with respect to the mutation would include known inhibitors of the oncogene of interest. In the case of the BRAF V600E mutation we expected that known BRAF V600E inhibitors such as vemurafenib and dabrafenib would have higher pGI50 values for those cell lines with BRAF V600E mutations than those cell lines that lacked them. We did not have any clear expectations with respect to the KRAS oncogene mutations since there are currently no approved drugs that selectively target cancer cells with mutated KRAS. However we planned to do literature research on any molecules that had high pGI50 values for cell lines with KRAS mutations.

Materials and Methods

Since the general strategy was to look for tested molecules that inhibited cell lines with particular mutations more than those that lacked those mutations a way of estimating such selectivity was needed in terms of some numerical quantity that could be sorted. The quantity that we chose was, for each molecule tested, to calculate the difference between the mean pGI50 for those cell lines with a mutation of interest ($\overline{\text{pGI50}}_{\text{mut}}$) versus the mean pGI50 of those cell lines that lacked the mutation and were assumed to have wild-type versions of the gene mutation in question ($\overline{\text{pGI50}}_{\text{wt}}$). Equation (1) shows the formula for this difference which we refer to by Diff_{mut} .

$$\text{Diff}_{\text{mut}} = \overline{\text{pGI50}}_{\text{mut}} - \overline{\text{pGI50}}_{\text{wt}} \quad (1)$$

The GI50 data from the September 2012 NCI60 data release were downloaded from the NCI website (NCI-DTP, 2012). This data set had screening data for 46,223 molecules. Each molecule was associated with varying numbers of cell line screens. We pruned the data to remove molecules that had been tested on less than 50 of the 60 cell lines. We also removed those molecules where the GI50 range was less than 1.2 \log_{10} units. These were similar filters to those used by Reinhold et al. (Reinhold et al., 2012) in order to remove “nonresponsive and out of proper range data.” However Reinhold et al. retained molecules tested against 35

or more cell lines whereas we only kept molecules that had been tested against 50 or more cell lines. We kept data with more cell lines in order to make it more probable that some of the mutated cell lines of interest would be present in each cell line profile for all molecules. After these filters 14,487 molecules remained, each of which had both been tested on 50 or more cell lines and had a range of pGI50 values that were $\geq 1.2 \log_{10}$ units.

The pGI50 values used to compute the means were standardized as “z-scores” using Equation (2) with the sample mean ($\hat{\mu}$) and sample standard deviation ($\hat{\sigma}$) (Kreyszig, 1999) calculated with respect to each cell line to try and account for different responses across different cell lines.

$$\text{z-score} = \frac{\text{pGI50} - \hat{\mu}}{\hat{\sigma}} \quad (2)$$

The two sets of mutations that we looked at were the V600E mutation of BRAF and any of the KRAS mutations at codons 12, 13, and 61 that had been sequenced in the Ikediobi study (Ikediobi et al., 2006). The KRAS mutations are shown in Table 3. These mutations were chosen because they are associated with cancer cell proliferation. The V600E mutation of BRAF has been shown to drive cancer cell proliferation and the mutated protein is the drug target for the approved melanoma drugs vemurafenib and dabrafenib (Bollag et al., 2012). In KRAS the codons 12, 13, and 61 are situated around the active site where the dephosphorylation reaction of GTP to GDP is catalyzed. Mutations at these three sites interfere with the ability of the KRAS GTPase to dephosphorylate GTP and are associated with cancer cell proliferation (Weinberg, 2014c).

Using Equation (1) a value was calculated for each of the 14,487 molecules, one for the KRAS Diff_{mut} and one for the BRAF Diff_{mut} . These values were then sorted in numerical descending order. Those values that were higher were interpreted as indicating that the molecule tended to inhibit growth more for cells within the NCI60 panel with that mutation.

Cell Line	Tissue	KRAS Mutation
HCT-15	Colon	G13D
MDA-MB-231/ATCC	Breast	G13D
OVCAR-5	Ovarian	G12V
A549/ATCC	Non-Small Cell Lung	G12S
HOP-62	Non-Small Cell Lung	G12C
RPMI-8226	Leukemia	G12A
SW-620	Colon	G12V
CCRF-CEM	Leukemia	G12D
NCI-H23	Non-Small Cell Lung	G12C
NCI-H460	Non-Small Cell Lung	Q61H
HCC-2998	Colon	A146T
HCT-116	Colon	G13D

Table 3: Cells with KRAS mutations in the NCI60 cell lines. The particular amino acid mutation is shown (Ikediobi et al., 2006).

Resampling was used to estimate p-values and confidence intervals for the Diff_{mut} values (Diaconis and Efron, 1983; Good, 2005). Permutation testing was done to estimate

the p-values and the bootstrap method was used to estimate the confidence intervals. The p-values were estimated by shuffling the z-scores for the individual cell line pGI50s for a screened molecule and then calculating the Diff_{mut} for these shuffled values. This was done for 1000 trials and the p-value estimated by counting the number of times a shuffled value exceeded the calculated value and dividing by 1000. So the resulting p-value was an estimate for the probability that the measured value would occur by chance. The bootstrap method was used to estimate the 95% confidence interval. That is, the data set was sampled using replacement for 1000 trials and the confidence interval estimated from the calculated values for the trials.

We cross-referenced the top-ranked molecules with the names and FDA status assigned in the Cell Miner data set (CellMiner, 2012). This FDA status listed whether a molecule was either an approved drug or in clinical trials, information which was useful in interpreting the top-ranked molecules in the data set.

We examined the 50 top-ranked molecules for both the KRAS and BRAF screens, looking at the molecular structure diagrams and FDA status among other information. For KRAS the p-values for these 50 molecules ranged from < 0.001 up to 0.11. For BRAF the p-values ranged from < 0.001 up to 0.033. We used this 50 molecule cutoff because it was a relatively small number to inspect visually and the p-values were relatively low which indicated that the values were less likely to have occurred by chance.

Experimental Details

The September 2012 GI50 data was downloaded from the NCI DTP website (NCI-DTP, 2012). We used these GI50 values for our calculations.

We downloaded the CellMiner version of the DTP NCI60 set (CellMiner, 2012) because it included extra information about many of the molecules in the data set which was helpful for interpreting results. The NCI DTP program also helpfully provided a file to map from the NCI molecule ids (NSC ID) to PubChem compound ids in response to an email request. This was helpful in being able to associate chemical structures downloaded from PubChem (Bolton et al., 2008) with the NSC IDs in the GI50 data set.

The data for the mutations in the cell lines was obtained by downloading the COSMIC (Catalogue of Somatic Mutations In Cancer) (Forbes et al., 2008) cell line project data set (COSMIC, 2013).

The calculations were done by means of short programs written in JavaScript using the Node.js server-side JavaScript implementation (Dahl, 2014). We chose JavaScript as the programming language because we anticipated building a web application to visualize results and expected that using JavaScript would simplify development (web browsers typically use JavaScript as the programming language and having server-side code in the same language as the client should simplify code sharing between client and server). GNU Make was used to coordinate the downloading of data sets and generating of results using the programs (Feldman, 2014; Bostock, 2013). The git revision control system was used to keep track of changes in the source files including the Makefile (Torvalds, 2005).

There were small differences between the names of cell lines in the downloaded data sets. For example in the COSMIC data the HT29 cell line was listed as “HT-29” whereas in the NCI DTP data set it was referred to as “HT29.” Therefore it was necessary to include

transformations in the code in order to standardize the cell line names.

The OpenBabel chemistry software library was used to generate the 2D structure diagrams from the PubChem 2D coordinates (OBoyle et al., 2011).

The R statistical package was used to generate 2D plots and to estimate some p values and confidence intervals (R Core Team, 2014).

As mentioned the results in this paper were generated using a Makefile. Results were generated on both the Debian Linux (running on Google Compute Engine) and Mac OS X (running on a late 2013 Mac Book Pro laptop) operating systems and results were very similar on both systems.

The molecular graphics image in Figure 5 was prepared using the UCSF Chimera package (Pettersen et al., 2004).

Results and Discussion

Table 4 shows the ten highest ranked molecules with respect to the BRAF V600E Diff_{mut} value and another molecule, 17-amino-17-demethoxygeldanamycin (NSC 255109), that was annotated in CellMiner as being in a clinical trial. The highest ranked molecule was selumetinib (Figure 1). Selumetinib (NSC 741078) is a MEK 1/2 (mitogen-activated protein kinase kinase 1/2 (Gray et al., 2013)) inhibitor (Troiani et al., 2012) so it inhibits one of the component signaling proteins in the RAS \rightarrow RAF \rightarrow MEK \rightarrow ERK pathway of which BRAF is also a component.

Rank	NSC ID	Name	FDA Status	Diff_{mut}	p-value	C.I.
1	741078	Selumetinib	Clinical Trial	2.45	< 0.001	1.7-3.1
2	761431	Vemurafenib	Approved	1.87	< 0.001	1.4-2.3
3	706829	-	-	1.40	< 0.001	0.7-2.1
4	354462	Hypothemycin	Clinical Trial	1.38	< 0.001	0.8-1.9
5	361127	Destruxin E	-	1.33	0.002	0.8-1.9
6	706031	-	-	1.26	< 0.001	0.6-2.0
7	299117	Sendanin	-	1.20	0.006	0.2-1.9
8	380856	-	-	1.12	0.004	0.5-1.7
9	644211	Chlorodestruxin E	-	1.10	0.002	0.5-1.6
10	662199	-	-	1.08	< 0.001	0.6-1.6
35	255109	-	Clinical Trial	0.80	0.007	0.4-1.2

Table 4: Molecules ranked by descending BRAF V600E difference. The confidence interval (C.I.) is at the 95% level.

The second-ranked molecule is vemurafenib (NSC 761431), with a Diff_{mut} value of 1.87. Vemurafenib (Figure 1) is a molecule that has been approved to treat melanoma with the BRAF V600E mutation and has been shown to selectively inhibit BRAF with the V600E mutation over wild-type BRAF (Bollag et al., 2012). Also vemurafenib has been observed to selectively inhibit cell lines that have the BRAF V600E mutation so a large Diff_{mut} value is consistent with this. (Bollag et al., 2012). Figure 2 shows a plot of the response of the NCI60 cell lines to vemurafenib. This plot shows that eight of the ten cells lines with

215 V600E mutations show greater growth inhibition than any of the cell lines without the V600E
216 mutation.

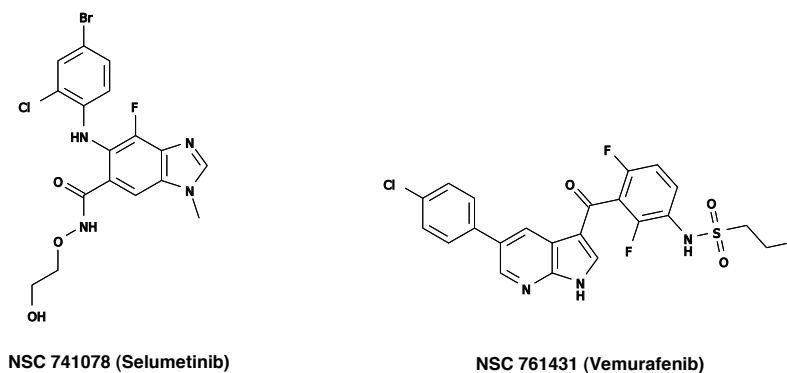


Figure 1: Structure diagrams for selumetinib and vemurafenib. Selumetinib was ranked in position one in the BRAF V600E screen and vemurafenib was ranked at position two (Table 4). Selumetinib is currently in clinical trials and vemurafenib has been approved for the treatment of melanoma with BRAF V600E mutations (Bollag et al., 2012)

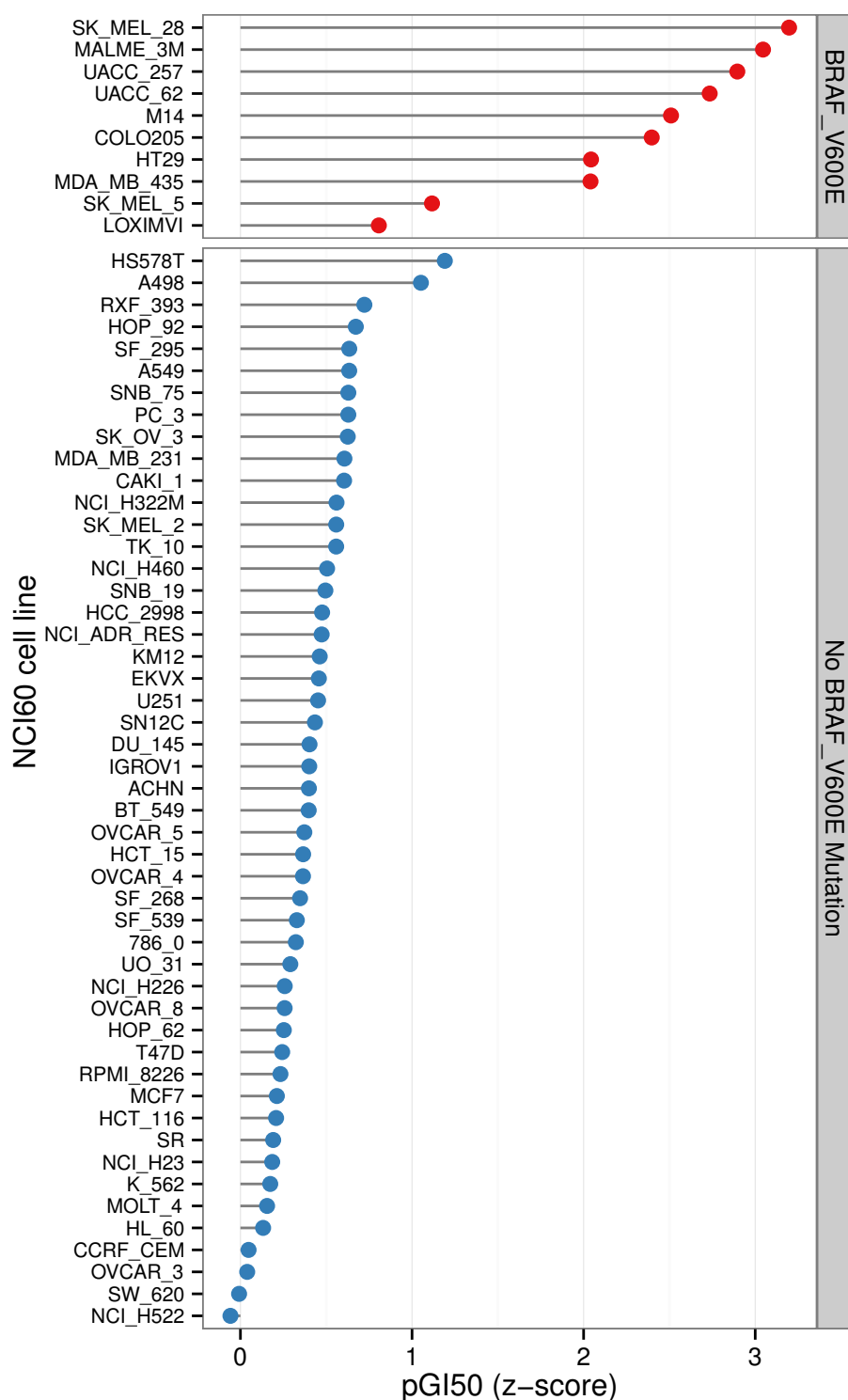


Figure 2: The response of individual NCI60 cell lines to vemurafenib (NSC 761431). The response of each cell line is sorted in descending order with respect to cell lines with and without the V600E mutation.

Staying within the top 50 ranks, the next highest-ranked molecule annotated as being either FDA approved or in a clinical trial is hypothemycin (NSC 354462) at rank 4, and after that 17-amino-17-demethoxygeldanamycin (NSC 255109) at rank 35. Hypothemycin has been reported to irreversibly inhibit the MEK and ERK proteins (Fukazawa et al., 2010). 17-Amino-17-demethoxygeldanamycin has been reported as an inhibitor of heat shock protein 90 (hsp90) (Ge et al., 2006). It appears that the mechanism of action of the high-ranked molecules selumetinib, vemurafenib, and hypothemycin is a result of each of these molecules interacting with one or more components of the RAS \rightarrow RAF \rightarrow MEK \rightarrow ERK pathway.

Table 5 shows the top 10 ranked molecules with respect to the KRAS Diff_{mut} value. In contrast to Table 4 the Diff_{mut} values are lower with the highest value 0.90 less than half of the highest BRAF V600E Diff_{mut} value of 2.45. Also the highest ranked KRAS Diff_{mut} values are less statistically significant than the highest ranked BRAF V600E Diff_{mut} values with higher p-values and confidence intervals at a lower range.

Rank	NSC ID	Name	FDA Status	Diff _{mut}	p-value	C.I.
1	751605	-	-	0.90	0.058	-0.2-2.0
2	726973	-	-	0.89	0.001	0.3-1.4
3	63878	Cytarabine	Approved	0.86	0.018	0.1-1.6
4	695267	-	-	0.86	0.049	-0.2-2.0
5	666783	-	-	0.84	0.002	0.4-1.3
6	692745	-	-	0.83	0.003	0.3-1.4
7	671554	-	-	0.83	0.001	0.1-1.8
8	613327	Gemcitabine	Approved	0.82	0.038	0.1-1.4
9	668297	-	-	0.81	0.007	0.5-1.2
10	696558	-	-	0.80	0.016	0.1-1.5

Table 5: Molecules ranked by descending KRAS difference. The confidence interval (C.I.) is at the 95% level.

Figure 3 shows the response of the third ranked molecule, cytarabine, to the NCI60 cell lines. This plot indicates that the growth of cell lines with a KRAS mutation are inhibited more often by cytarabine (NSC 63878) than cell lines that lack a KRAS mutation. This effect is clearly not as pronounced as that of vemurafenib with respect to cell lines with a BRAF V600E mutation.

In Chapter 21 of the Twelfth Edition of Goodman and Gilman’s “The Pharmacological Basis of Therapeutics” it is noted that “20% of AML patients have leukemic cells with a k-RAS mutation, and these patients seem to derive greater benefit from high dose Ara-C regimens than do patients with wild-type k-RAS.” This comment is based on a study by Neubauer et al. (Goodman et al., 2011; Neubauer et al., 2008). “Ara-C” is a synonym for cytarabine, AML stands for acute myeloid leukemia and k-RAS is another way of writing KRAS. This observation seems to be consistent with the result that cytarabine seems to preferentially inhibit the NCI60 cell lines with KRAS mutations. Neubauer et al. cite work by Koo et al that shows that cytarabine helps to induce apoptosis in a cell with a RAS oncogene through some mechanism involving the S-phase checkpoint (Koo et al., 1999). It is not clear to us that the molecular details of how this apoptosis takes place are known.

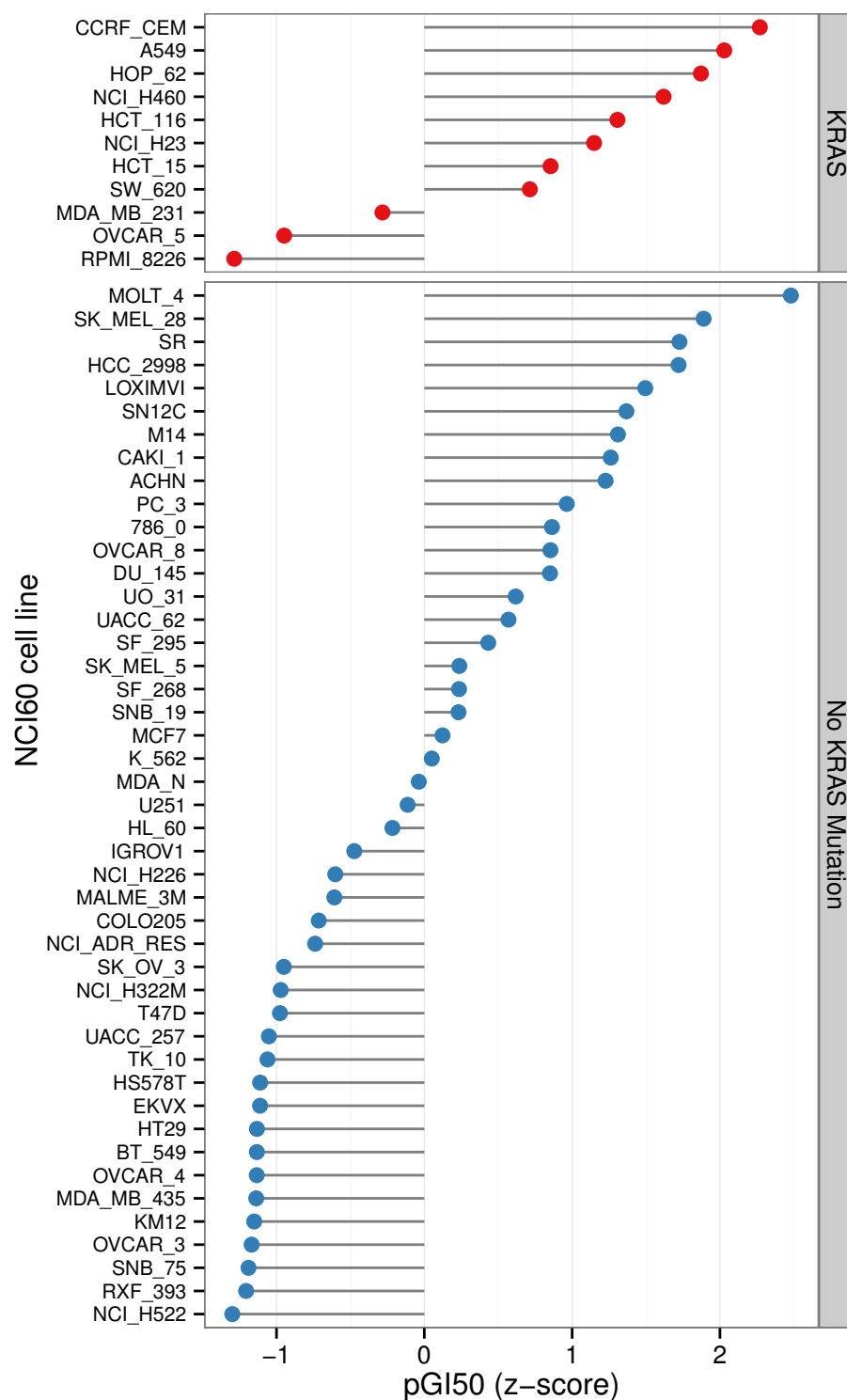


Figure 3: The response of individual NCI60 cell lines to cytarabine (NSC 63878). The response of each cell line is sorted in descending order with respect to cell lines with and without the KRAS mutations at codons 12, 13, or 61.

Several of the top ranked molecules with respect to KRAS are cytidine analogs which are shown in Figure 4. This figure shows four such derivatives in the top fifty ranked molecules with three appearing in the top ten. Cytarabine and gemcitabine are approved cancer chemotherapy drugs. These drugs are classified as antimetabolites that interfere with DNA synthesis in cancer cells (Trevor et al., 2009). Wang et al., in a review of small molecules that modulate the activity of mutant KRAS, note that several cytidine analogs were found to selectively kill cell lines with KRAS mutations. They also note that the “exact target at a molecular level” of these compounds is unknown with respect to targeting cell lines with mutant KRAS (Wang et al., 2013).

It is not clear to us why cytidine analogs might preferentially inhibit the growth of cell lines with KRAS mutations. Possibly this is described in the scientific literature but does not seem to be explained in two recent review of KRAS inhibitors (Wang et al., 2013, 2012). It seems to be an interesting coincidence that cytidine analogs should inhibit the growth of cell lines with mutant KRAS which contain the cytosine base pair guanosine in the form of guanosine triphosphate (GTP). GTP binds with femtomolar affinity to KRAS, and its binding stabilizes the active form of KRAS. Guanosine diphosphate stabilizes the inactive form of KRAS (Wang et al., 2012). We speculate that perhaps the cytidine analogs influence the conformation of KRAS in some way that affects function through a molecular interaction with the guanosine of GTP. However we know of no experimental evidence to support this idea.

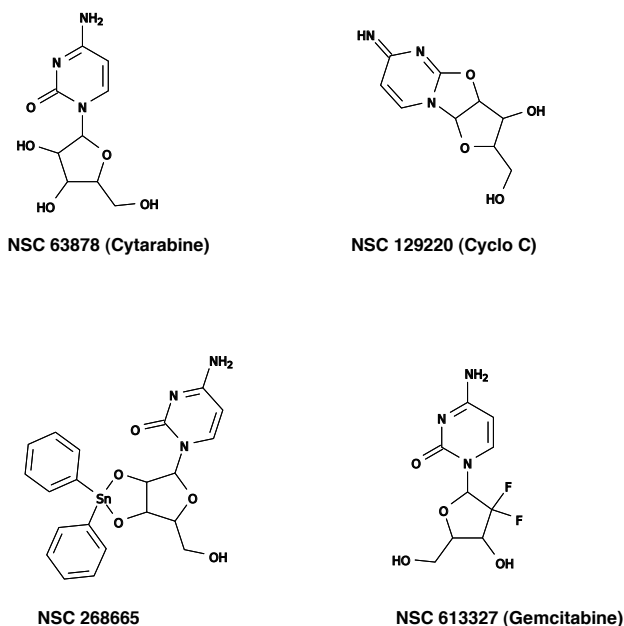


Figure 4: Several cytidine derivatives which are ranked highly with respect to the KRAS Diff_{mut} value. NSC 63878 (cytarabine) was ranked at position 3, NSC 613327 (gemcitabine) at position 8, NSC 268665 at position 19, and NSC 129220 at position 22.

Figure 5 shows a crystal structure of a G12D HRAS mutant (PDB ID-1AGP) showing a GTP analog, guanosine 5'-(beta,gamma-imido)triphosphate, bound. This structure shows that the guanosine part of the molecule appears to be buried in a cavity but is partly exposed

and perhaps could interact with another molecule such as a cytidine analog as we speculate. Since HRAS and KRAS are very similar proteins (Schubbert et al., 2007) it seems quite possible that GTP binds to KRAS in a similar manner.

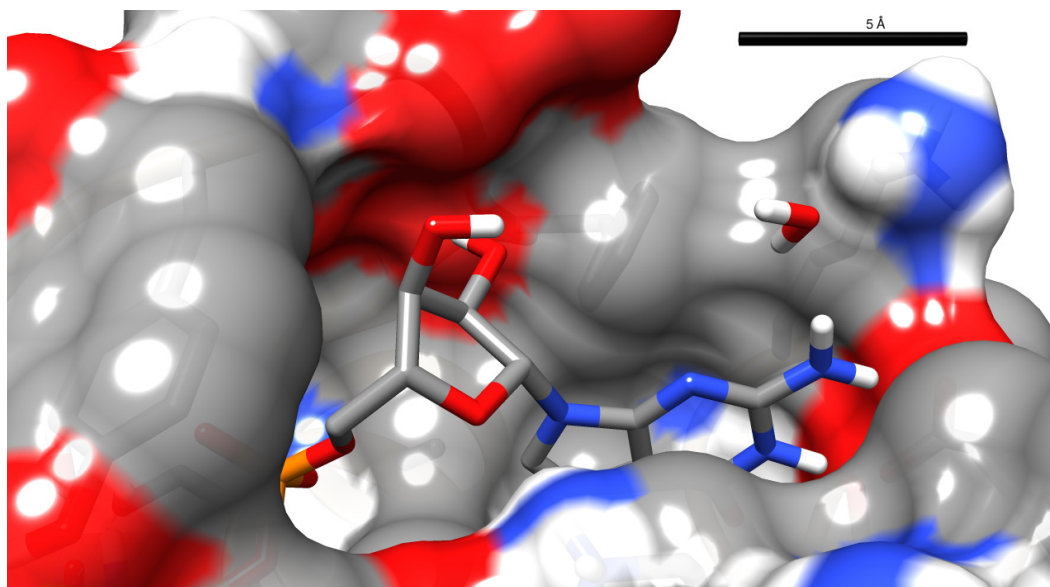


Figure 5: A view of the crystal structure (PDB ID-1AGP) of the G12D HRAS mutant is shown with a GTP analog, guanosine 5'-(beta,gamma-imido)triphosphate, bound. This GTP analog appears to be bound in such a way in which its guanosine part may be able to interact with other molecules at the surface of the protein. Colors represent the atomic elements, carbon (grey), hydrogen (white), nitrogen (blue), oxygen (red), and phosphorus (orange).

The conjecture that cytosine derivatives interact with KRAS bound to either GTP or GDP is a highly speculative conjecture and most likely wrong given the lack of direct evidence. However the question of why cytosine analogs appear to preferentially inhibit the growth of cell lines with KRAS mutations does seem an interesting one given the importance of KRAS as a potential drug target.

The main hypothesis of this study was that there were molecules that had been screened in the NCI60 cell line panel that selectively inhibited the growth of cell lines with a particular mutation more than those that lacked that mutation. We looked at two oncogenes, mutated BRAF and mutated KRAS and used a simple method to rank screened molecules in order of how much those molecules inhibited the growth of cell lines with a particular mutation.

In the case of BRAF highly ranked molecules included selumetinib, vemurafenib and hypothemycin. All three of these molecules inhibited proteins in the RAS → RAF → MEK → ERK pathway which indicates that the ranking was sensible. Vemurafenib was designed to inhibit BRAF with the V600E mutation and has been approved by the US FDA to treat melanoma with this specific mutation. We expected that the other approved BRAF V600E inhibitor dabrafenib would also rank highly. However the screening results for this molecule were not in the 2012 NCI60 data set.

In the case of KRAS highly ranked molecules included a number of drugs that were approved and in clinical trials for the treatment of cancer. The highest ranked molecules

for KRAS has lower Diff_{mut} values than those for BRAF and these values also had higher p-values so it is not clear whether the molecules that were highly ranked really selectively inhibited cell lines with KRAS mutations. However it seems that it might be worthwhile to investigate some of these highly ranked molecules. Several of these molecules are cytidine derivatives. It is quite possible that one cytidine derivative might have been highly ranked by chance but less likely that several different cytidines would be highly ranked by chance. Therefore it seems worthwhile to look into why such molecules would selectively inhibit the growth of cell lines with KRAS mutations.

The NCI60 screen is a phenotypic small-molecule screen, a screen in which some amount of a small-molecule is added to a cell or organism and observed to see if there is a change in phenotype (Eggert, 2013). Eggert mentions some potential advantages of this approach. One potential advantage cited is that phenotypic screens can potentially “target any protein (or other entity, such as lipid or nucleic acid) in its biological context” (Eggert, 2013). This suggests that phenotypic screens are relatively unbiased by not requiring the researcher to focus on particular targets in the cell. In this paper we have proceeded to look at molecules in the NCI60 screen that appear to inhibit the growth of cell lines with particular oncogene mutations. While this is a bias for a particular oncogene it is unbiased with regard to the mechanism by which the oncogene is targeted. For example in the case of the BRAF V600E screen three molecules that were highly ranked, vemurafenib, selumetinib, and hypothemycin, are indicated by the literature to interact with different protein components of the $\text{RAS} \rightarrow \text{RAF} \rightarrow \text{MEK} \rightarrow \text{ERK}$ pathway as described above.

There are limitations for using cancer cell lines as a model for human cancer pharmacology because of the fact that cancer cell lines that are able to grow in culture are not necessarily representative of cancer cells in vivo. In spite of this such cell lines may still be useful for drug discovery as long as these limitations are recognized (Weinstein, 2012; Weinberg, 2014b).

Related Work

The work described in this paper has similarities to previous work by Ikediobi (Ikediobi, 2008) and Meltzer et al. (Abaan et al., 2013). These two studies are similar because they are also based on the NCI60 data set and look at the cancer-related gene mutations within the cell lines. In particular both studies ranked screened molecules in descending order of the difference between the mean GI50s of mutated and wild-type genes in the same way that we have described. A difference is that these studies did not normalize the GI50s as we have done, something that we suppose will not make a big difference to the results.

Ikediobi looked at “relationships between mutations in cancer genes and drug activity in the NCI-60 Cell lines.” Ikediobi computed the difference between the means of mutant and wild type cell lines in the NCI-60 for selected oncogene mutations for screened small molecules. These mutations were sequenced by Ikediobi et al and are the same ones we are using. (Ikediobi et al., 2006). The p-values for each small molecule tested were estimated using the Wilcoxon rank-sum test.

The most significant (p-values < 0.05) drug-gene relationships were found for BRAF and CDKN2A. The KRAS drug-gene relationship was not found to be significant, something that is in contrast to what we found using the same data and therefore requires more investigation on our part. For example it would be interesting to calculate the p-value using the Wilcoxon

method on our highest ranked KRAS inhibitors.

Ikediobi showed the ten most significant compounds associated with the CDKN2A mutations. An interesting coincidence is that cytarabine (NSC 63878) appears in this list as well as another cytosine derivative aracytidine 5'-phosphate (NSC 99445).

In the case of molecules that preferentially inhibit cell lines with BRAF mutations, several of the highest ranked molecules from the Ikediobi study were found in our highest ranked list such as NSC 706829 and NSC 354462, both MEK inhibitors according to Ikediobi (Ikediobi, 2008).

Meltzer et al. describe a study in which they sequenced the exomes of the NCI60 cell lines and reported on the mutations found in genes and the screened small molecules associated with inhibiting activity in cell lines with particular gene mutations. In particular they report small molecules with statistically significant large mean \log_{10} GI50s between mutant and wild-type BRAF. (The method they used to calculate statistical significance was not clear to us from our reading of the paper). They reported that selumetinib (NSC 741078), vemurafenib (NSC 761431), and hypothemycin (NSC 354462) were highly ranked which agrees with our results as shown in Table 4.

Further Work

The estimates of Diff_{mut} may overestimate the “true” values because of the problem that when looking at a large number of values some may be high because of chance random noise. An “empirical Bayes” method suggested by Efron may be a good way of adjusting the values to lower the contribution from random noise (Efron, 2013). We plan to investigate implementing this method on order to improve the quality of the calculations.

Conclusions

Our results indicate that this method gives useful estimates for how much a screened small molecule preferentially inhibits the growth of cancer cell lines with a particular KRAS or BRAF cancer-related mutation. We defined a quantity, Diff_{mut} , that estimates how much more a given small molecule inhibits cell lines with a mutation of interest than cell lines without that mutation. We ranked the small molecules in descending order of Diff_{mut} and then tried to explain the ranking of the highest ranked molecules in terms of independent facts about those molecules.

In the BRAF ranking the highest ranked molecules included vemurafenib and selumetinib. Vemurafenib is a drug designed and approved to treat melanoma with the BRAF V600E mutation. Selumetinib is a MEK 1/2 inhibitor and MEK 1/2 in a protein which shares a signaling pathway with BRAF. Therefore these rankings seem to make sense with respect to BRAF.

With the KRAS rankings the highest ranked molecules included cytarabine, gemcitabine, and other cytidine analogs. Cytarabine is an approved chemotherapy drug that has been observed to be of some benefit to AML patients with a KRAS mutation (Goodman et al., 2011; Neubauer et al., 2008). In addition Wang et al. note that several cytidine analogs were found to kill cell lines with KRAS mutations (Wang et al., 2013). Therefore it seems that there is some evidence to suggest that the high ranking of these cytidine analogs concords with results independent of the NCI60 data.

Literature research of the top hits gave clues as to the mechanism of action of the growth inhibition. In general there may be no relevant information in the literature on a particular highly ranked compound and understanding the mechanism of action would require experimental investigation. However perhaps ranking would help to focus attention on likely relevant compounds.

Acknowledgements

We are grateful for a generous credit from Google, Inc. for the use of the Google Compute Engine Service (Google, 2012).

References

- Abaan, O. D., Polley, E. C., Davis, S. R., Zhu, Y. J., Bilke, S., Walker, R. L., Pineda, M., Gindin, Y., Jiang, Y., Reinhold, W. C., Holbeck, S. L., Simon, R. M., Doroshow, J. H., Pommier, Y., and Meltzer, P. S. (2013). The exomes of the NCI-60 panel: A genomic resource for cancer biology and systems pharmacology. *Cancer Research*, 73(14):4372–4382.
- Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K., and Hirth, P. (2012). Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nature Reviews Drug Discovery*, 11(11):873–886.

- Bolton, E. E., Wang, Y., Thiessen, P. A., and Bryant, S. H. (2008). PubChem: integrated platform of small molecules and biological activities. *Annual reports in computational chemistry*, 4:217–241.
- Bostock, M. (2013). Why use make. <http://bost.ocks.org/mike/make/>.
- CellMiner (2012). Cellminer web application. <http://discover.nci.nih.gov/cellminer/>. [Online; accessed 4-September-2013].
- COSMIC (2013). COSMIC cell lines project. ftp://ftp.sanger.ac.uk/pub/CGP/cell_lines_project/data_export/CosmicCellLineProject_v66_250713.tsv. [Online; accessed 4-September-2013].
- Dahl, R. L. (2014). Node.js. nodejs.org. Version 0.10.28/May 1, 2014.
- Diaconis, P. and Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, 248(5):116–130.
- Efron, B. (2013). A 250-year argument: belief, behavior, and the bootstrap. *Bulletin of the American Mathematical Society*, 50(1):129–146.
- Eggert, U. S. (2013). The why and how of phenotypic small-molecule screens. *Nature chemical biology*, 9(4):206–209.
- Feldman, S. (2014). Gnu make. <http://www.gnu.org/software/make/>.
- Forbes, S., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J., Futreal, P., and Stratton, M. (2008). The catalogue of somatic mutations in cancer (cosmic). *Current protocols in human genetics*, pages 10–11.
- Fukazawa, H., Ikeda, Y., Fukuyama, M., Suzuki, T., Hori, H., Okuda, T., and Uehara, Y. (2010). The resorcylic acid actone hypothemycin selectively inhibits the mitogen-activated protein kinase kinase-extracellular signal-regulated kinase pathway in cells. *Biological and Pharmaceutical Bulletin*, 33(2):168–173.
- Ge, J., Normant, E., Porter, J. R., Ali, J. A., Dembski, M. S., Gao, Y., Georges, A. T., Grenier, L., Pak, R. H., Patterson, J., et al. (2006). Design, synthesis, and biological evaluation of hydroquinone derivatives of 17-amino-17-demethoxygeldanamycin as potent, water-soluble inhibitors of Hsp90. *Journal of medicinal chemistry*, 49(15):4606–4615.
- Good, P. (2005). *Permutation, parametric and bootstrap tests of hypotheses*. Springer.
- Goodman, L. S. et al. (2011). *Goodman and Gilman's the pharmacological basis of therapeutics*. McGraw-Hill New York, 12th edition.
- Google (2012). Google compute engine. <https://cloud.google.com/products/compute-engine/>.
- Gray, K. A., Daugherty, L. C., Gordon, S. M., Seal, R. L., Wright, M. W., and Bruford, E. A. (2013). Genenames.org: the hgnc resources in 2013. *Nucleic Acids Research*, 41(D1):D545.

- Ikediobi, O. N. (2008). *Characterization of Cancer Gene Mutations in Human Cancer Cell lines for Correlation with Drug Activity*. PhD thesis, University of Cambridge.
- Ikediobi, O. N., Davies, H., Bignell, G., Edkins, S., Stevens, C., O'Meara, S., Santarius, T., Avis, T., Barthorpe, S., Brackenbury, L., et al. (2006). Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Molecular cancer therapeutics*, 5(11):2606–2612.
- Koo, H.-M., McWilliams, M. J., Alvord, W. G., and Woude, G. F. V. (1999). Ras oncogene-induced sensitization to 1- β -d-arabinofuranosylcytosine. *Cancer research*, 59(24):6057–6062.
- Kreyszig, E. (1999). *Advanced engineering mathematics*. Wiley.com, 8th edition.
- Namba, H., Nakashima, M., Hayashi, T., Hayashida, N., Maeda, S., Rogounovitch, T. I., Ohtsuru, A., Saenko, V. A., Kanematsu, T., and Yamashita, S. (2003). Clinical implication of hot spot BRAF mutation, V599E, in papillary thyroid cancers. *Journal of Clinical Endocrinology & Metabolism*, 88(9):4393–4397.
- NCI-DTP (2012). DTP screening results. http://dtp.nci.nih.gov/docs/cancer/cancer_data.html. [Online; accessed 17-December-2012].
- Neubauer, A., Maharry, K., Mrózek, K., Thiede, C., Marcucci, G., Paschka, P., Mayer, R. J., Larson, R. A., Liu, E. T., and Bloomfield, C. D. (2008). Patients with acute myeloid leukemia and ras mutations benefit most from postremission high-dose cytarabine: a cancer and leukemia group b study. *Journal of Clinical Oncology*, 26(28):4603–4609.
- Nguyen, K. (2013). Market watch: upcoming market catalysts in Q2 2013. *Nature Reviews Drug Discovery*, 12(4):254–254.
- OBoyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):1–14.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reinhold, W. C., Sunshine, M., Liu, H., Varma, S., Kohn, K. W., Morris, J., Doroshow, J., and Pommier, Y. (2012). Cellminer: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer research*, 72(14):3499–3511.
- Schubbert, S., Shannon, K., and Bollag, G. (2007). Hyperactive Ras in developmental disorders and cancer. *Nature Reviews Cancer*, 7(4):295–308.
- Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer*, 6(10):813–823.

- Stephen, A. G., Esposito, D., Bagni, R. K., and McCormick, F. (2014). Dragging Ras back in the ring. *Cancer cell*, 25(3):272–281.
- Thompson, H. (2013). US National Cancer Institute’s new Ras project targets an old foe. *Nature medicine*, 19(8):949–950.
- Torvalds, L. (2005). Git. <http://git-scm.com>.
- Trevor, A. J., Katzung, B. G., and Masters, S. B. (2009). *Basic & clinical pharmacology*. McGraw-Hill Medical.
- Troiani, T., Vecchione, L., Martinelli, E., Capasso, A., Costantino, S., Ciuffreda, L., Morgillo, F., Vitagliano, D., D’Aiuto, E., De Palma, R., et al. (2012). Intrinsic resistance to selumetinib, a selective inhibitor of MEK1/2, by cAMP-dependent protein kinase A activation in human lung and colorectal cancer cells. *British journal of cancer*, 106(10):1648–1659.
- Wang, W., Fang, G., and Rudolph, J. (2012). Ras inhibition via direct Ras binding—is there a path forward? *Bioorganic & medicinal chemistry letters*.
- Wang, Y., Kaiser, C. E., Frett, B., and Li, H.-y. (2013). Targeting mutant KRAS for anticancer therapeutics: A review of novel small molecule modulators. *Journal of Medicinal Chemistry*, 56(13):5219–5230.
- Weinberg, R. A. (2014a). *The biology of cancer*, chapter Cytoplasmic Signaling Circuitry Programs Many of the Traits of Cancer. Garland Science New York.
- Weinberg, R. A. (2014b). *The biology of cancer*, chapter The Rational Treatment of Cancer. Garland Science New York.
- Weinberg, R. A. (2014c). *The biology of cancer*, chapter Growth Factors, Receptors, and, Cancer. Garland Science New York.
- Weinstein, J. N. (2012). Drug discovery: Cell lines battle cancer. *Nature*, 483(7391):544–545.