

12-5-2018

Canvass: A Crowd-Sourced, Natural-Product Screening Library for Exploring Biological Space

Sara E. Kearney

National Center for Advancing Translational Sciences

Gergely Zahoránszky-Kóhalmi

National Center for Advancing Translational Sciences

Kyle R. Brimacombe

National Center for Advancing Translational Sciences

Mark J. Henderson

National Center for Advancing Translational Sciences


Caitlin Lynch

National Center for Advancing Translational Sciences

See next page for additional authors

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Follow this and additional works at: https://uknowledge.uky.edu/chemistry_facpub

 Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#), [Biology Commons](#), [Chemistry Commons](#), and the [Pharmacy and Pharmaceutical Sciences Commons](#)

Repository Citation

Kearney, Sara E.; Zahoránszky-Kóhalmi, Gergely; Brimacombe, Kyle R.; Henderson, Mark J.; Lynch, Caitlin; Zhao, Tongan; Wan, Kanny K.; Itkin, Zina; Dillon, Christopher; Shen, Min; Cheff, Dorian M.; Lee, Tobie D.; Bougie, Danielle; Cheng, Ken; Coussens, Nathan P.; Dorjsuren, Dorjbal; Eastman, Richard T.; Huang, Ruili; Iannotti, Michael J.; Karavadhi, Surendra; Klumpp-Thomas, Carleen; Roth, Jacob S.; Sakamuru, Srilatha; Sun, Wei; Titus, Steven A.; Yasgar, Adam; Zhang, Ya-Qin; Zhao, Jinghua; Andrade, Rodrigo B.; Brown, M. Kevin; and Grossman, Robert B., "Canvass: A Crowd-Sourced, Natural-Product Screening Library for Exploring Biological Space" (2018). *Chemistry Faculty Publications*. 145.

https://uknowledge.uky.edu/chemistry_facpub/145

Authors

Sara E. Kearney, Gergely Zahoránszky-Kóhalmi, Kyle R. Brimacombe, Mark J. Henderson, Caitlin Lynch, Tongan Zhao, Kanny K. Wan, Zina Itkin, Christopher Dillon, Min Shen, Dorian M. Cheff, Tobie D. Lee, Danielle Bougie, Ken Cheng, Nathan P. Coussens, Dorjbal Dorjsuren, Richard T. Eastman, Ruili Huang, Michael J. Iannotti, Surendra Karavadhi, Carleen Klumpp-Thomas, Jacob S. Roth, Srilatha Sakamuru, Wei Sun, Steven A. Titus, Adam Yasgar, Ya-Qin Zhang, Jinghua Zhao, Rodrigo B. Andrade, M. Kevin Brown, and Robert B. Grossman

Canvass: A Crowd-Sourced, Natural-Product Screening Library for Exploring Biological Space**Notes/Citation Information**

Published in *ACS Central Science*, v. 4, issue 12, p. 1727-1741.

Copyright © 2018 American Chemical Society

This is an open access article published under an [ACS AuthorChoice License](#), which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

Due to the large number of authors, only the first 30 and the authors affiliated with the University of Kentucky are listed in the author section above. For the complete list of authors, please download this article or visit: <https://doi.org/10.1021/acscentsci.8b00747>

Digital Object Identifier (DOI)

<https://doi.org/10.1021/acscentsci.8b00747>



Canvass: A Crowd-Sourced, Natural-Product Screening Library for Exploring Biological Space

Sara E. Kearney,^{a,†} Gergely Zahoránszky-Kóhalmi,^{a,†} Kyle R. Brimacombe,^a Mark J. Henderson,^a Caitlin Lynch,^a Tongan Zhao,^a Kanny K. Wan,^{a,b,ib} Zina Itkin,^a Christopher Dillon,^a Min Shen,^a Dorian M. Cheff,^a Tobie D. Lee,^a Danielle Bougie,^a Ken Cheng,^a Nathan P. Coussens,^a Dorjbal Dorjsuren,^a Richard T. Eastman,^a Ruili Huang,^a Michael J. Iannotti,^a Surendra Karavadhi,^a Carleen Klumpp-Thomas,^a Jacob S. Roth,^a Srilatha Sakamuru,^a Wei Sun,^a Steven A. Titus,^a Adam Yasgar,^a Ya-Qin Zhang,^a Jinghua Zhao,^a Rodrigo B. Andrade,^{c,ib} M. Kevin Brown,^{d,ib} Noah Z. Burns,^{e,ib} Jin K. Cha,^{f,ib} Emily E. Mevers,^g Jon Clardy,^{g,ib} Jason A. Clement,^h Peter A. Crooks,ⁱ Gregory D. Cuny,^j Jake Ganor,^k Jesus Moreno,^l Lucas A. Morrill,^l Elias Picazo,^l Robert B. Susick,^l Neil K. Garg,^{l,ib} Brian C. Goess,^m Robert B. Grossman,^{n,ib} Chambers C. Hughes,^{o,ib} Jeffrey N. Johnston,^{p,ib} Madeleine M. Joulie,^q A. Douglas Kinghorn,^{r,ib} David G.I. Kingston,^{s,ib} Michael J. Krische,^t Ohyun Kwon,^{l,ib} Thomas J. Maimone,^{u,ib} Susruta Majumdar,^{v,w} Katherine N. Maloney,^x Enas Mohamed,^y Brian T. Murphy,^{z,ib} Pavel Nagorny,^{A,ib} David E. Olson,^{B,C,D,ib} Larry E. Overman,^{E,ib} Lauren E. Brown,^{F,ib} John K. Snyder,^F John A. Porco, Jr.,^{F,ib} Fatima Rivas,^G Samir A. Ross,^{y,ib} Richmond Sarpong,^{H,ib} Indrajeet Sharma,^{I,ib} Jared T. Shaw,^{B,ib} Zhengren Xu,^J Ben Shen,^{J,ib} Wei Shi,^{K,ib} Corey R.J. Stephenson,^{A,ib} Alyssa L. Verano,^L Derek S. Tan,^{L,M,ib} Yi Tang,^{I,ib} Richard E. Taylor,^N Regan J. Thomson,^{O,ib} David A. Vosburg,^{b,ib} Jimmy Wu,^{P,ib} William M. Wuest,^{Q,R} Armen Zakarian,^{S,ib} Yufeng Zhang,^T Tianjing Ren,^T Zhong Zuo,^{T,ib} James Inglese,^{a,ib} Sam Michael,^a Anton Simeonov,^a Wei Zheng,^a Paul Shinn,^a Ajit Jadhav,^a Matthew B. Boxer,^{a,U} Matthew D. Hall,^{*,a,ib} Menghang Xia,^a Rajarshi Guha,^{a,V} and Jason M. Rohde^{*,a,W,ib}

^aNational Center for Advancing Translational Sciences, National Institutes of Health, 9800 Medical Center Drive, Rockville, Maryland 20850, United States

^bDepartment of Chemistry, Harvey Mudd College, 301 Platt Boulevard, Claremont, California 91711, United States

^cDepartment of Chemistry, Temple University, 1901 North 13th Street, Philadelphia, Pennsylvania 19122, United States

^dDepartment of Chemistry, Indiana University, 800 East Kirkwood Avenue, Bloomington, Indiana 47405, United States

^eDepartment of Chemistry, Stanford University, 333 Campus Drive, Stanford, California 94305, United States

^fDepartment of Chemistry, Wayne State University, 5101 Cass Avenue, Detroit, Michigan 48202, United States

^gDepartment of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, United States

^hNatural Products Discovery Institute, Baruch S. Blumberg Institute, 3805 Old Easton Road, Doylestown, Pennsylvania 18902, United States

ⁱUniversity of Arkansas for Medical Sciences, 4301 West Markham Street 522, Little Rock, Arkansas 72205, United States

^jDepartment of Pharmacological and Pharmaceutical Sciences, University of Houston, 4849 Calhoun Road, Houston, Texas 77204, United States

^kDiamond Age Corp., 344 East Louisiana Street, McKinney, Texas 75069, United States

^lDepartment of Chemistry and Biochemistry, UCLA, 607 Charles E. Young Drive East, Los Angeles, California 90095, United States

^mDepartment of Chemistry, Furman University, 3300 Poinsett Highway, Greenville, South Carolina 29613, United States

ⁿDepartment of Chemistry, University of Kentucky, Lexington, Kentucky 40506, United States

^oScripps Institution of Oceanography, UCSD, 9500 Gilman Drive, La Jolla, California 92093, United States

^pDepartment of Chemistry, Vanderbilt University, 7330 Stevenson Center, Nashville, Tennessee 37235, United States

^qDepartment of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104, United States

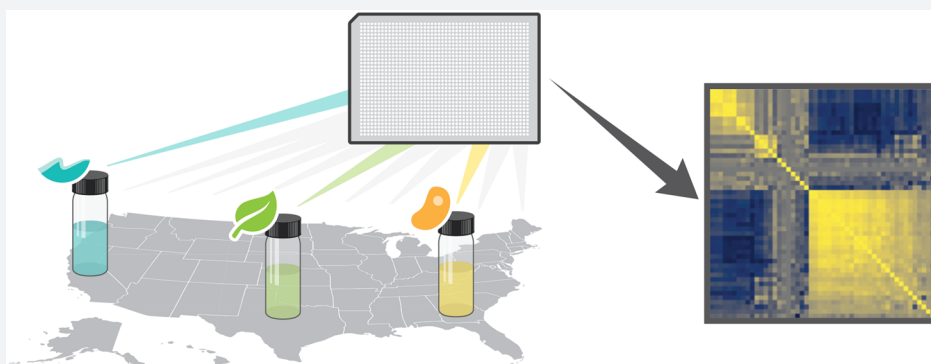
^rCollege of Pharmacy, The Ohio State University, 500 West 12th Avenue, Columbus, Ohio 43210, United States

Received: October 12, 2018

Published: December 5, 2018

- ^sDepartment of Chemistry, Virginia Tech, 900 West Campus Drive, Blacksburg, Virginia 24061, United States
- ^tChemistry Department, The University of Texas at Austin, 105 East 24th Street STOP A5300, Austin, Texas 78712, United States
- ^uDepartment of Chemistry, University of California Berkeley, 826 Latimer Hall, Berkeley, California 94720, United States
- ^vDepartment of Molecular Pharmacology and Neurology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, United States
- ^wCenter for Clinical Pharmacology, St Louis College of Pharmacy and Washington University School of Medicine, 2 Pharmacy Place, St. Louis, Missouri 63110, United States
- ^xDepartment of Chemistry, Point Loma Nazarene University, 3900 Lomaland Drive, San Diego, California 92106, United States
- ^yUniversity of Mississippi School of Pharmacy, 2500 North State Street, Jackson, Mississippi 39216, United States
- ^zCollege of Pharmacy, Department of Medicinal Chemistry and Pharmacognosy, University of Illinois at Chicago, 900 South Ashland Avenue, Chicago, Illinois 60607, United States
- ^ADepartment of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109, United States
- ^BDepartment of Chemistry, University of California, Davis, One Shields Avenue, Davis, California 95616, United States
- ^CSchool of Medicine, Department of Biochemistry and Molecular Medicine, University of California, Davis, 2700 Stockton Boulevard, Suite 2102, Sacramento, California 95817, United States
- ^DCenter for Neuroscience, University of California, Davis, 1544 Newton Court, Davis, California 95618, United States
- ^EDepartment of Chemistry, University of California, Irvine, Irvine, California 92697, United States
- ^FDepartment of Chemistry and Center for Molecular Discovery (BU-CMD), Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, United States
- ^GDepartment of Chemical Biology and Therapeutics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, United States
- ^HDepartment of Chemistry, University of California Berkeley, 841-A Latimer Hall, Berkeley, California 94720, United States
- ^IDepartment of Chemistry and Biochemistry, and Institute of Natural Products and Research Technologies, University of Oklahoma, 101 Stephenson Parkway, Norman, Oklahoma 73019, United States
- ^JDepartment of Chemistry, Florida Campus, The Scripps Research Institute, 130 Scripps Way, Jupiter, Florida 33458, United States
- ^KDepartment of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, United States
- ^LPharmacology Graduate Program, Weill Cornell Graduate School of Medical Sciences, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, United States
- ^MChemical Biology Program, Sloan Kettering Institute and Tri-Institutional Research Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, United States
- ^NDepartment of Chemistry and Biochemistry and the Warren Family Research Center for Drug Discovery and Development, University of Notre Dame, 305 McCourtney Hall, Notre Dame, Indiana 46556, United States
- ^ODepartment of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United States
- ^PDepartment of Chemistry, Dartmouth College, Hanover, New Hampshire 03755, United States
- ^QDepartment of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, Georgia 30322, United States
- ^REmory Antibiotic Resistance Center, Emory University School of Medicine, 201 Dowman Drive, Atlanta, Georgia 30322, United States
- ^SSanta Barbara Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States
- ^TSchool of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Sha Tin, New Territories, Hong Kong SAR

S Supporting Information



continued...

ABSTRACT: Natural products and their derivatives continue to be wellsprings of nascent therapeutic potential. However, many laboratories have limited resources for biological evaluation, leaving their previously isolated or synthesized compounds largely or completely untested. To address this issue, the Canvass library of natural products was assembled, in collaboration with academic and industry researchers, for quantitative high-throughput screening (qHTS) across a diverse set of cell-based and biochemical assays. Characterization of the library in terms of physicochemical properties, structural diversity, and similarity to compounds in publicly available libraries indicates that the Canvass library contains many structural elements in common with approved drugs. The assay data generated were analyzed using a variety of quality control metrics, and the resultant assay profiles were explored using statistical methods, such as clustering and compound promiscuity analyses. Individual compounds were then sorted by structural class and activity profiles. Differential behavior based on these classifications, as well as noteworthy activities, are outlined herein. One such highlight is the activity of (–)-2(S)-cathafoline, which was found to stabilize calcium levels in the endoplasmic reticulum. The workflow described here illustrates a pilot effort to broadly survey the biological potential of natural products by utilizing the power of automation and high-throughput screening.

■ INTRODUCTION

Throughout history, nature has served as our primary source of medicines and continues to be one of the richest sources of new therapeutics. Either directly or as inspiration, natural products account for 50% to 70% of all small-molecule pharmaceutical agents currently in clinical use.¹ While their influence has been most profound in the treatment of infectious diseases and cancer, natural products have also found utility in other therapeutic areas such as pain, inflammation, and cardiovascular disorders. Yet, many pharmaceutical companies have diminished or abandoned natural products research throughout recent decades for a variety of reasons, ranging from the promise of emerging technologies (e.g., combinatorial chemistry), to concerns about international regulations of access to natural products and their sources.² In parallel, decreases in funding agency support for natural-products-related research have contributed to this contraction. However, within the same time period, natural compounds still continue to be both a significant source and point of inspiration for new medicines.¹

Despite this dichotomy, the pendulum is swinging back toward natural products within both industry and academia. New screening libraries are being designed to incorporate key features of natural products, including scaffold diversity and stereochemistry.² Strategic prefractionation methods have also facilitated high-throughput screening of natural product extracts.^{2,3} Genome mining with the goal of discovering “hidden” natural products within microbial genomes⁴ has fostered a great deal of excitement, and is the foundational approach of a number of pharmaceutical companies. The academic sector has also begun to see a recovery in the funding climate, as reflected in the creation of the Center for High-Throughput Functional Annotation of Natural Products (HiFAN).⁵ HiFAN is a collaborative, international, multi-institute center established to determine the mechanism of action of natural products and botanicals, with the intention of making platform technologies and data available community-wide. Together, these and other developments bode well for the renewed interest in nature as a rich resource for biologically relevant chemical matter.

While this resurgence has significant potential, especially to address the imminent threat of antibiotic resistance,⁶ we hypothesize that therapeutic opportunities for natural products across other disease indications have been underexplored. All too often, isolation scientists and synthetic chemists in academic laboratories isolate or synthesize natural products and test them against a single representative cancer cell line or bacterial strain, or in some instances, never test their compounds in any biological assay at all, missing out entirely on the potential to discover a valuable, new therapeutic. Do storage freezers in

laboratories engaged in natural product synthesis throughout the world contain the next advancements in human health? Toward realizing the potential of purified natural products, we established the Canvass natural product screening pilot initiative to provide the scientific community with a mechanism to evaluate the biological activities of natural products in a diversity of *in vitro* assays.

With the Canvass pilot program, we set out to crowd-source a diverse set of purified natural products by inviting academic investigators and companies to submit their natural products to the National Center for Advancing Translational Sciences (NCATS). Upon assembly, the Canvass library was compared to other relevant, well-studied chemical libraries. We then sought to broadly explore, or “canvass”, the library’s biological activity in an assortment of robust assays using quantitative high-throughput screening (qHTS).⁷ Due to the broad scope of disease-relevant mechanisms investigated by NCATS, we were able to screen the library against a wide range of assays. The resulting data set from 50 different assays was then systematically analyzed to identify overall trends and specific natural products with interesting biological activities. Project teams at NCATS further investigated the activities of several compounds using established workflows, and the full data set was made available through the Canvass website (<https://tripod.nih.gov/canvass>).

■ RESULTS

Canvass Library. The Canvass library of 346 natural products was assembled through a broad solicitation of both the academic and private sectors via the Canvass website. The pure (>85% purity by liquid chromatography/mass spectrometry [LC/MS]) compounds were submitted by 45 academic laboratories or companies around the world. We collected pure natural products, rather than natural product extracts, to circumvent deconvolution and structure elucidation due to time and resource limitations. The library was formatted into 1536-well plates and evaluated in 50 assays by qHTS in an 11-point concentration series.^{7,8} We manually classified⁹ the Canvass compounds using a set of 12 well-known structural classes, the distribution of which is summarized in Figure 1a.

Physicochemical Property Distributions. To ascertain the similarity of the Canvass library to existing drug collections, we analyzed and compared the structural features and physicochemical properties of the compound collection to publicly available compound collections known to contain druglike compounds or natural products. We first examined the physicochemical properties of the library in comparison to three well-known small-molecule libraries: the DrugBank Approved Drugs (2073 compounds, database version: 2.0.9),¹⁰ the ChEMBL¹¹

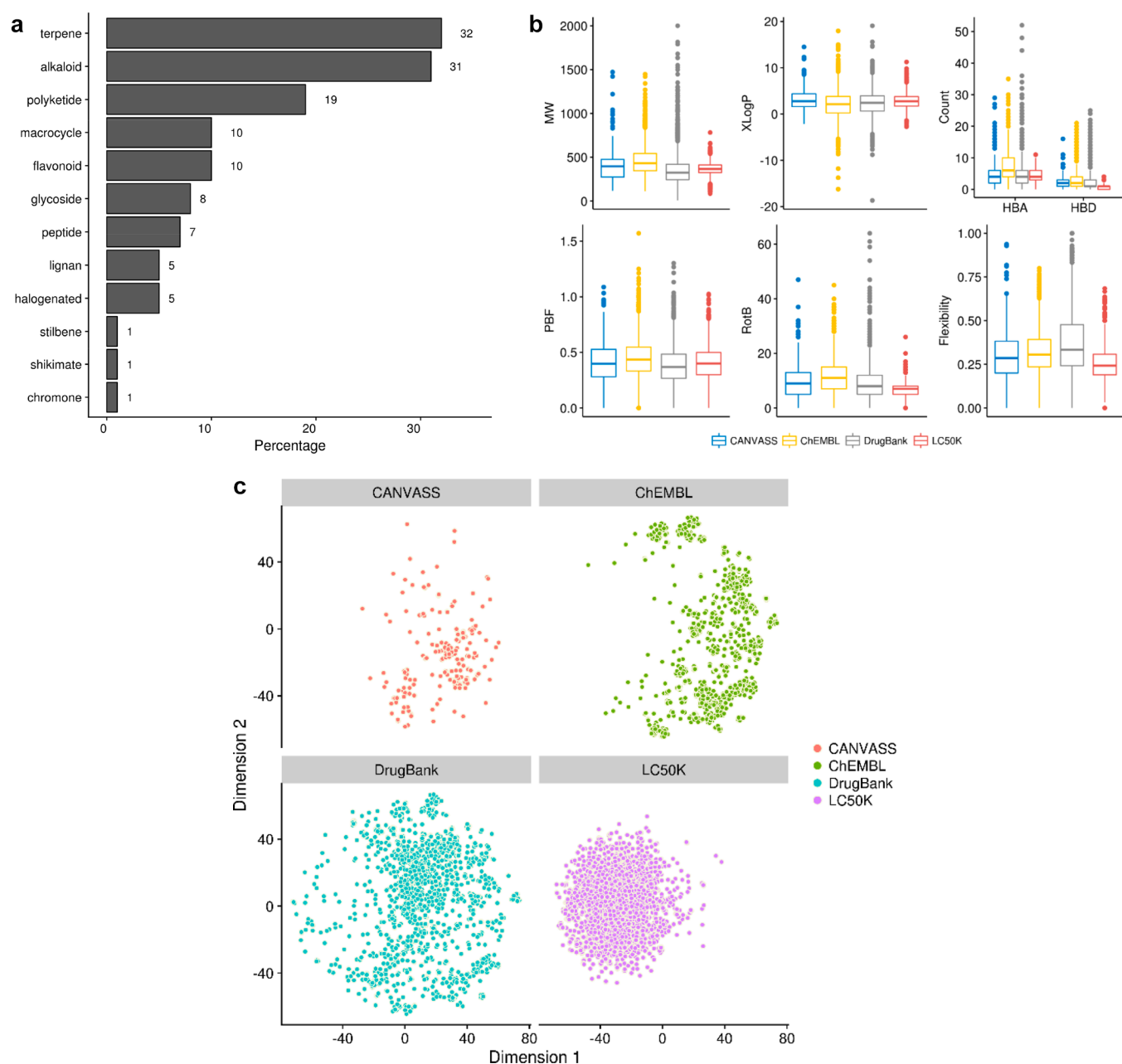


Figure 1. (a) Distribution of structural classes within the Canvass library. (b) Physicochemical properties of chemical libraries; MW = molecular weight, HBA = H-bond acceptor, HBD = H-bond donor, RotB = number of rotatable bonds, PBF = plane of best fit. (c) Chemical space overlap of the Canvass library with three other libraries in a 1024D fingerprint space reduced to two dimensions using tSNE. ECFP-6 fingerprints were computed using the CDK; tSNE = t-distributed stochastic neighbor embedding, ECFP-6 = extended connectivity fingerprint of diameter = 6, CDK = Chemistry Development Kit.

natural product set (1921 compounds, database version: ChEMBL 23), and a random subset of 3000 molecules from the Life Chemicals Diversity Set of 50K molecules (LC50K, 50 240 compounds). Selecting a subset of the LC50K library was necessary to reduce the dominance of the chemical space of such a large and diverse library. These libraries, spanned by compounds representing the entry-points and end-points of the drug discovery pipeline, helped us evaluate how the Canvass collection fits within drug discovery space. Specifically, we used the collections of ChEMBL natural products and LC50K molecules to represent the entry-points of drug discovery space. The ChEMBL collection generally represented natural products which historically have served as a rich source of drugs or starting points in

lead-optimization efforts, while the LC50K molecules exemplified the engineered libraries of diverse compounds characterized by desirable properties for lead-discovery purposes. Meanwhile, since the DrugBank set covers the majority of approved small-molecule drugs, it represented the end-points space.

We computed seven physicochemical properties: molecular weight (MW), H-bond acceptor (HBA) and H-bond donor (HBD) counts, XLogP,¹² the number of rotatable bonds (RotB), the plane of best fit (PBF),¹³ and fraction of rotatable bonds (flexibility).¹⁴ While the first five are relevant in a drug discovery setting, the PBF and flexibility descriptors characterize the three-dimensionality of the molecules. As shown by Meyers et al.,¹⁵ many synthetic scaffolds tend toward flatness, and there

has been increasing interest to enhance the three-dimensionality of molecules in screening libraries.¹⁶ Figure 1b summarizes the distribution of these properties for the Canvass collection versus those of the other libraries. While the medians of the properties are well-aligned between the Canvass and LC50K collections, it should be noted that Canvass compounds represent a more diverse physicochemical space. The accordance of median property values toward the ChEMBL and DrugBank collections shows a mixed picture. There is an almost perfect split in the number of cases where the Canvass physicochemical properties are more closely aligned with either the ChEMBL or DrugBank collections. Surprisingly, the distribution of PBF is similar between the four libraries.¹⁵

Chemical Space Overlap. We next examined the chemical space overlap of the Canvass library with the three comparator libraries. We considered two distinct chemical spaces: the physicochemical 7-dimensional descriptor space defined above, and a 1024-dimensional fingerprint space emphasizing structural features. In both cases, we computed a reduced 2-dimensional (2D) space using t-distributed stochastic neighbor embedding (tSNE).¹⁷ The results for the physicochemical space analysis are presented in the Supporting Information (see Figure S1). In the physicochemical space, the Canvass library is very similar to the other libraries, even though they may not be specifically natural-product-like, which is in line with the property distributions summarized in Figure 1b. We quantified the overlap between pairs of libraries using Thornton's separability index (*S*),¹⁸ resulting in the following values: 0.81, 0.82, and 0.89 for Canvass versus the ChEMBL natural products, DrugBank, and the LC50K subset, respectively. A larger index represents a larger separation in terms of likeness. These indices support previous observations regarding physicochemical property distributions that Canvass compounds are well-aligned with the other libraries in terms of their physicochemical properties. As expected, the closest set to Canvass in physicochemical space is the ChEMBL natural products set.

We then computed 1024-bit ECFP-6¹⁹ fingerprints using the Chemistry Development Kit (CDK)²⁰ for all compounds and examined the overlap in fingerprint space (Figure 1c). We observed that the embedding of Canvass compounds in this chemical space shows a resemblance to that of the ChEMBL natural products, as might be expected. Further, the chemical space occupied by Canvass and the DrugBank compounds shows a significant overlap. The quantitated overlap (using the separability index) reflects similar observations that we made regarding the physicochemical space, only, in this chemical space, Canvass overlaps to the highest degree with the DrugBank library (0.93), followed by ChEMBL natural product (0.95) and the LC50K (0.99) libraries. While this is somewhat unexpected, it may indicate that the Canvass library contains a number of structural elements in common with approved drugs.

Summary of the Assay Panel. The Canvass library was screened in qHTS format with 11-point dose–response against 50 assays covering a variety of readouts, modalities, and targets (either specific protein targets or biological processes) in both cell-based and biochemical assays. The bulk (33) of the assays focused on viability (e.g., cytotoxicity, cell proliferation, or membrane integrity), while 11 assays probed specific pathways (e.g., hypoxia-inducible factor 1- α [HIF1] signaling, or calcium modulation), and the remaining 6 assays were designed for specific biochemical targets (e.g., mutant isocitrate dehydrogenase 1 [mIDH1] or ATPase family AAA domain-containing protein 5 [ATAD5]). All cell-based assays were measured at a

single end-point, with the exception of the apoptosis assays using Caspase-Glo, which were measured at three time-points (12, 18, and 24 h). This screen generated over 210 000 data points. Though it is worth noting that, while we ran 50 individual assays, this number includes counter-screens associated with other assays. An example is the secreted endoplasmic reticulum calcium-monitoring proteins (SERCaMP) assay designed to detect endoplasmic reticulum (ER) calcium dysfunction.²¹ The primary assay identifies compounds that prevent depletion of the ER calcium store and is accompanied by secretion and viability counter-assays, each of which is designed to eliminate false positives from the primary assay.

A variety of quality control (QC) metrics²² were computed for each assay (focusing on the primary readout only) including *Z'*-factor,²³ signal-to-background (S/B), and the coefficient of variation (CV). QC measure values associated with plates of $Z' \leq -10$ were treated as outliers and accordingly excluded from the analysis. Figure 2a–c summarizes the *Z'*, S/B, and CV for all assays, grouped by their type (pathway, target, and viability). In general, assay performance²² was good ($0 < Z' \leq 0.5$) to excellent ($Z' > 0.5$) for all the assays in the panel, with a few exceptions. For instance, in the caspase-HEK293 apoptosis assay, the control compound (doxorubicin) did not elicit sufficient signal, which necessitated normalization using the maximum value from the sample wells. As a result, the *Z'* is not relevant for this particular assay. For assays with multiple readouts, we report only the median QC measure for the main readout. Except for the apoptosis assays, the viability assays tended to exhibit slightly better performance across all metrics than the other two assay classes (Figure 2d). A separate plot was made for better visibility of QC measure values for assays of $Z' \geq 0$ and of $SB < 60$ (see Figure S2).

Clustering Assays. Several features stand out from a pairwise correlation matrix of a vector representation of the assays (Figure 3, see the Experimental Methods section for details). At a high level, four clusters of assays are apparent. The largest cluster is composed mainly of cytotoxicity assays (purple) with several pathway-specific (green) and one target-specific (red) assays. The cytotoxicity assays exhibit a negative correlation with a number of other assays (e.g., apoptotic assays), which can be largely attributed to the normalization scheme. Agonist assays have positive normalized areas under the curve (nAUCs), whereas antagonist assays have negative nAUCs. However, the observed negative correlations are modest. The second major cluster is composed of the three target-specific (red), two pathway and two cytotoxicity assays that exhibit overall poor correlation with any other. This is likely indicative of the orthogonal nature of biological or chemical processes captured by these assays. For instance, counter-assays associated with different screening technologies, such as AlphaLISA or fluorescence, are poorly correlated with each other, as expected. However, the diaphorase and redox counter-assays that are also in this cluster are correlated to some degree as one would expect; the negative correlation in this case is due to the normalization schemes. Two smaller clusters are characterized by a high correlation among the associated assays. One cluster includes the apoptotic assays, membrane integrity, protease, and HIF1 assays. The other cluster is composed of the p53, ATAD5, and CAR assays. Similar observations can be made when clustering is performed with the help of $\log AC_{50}$ and efficacy values of samples (see Figures S5 and S6). Overall, the screening results in the Canvass assay panel confirm general expectations based on the nature of the individual assays.

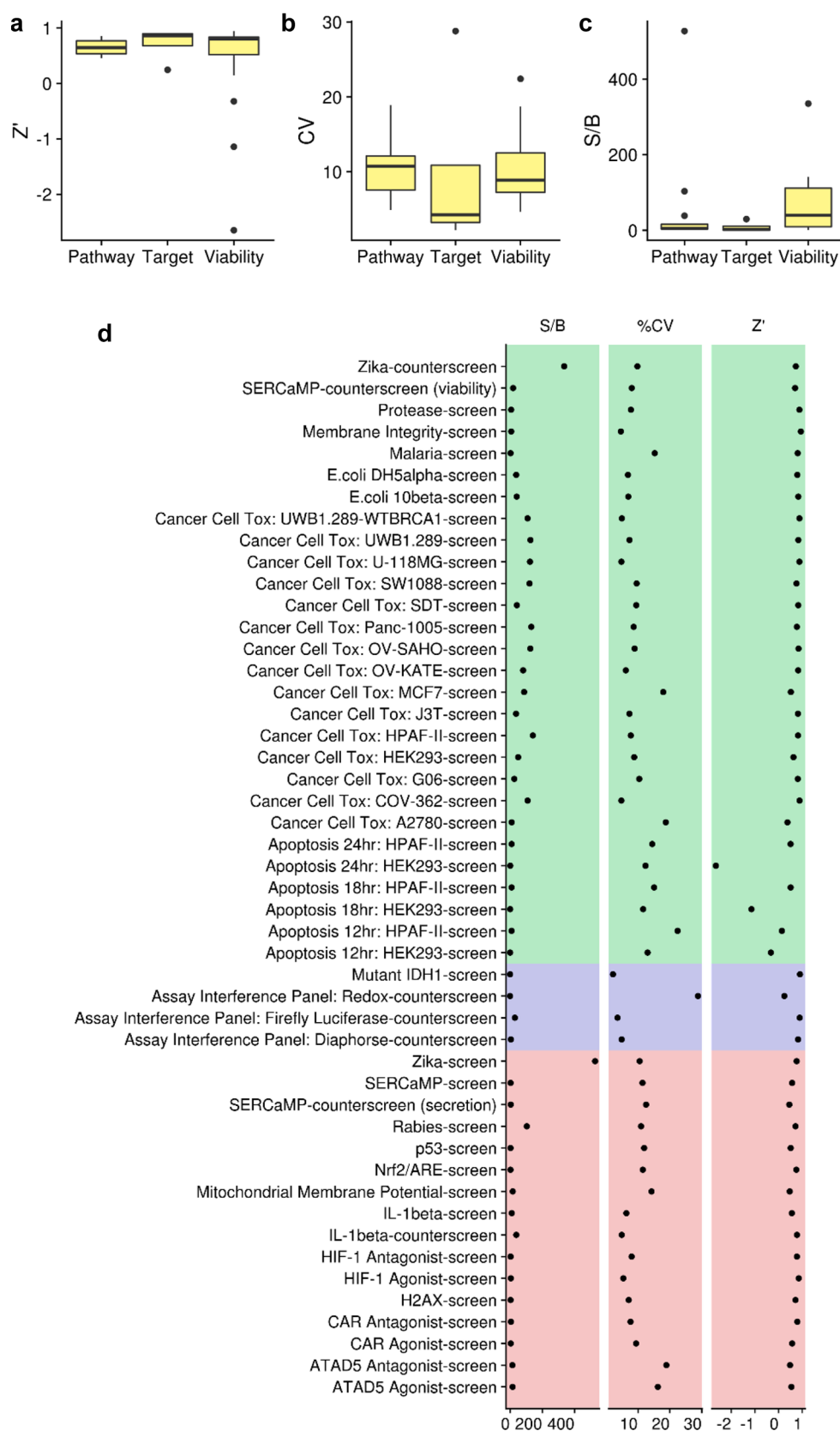


Figure 2. (a–c) Distribution of median assay quality control measures among the three assay classes (pathway, target, and viability). (d) Summary of quality control metrics for the Canvass assay panel, characterized as a pathway-based (pink), a target-based (blue), or a viability assay (green). For all metrics the median values, across all plates run in the assay, are reported.

Promiscuity Analysis. Promiscuous compounds can pose challenges in screening campaigns.²⁴ Promiscuity can be due to

assay interference (such as quenching and autofluorescence) or intractable mechanisms of action, including nonspecific

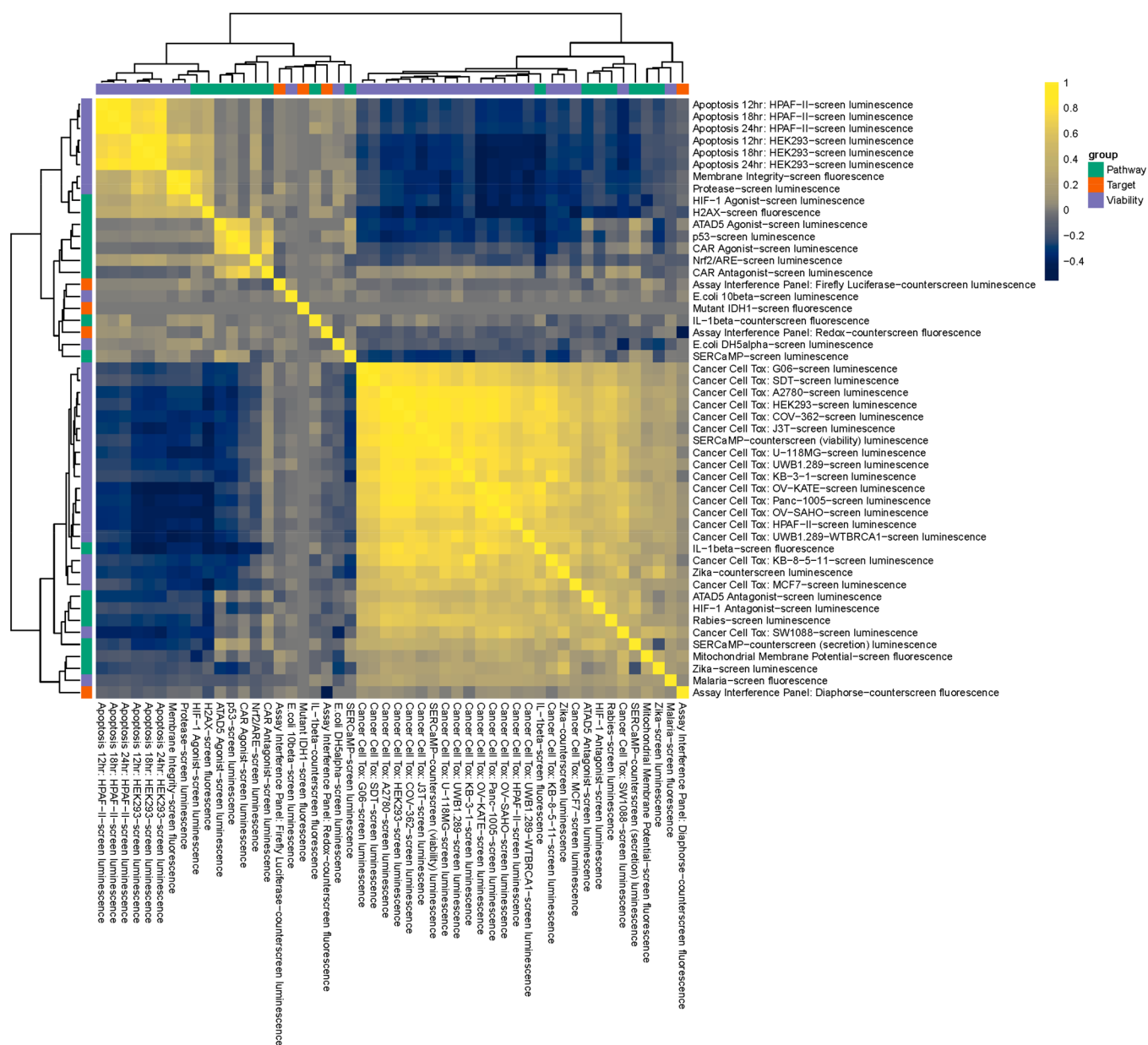


Figure 3. Heatmap representation of the clustering of the assays, based on the Pearson correlation matrix computed from z-scored compound nAUCs. Pearson correlation can have a value between -1 and 1 , where 0 means no correlation, 1 means completely positively correlated, and -1 means completely negatively correlated.

reactivity, redox, or aggregation by pan-assay interference compounds (PAINS),²⁵ and can be evaluated from the hit-rate among all high-throughput screens run during a given period.^{24,26,27} Given that we screened the Canvass library in 50 assays, we characterized the promiscuity of these compounds based on their nAUC values. We considered the absolute value of nAUCs and ignored the pharmacological action (inhibitor, agonist, or antagonist) of the individual compounds. Using this parameter, we defined a compound as promiscuous using two rules: (i) the transformed nAUC value falls into the 90th percentile in a given assay, and (ii) the first condition holds true for at least 40% of the assays. This rule identified 49 compounds as promiscuous, and these are summarized in Figure S3 (compounds listed in Table S2). However, given that the majority of the assays in which these compounds are active are cytotoxicity assays, rather than target-specific assays, the

commonality of the assay end-points may unfairly emphasize their promiscuity. A number of these compounds, however, do appear as hits in target-specific assays (e.g., ATAD5, constitutive androstane receptor [CAR], SERCaMP), but their activity could have been driven by toxicity. To identify compounds not captured by the use of nAUC values, similar promiscuity analyses were carried out using the log AC_{50} data and the absolute value of the efficacy data. For the promiscuity analysis based on log AC_{50} data, compounds with a log AC_{50} value lower than the 10th percentile were considered, and the analysis did not reveal additional promiscuous compounds. The promiscuity analysis of the efficacy data was performed in an analogous manner to the nAUC analysis, and it revealed 12 additional promiscuous compounds that might be associated with cytotoxicity or aggregation at high concentrations (Figure S3).

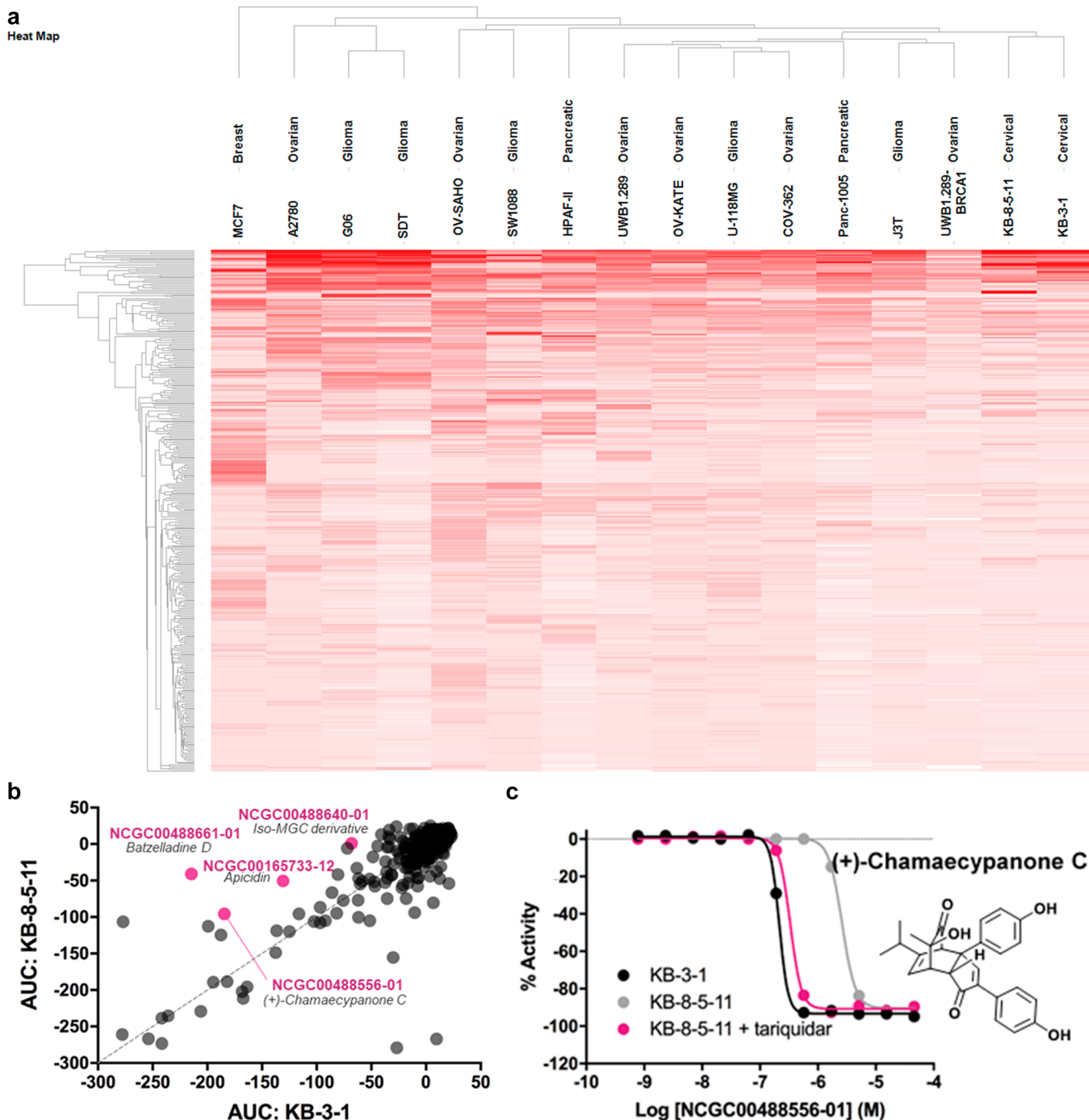


Figure 4. (a) Comparison of the cytotoxicity of each compound (rows) in 16 cancer cell lines (columns). The heatmap was generated based on the area under the dose–response curve (AUC). Dark red indicates a more potent and efficacious compound. (b) AUC correlation plot of KB-8-5-11 vs KB-3-1. AUC for each Canvass compound is represented by a dot with prospective P-gp substrates highlighted (pink) above the unity line (dashed). (c) Dose–response activity of (+)-chamaecypanone C, a prospective P-gp substrate identified in the Canvass library screen. This compound showed selective killing against KB-3-1 (black), resistance in KB-8-5-11 (gray), and reversal of resistance to levels approaching that of KB-3-1 in KB-8-5-11 + 1 μ M tariquidar (pink).

Cytotoxicity Panel Overview. We profiled the cytotoxicity of the Canvass library against a collection of 16 cell lines representing a range of malignancies. The primary motivation for assessing cell killing was the well-known contribution of natural product sources to the chemotherapeutic pharmacopoeia. A secondary goal was to identify cytotoxic compounds that may produce artifacts in other cell-based assays performed as part of the library profiling. Sensitivity varied across all

lines (Figure 4a), though some compounds demonstrated near pan-activity: herboxidiene (NCGC00488492), strophanthidin 3-*O*- β -glucopyranosyl-(1,2)-*O*- β -diginopyranosyl-(1,4)-*O*- β -cymaropyranosyl-(1,4)-*O*- β -digitoxopyranoside (NCGC00488465), and lactimidomycin (NCGC00488635). There was no clear clustering of sensitivity by tissue-of-origin, although the sensitivity of the canine glioma cell lines G06 and SDT closely correlated. Of the compounds in the library, 49% demonstrated

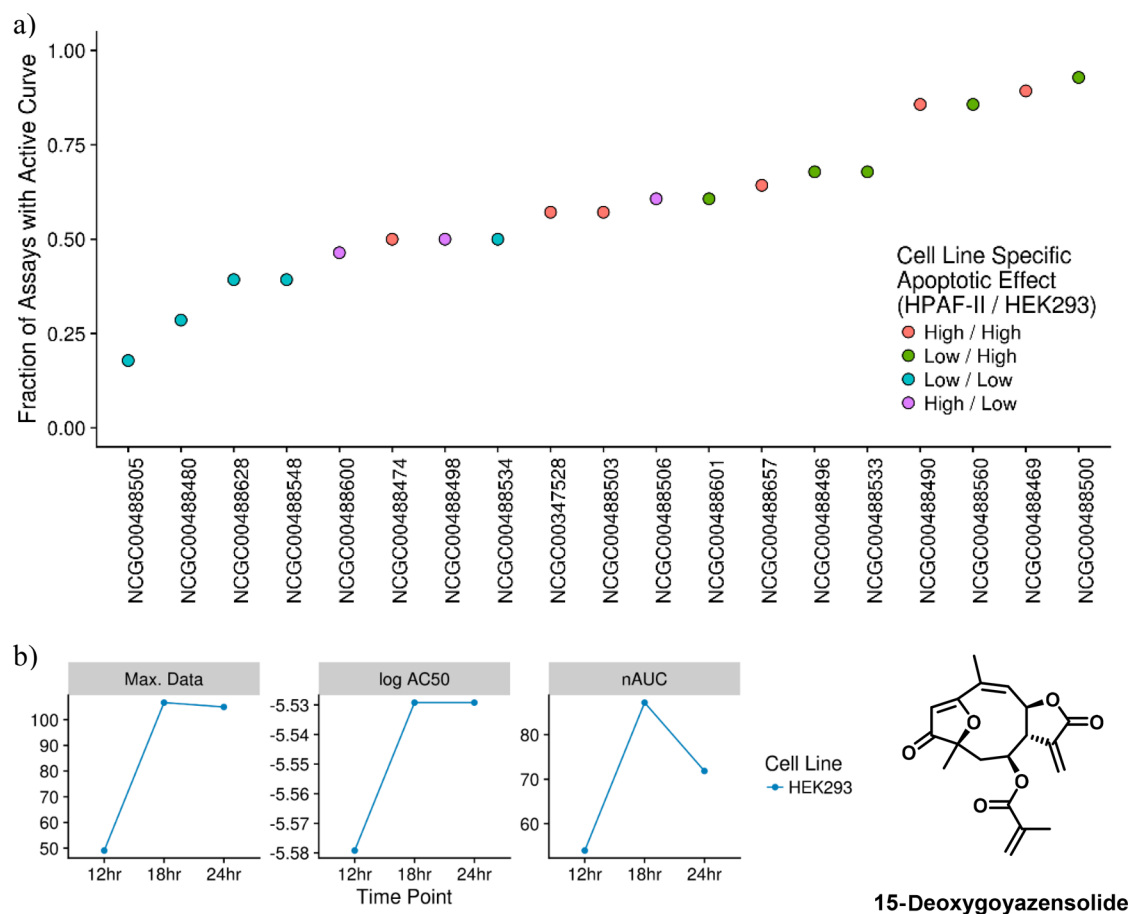


Figure 5. Compounds that induce apoptosis: (a) 19 compounds were identified as able to induce apoptosis as they were active at all three time-points in the apoptosis assay. The graph shows the fraction of the 27 non-caspase viability assays in which these 19 compounds are active. Individual points are colored by the maximum reading (Max. Data) observed across the two caspase assays (cell lines: HPAF-II and HEK293). The observed maximum readings were grouped based on quartiles (low—less than equal to median; and high—greater than median). Inactive samples were assigned to category “low” for clarity. (b) Curve fit parameters for 15-Deoxygoyazensolide the most potent compound in the Canvass library exhibiting apoptotic activity at all three time-points in any of the two cell lines (HEK293, HPAF-II) and its structure; nAUC = normalized area under the curve.

class 1 or 2 curves⁷ with maximum response over 50% against at least one cell line.

Natural product cytotoxins are susceptible to efflux by multidrug-resistance transporters. To identify P-glycoprotein (P-gp) substrates, we tested compounds against a P-gp over-expressing cell line, KB-8-5-11, and its nonexpressing parental counterpart, KB-3-1. Inhibition of P-gp reverses resistance to P-gp substrates. KB-8-5-11 cells cotreated with 1 μ M tariquidar (KB-8-5-11 + tariquidar), a known P-gp inhibitor, were also tested to confirm the P-gp substrates. P-gp substrates demonstrate reduced cell killing against KB-8-5-11 cells, and greater activity against the parent KB-3-1 cell line. Comparison of activity (AUC) against the two cell lines revealed a small number of compounds less active against the P-gp-expressing cell line (above the unity line, Figure 4b), and most compounds did not demonstrate significant cytotoxicity (clustered at the origin, Figure 4b). Confirmation retesting revealed 4 substrates among 40 compounds demonstrating cytotoxicity (pink spheres, Figure 4b): batzelladine D (NCGC00488661), (+)-chamaecypanone C (NCGC00488556, Figure 4c), apicidin (NCGC00165733, Figure S4), and an isomigrastatin derivative (NCGC00488640). Concentration–response curves show that KB 8-5-11 cells were resistant to cell killing, but were sensitized by the P-gp inhibitor tariquidar.

Apoptotic Behavior. The Canvass assay panel enables us to probe a wide variety of pharmacological responses, but at fixed

time-points. The inclusion of the apoptosis assays (measured using Caspase-Glo) read at three time-points (12, 18, and 24 h) allowed us to explore the apoptotic response to the compounds over time. The majority of the 346 Canvass compounds did not induce apoptosis at any of the time-points (just 33 compounds exhibited an active dose–response at one or more time-points), and only 19 compounds induced caspase activity in a dose–response manner at all three time-points in any of the two assays (Figure 5a). In general, these 19 compounds did not exhibit high potencies, with the exception of 15-deoxygoyazensolide (NCGC00488496) (see Figure 5b). Nevertheless, we categorized the observed activities in relation to the detected maximal response signal (“Max. Data”). This schema gave rise to four categories. Compounds characterized by “high” apoptotic activity in HPAF-II but “low” in HEK293 assays may have implications for biasing apoptosis induction toward cancer as opposed to normal cells. Three compounds were found that fall under this group of interest (NCGC00488600, NCGC00488498, NCGC00488506).

We also examined the correlation between the 19 compounds inducing apoptosis at all three time-points and their activities in the remaining 27 viability assays. We computed the fraction of assays in which each of these compounds exhibited an active curve class. These values ranged from 0.18 (chatancin, NCGC00488505)

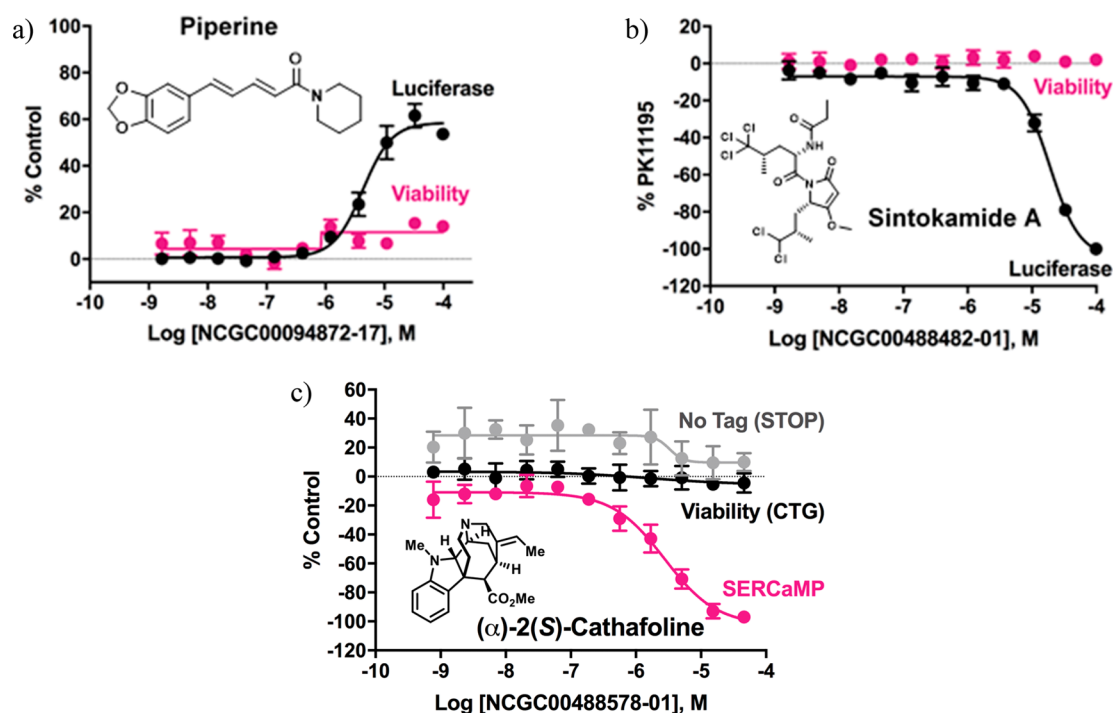


Figure 6. (a) Dose–response curves for the CAR activator (NCGC00094872) identified in the Canvass library. The stable cell line treated with compound and CITCO for 24 h in 1536-well plates was treated with ONE-Glo, and the luminescence intensity was calculated (black). The efficacies were compared to the CITCO positive controls, and the viability was also detected using fluorescence in the same well (pink). Data are expressed as mean \pm standard error of the mean for triplicate assays. (b) Dose–response curves for the CAR deactivator (NCGC00488482) using same method in part a, except cells were cotreated with PK11195 instead of CITCO, and PK11195 was used as positive controls. (c) 2(S)-Cathafoline is active in the secreted ER calcium-monitoring protein (SERCaMP) assay (pink), indicating the compound stabilizes ER calcium. Activity was examined in the primary SERCaMP assay (pink), a secretion counter-screen (gray), and a viability counter-screen (black). Mean activity \pm SD ($n = 3$) is shown.

to 0.93 ((+)-cryptocaryol A, NCGC00488500). The values for all compounds are summarized in Figure 5a, where we overlaid the maximum efficacy observed in the caspase assays for each compound. In general, compounds that strongly induce apoptosis tended to also show cytotoxic behavior across most of the viability assays.

Activity Highlights. Nuclear receptors (NRs) are proteins that regulate physiological homeostasis, metabolism, and development.²⁸ One crucial NR is the constitutive androstane receptor (CAR, NR1I3), which regulates many drug-metabolizing enzymes and transporters while also playing an essential role in energy homeostasis.²⁹ In this study, we screened the Canvass library of natural products for activation/deactivation of CAR employing a double stable reporter cell line (HepG2-CYP2B6-hCAR).³⁰ To identify activators, we added 0.75 μ M 1-(2-chlorophenylmethylpropyl)-3-isoquinoline-carboxamide (PK11195),³¹ a known CAR inhibitor, while adding 50 nM 6-(4-chlorophenyl)imidazo[2,1-*b*][1,3]thiazole-5-carbaldehyde *O*-(3,4-dichlorobenzyl)oxime (CITCO, a known CAR activator),³² to identify any deactivators in the Canvass collection.

Piperine (NCGC00094872), the pungent component of black and long peppers, and (+)-herboxidiene (NCGC00488492) were identified as strong, potent activators of CAR with efficacies of 63% and 609% (relative to CITCO) and EC_{50} values of 4.9 and 0.12 μ M, respectively (Figure 6a). Sintokamide A (NCGC00488482) produced a very strong deactivation of CAR with an efficacy of 102% (relative to PK11195) and an IC_{50} of 9.3 μ M (Figure 6b). This compound also did not induce cytotoxicity, as a multiplexed system was used in the same well to identify luminescence and viability concurrently. All three of these compounds, however, also elicited a very strong response for

pregnane X receptor (PXR, NR1I2) activation. However, a cross-talk exists between PXR and CAR, complicating the deconvolution of each receptor's independent contributions. Therefore, further studies are necessary to more accurately identify the specific NR-related behavior of these compounds both *in vitro* and *in vivo*.

Canvass was also screened for activity on the endoplasmic reticulum (ER) calcium store. Small molecules that stabilize ER calcium are of interest for diverse pathologies, including cardiovascular diseases and diabetes.³³ The secreted ER calcium-monitoring protein (SERCaMP) assay³⁰ was utilized to assess ER calcium homeostasis by measuring a reporter protein that is secreted in response to ER calcium depletion.²¹ SH-SY5Y neuroblastoma cells were treated with thapsigargin to deplete the ER calcium store, and compounds that reduced SERCaMP secretion, indicative of ER calcium stabilization, were identified. The screen was run in duplicate on separate days, with 30 compounds showing activity in both experiments. False positives attributable to nonspecific effects on the secretory pathway or cell viability were identified using a set of counter-screens, and 17 compounds (57%) were active in one or both of the counter-screens. The most favorable profile was observed for (–)-2(S)-cathafoline (NCGC00488578), with an AC_{50} of 2.8 μ M in the SERCaMP assay and no activity in the counter-screens (Figure 6c), suggesting that the compound stabilizes ER calcium levels. Analyzing the full activity profile of (–)-2(S)-cathafoline revealed that it is highly selective for ER calcium modulation; it was not a hit in any other assays in the panel. Further studies are underway to examine the effect of this compound in disease models associated with ER calcium dysregulation.

DISCUSSION

With generous contributions from laboratories throughout the world, we assembled a diverse set of 346 natural products for biological evaluation. Comparison of the Canvass library to three other chemical libraries (DrugBank, LC50K, and ChEMBL natural products) revealed that all four inhabit very similar physicochemical property space, though the LC50K subset tends to be focused with respect to most parameters. Given that DrugBank is a drug library and LC50K is designed for lead discovery purposes, their overlap with ChEMBL natural products and the Canvass libraries supports the observation that natural products are the origins for many drugs. Considering both the physicochemical space and fingerprint structural comparisons, Canvass compounds are most similar to druglike molecules in the DrugBank library. Despite some degree of structural overlap with the ChEMBL natural product set, it is evident that the Canvass set samples novel structural space compared to the other three libraries. Though the Canvass compounds comprised of natural products, their properties are more aligned with drugs than with the ChEMBL natural products, earning them a unique position among chemical collections. This suggests that the Canvass library is a good complement to the ChEMBL natural product set for drug discovery purposes.

The Canvass natural product library was screened in 50 distinct assays to broadly canvass the pharmacological activity of the compounds. The assortment of assays utilized in this pilot study cover a wide biological scope, and they were selected because they have been successfully implemented for qHTS at NCATS. Cytotoxicity was heavily emphasized given the history of natural products as anticancer and anti-infective agents. Targets with unique pharmacological potential, including CAR and Nrf2/ARE, were also examined. The balance of the screen probed developing pharmacological mechanisms in other areas such as inflammation (inflammasome) and cardiovascular as well as rare diseases (SERCaMP).

Beyond the facets highlighted through the analyses, notable activities also emerged for a small subset of compounds. We discovered three natural products that modulate the nuclear receptor CAR. Interestingly, the molecules highlighted above are structurally quite distinct from one another; one of them, piperine, is a component of the dietary spice pepper, and its observed activity has brought to light an unappreciated pharmacological impact of this widely consumed natural product.

The strategic utility of specific counter-screens facilitated the identification of selectively-active compounds, as exemplified by (-)-2(S)-cathafoline in the SERCaMP assay. The calcium-stabilizing effect of this molecule is significant, especially given that cathafoline was highly selective across our panel; it was not active in any other Canvass assays. Though it is a member of a large family of indole alkaloids, little was known about the biological activity of cathafoline prior to this Canvass study. Fortuitously, the Garg laboratory's prior total synthesis campaign³⁴ provided access to synthetic cathafoline, which, when submitted in response to our call for natural products, enabled the discovery of cathafoline's promising *in vitro* activity. Studies are underway to elucidate the biological target of cathafoline and to realize the full potential of its calcium modulatory effects.

Investigations of potent natural products continue to reveal key biological targets with therapeutic relevance. For example, the isolation and biological evaluation of peptides isolated from

viper snake venom helped bring to light the importance of the angiotensin converting enzyme (ACE) in regulating blood pressure. This discovery inspired the design of the drug captopril for the treatment of hypertension.³⁵ There is no doubt that natural products will continue to impact the numerous unmet medical needs of both ubiquitous illnesses and rare diseases alike, especially as we have only just begun to understand new and developing areas of disease biology, such as the human microbiome. It is our hope that screening natural product libraries for a wide range of biological activities, as with the Canvass pilot program, will accelerate the development of therapeutics in many disease areas. The Canvass pilot program serves as an example of how to successfully leverage resources and expertise throughout the scientific community for the evaluation of natural products in drug discovery, and it will serve as a framework for larger scale investigations in the future.

EXPERIMENTAL METHODS

The Canvass library of 346 natural products was assembled through a broad solicitation of both the academic and private sectors via the Canvass website, with heavy emphasis on natural product isolation and total synthesis laboratories. To streamline the pursuit of biological discoveries and minimize deconvolution, only pure natural products were accepted. The NCATS Compound Management team provided our collaborators with tared, barcoded vials for natural product submissions to facilitate both the receipt and processing of samples. Quality control (QC) for purity >85% was managed in two ways: by requiring the cosubmission by PIs of recent analyses of submitted natural products utilizing either LC/MS or ¹H NMR, and subsequently upon receipt of samples, NCATS performing LC/MS analyses of all submissions. By implementing two QC checks, we minimized QC failures for natural products which are unstable to our standard LC/MS QC protocols. In total, 45 PIs or institutions submitted a sum of 346 natural products to assemble the Canvass library, which was subsequently evaluated in 50 whole-cell or biochemical assays. This library continues to serve as a valuable resource in screening campaigns following the work described herein.

Data generated by the Canvass program has been made available to our collaborators via the website, located at <https://tripod.nih.gov/canvass/>. The website provides open access to both general and detailed overviews of the program as a whole, as well as a subset of the assays which were planned at the outset, and a portal for questions and answers. By registering, collaborators' accounts allowed access to secure and private interfaces to expedite compound submissions and to simplify data browsing. Importantly, data generated for a given set of compound submissions were initially only made accessible to the submitters. The application provides the traditional tabular view of compound activities, with a variety of visualizations to provide high-level summaries of activity in the assay panel, along with detailed views for individual compounds. With this publication, the full data set presented herein is now available at the Canvass website (<https://tripod.nih.gov/canvass/>).

DATA SETS AND COMPUTATIONAL METHODS

Chemical Libraries. The Canvass library consists of 346 compounds that can be described as natural products. To better characterize the chemical space represented by these compounds, we included three additional chemical libraries in our analysis: the DrugBank Approved Drugs (2073 compounds,

database version: 2.0.9), the ChEMBL natural products (1921 compounds, database version: ChEMBL 23), and the Life Chemicals Diversity Set of 50K molecules (LC50K, 50 240 compounds). The LC50K collection was included to provide a baseline for the comparison of molecular properties. To this end, a subset consisting of 3000 molecules was sampled randomly from the LC50K collection and used in subsequent analyses. Chemical libraries were subjected to the same standardization scheme as part of a KNIME workflow including community nodes originating from CDK cheminformatics suites.^{36–39} Standardization steps involved keeping only the largest component of compounds.

qHTS Data Processing Pipeline. While the data processing details for individual assays can be found in their respective references, we briefly outline the processing pipeline here. Plate data were normalized to the per-plate positive and negative controls, and dose–response curves were fitted using a grid-based algorithm.^{40,41} Dose–response curves were then assigned a curve class (see Seethala and Zhang^{40,6} for more detailed definitions of curve classes), which is a heuristic classification that allows us to easily identify good quality versus poor quality dose–response curves. To assign one of seven possible classes, the curve is evaluated in terms of features like asymptote definition, the R^2 of the curve fit, and efficacy. Furthermore, for assays measuring an increase in signal (e.g., agonist assays), curve classes are positive, and for those measuring a decrease in signal (e.g., inhibitor or antagonist assays), curve classes are negative. Broadly, we consider curve classes of 1.1, 1.2, 2.1, and 2.2 (and the corresponding negative values) as good quality active dose–responses, curve class of 4 as inactive (i.e., no dose response), and all others as inconclusive dose–responses. We only deviated from this classification in the case of Caspase assays, where good quality active dose–responses curves were considered those that are associated with a curve class value different from 4.

AUC Computation. In the Canvass assays, the more potent and efficacious a compound is, the larger its absolute area under the curve (AUC) value. Moreover, the sign of the AUC reflects the pharmacological action of a compound (e.g., inhibitors can be distinguished from activators). It should be noted that automatically generated raw AUC values can be misleading in certain cases. For the case of an inhibitory dose–response curve, which starts in a positive response range, the absolute AUC can be small due to the nature of numerical integration. A similarly misleading absolute AUC value is obtained for activation dose–response curves, which start in the negative response range. Such anomalies are typically the result of noisy assay data. To remedy such cases, we found it necessary to transform the raw AUC values to better reflect the magnitude of biological responses independent of whether the curve exhibits negative or positive response values (due to the plate normalization scheme). Accordingly, in the case of inhibition curves, the maximal y -value of a curve is subtracted from all y -values. In an analogous manner, the minimal y -value was subtracted from the y -values of activation curves. The resultant normalized AUC values are denoted by nAUC to distinguish them from raw AUC values.

Assay Clustering. To cluster the assays, we first represented each assay as a 346-element vector of compound nAUC values. Next, we computed the pairwise Pearson correlation matrix, which was then clustered using hierarchical clustering with complete linkage. We also computed clusterings using potency (setting it to 10 μ M for compounds that were inactive in an assay) and efficacy (Figures S5 and S6).

Descriptor Computation. Physicochemical descriptors (molecular weight, rotatable bond number, flexibility,¹⁴ number of hydrogen-bond donors/acceptors, clogP,⁴² TPSA⁴³) and ECFP6 fingerprints were computed using CDK nodes for KNIME (version 3.4.2) and processed in R statistical suite (version 3.4.4) using the fingerprint package.

Library Overlap. Thornton's separability index (S)¹⁸ is a class separability measure designed to discriminate between objects from different classes and is commonly used to characterize the quality of clustering. We applied this measure to quantify the degree of overlap of two libraries in a predefined chemical space. S is defined as the fraction of compounds for which their nearest neighbor is not from the same library as themselves. Thus, S ranges from 1 (two libraries completely overlap with each other) to 0 (two libraries have no overlap). Note that these limits are somewhat dependent on the actual spatial distribution of compounds in a given chemical space. Nonetheless, for the purposes of comparing library overlap in a relative manner, the use of S is sufficient. We implemented S using R 3.4.4 (code available at <https://spotlite.nih.gov/gzahoranszky/CANVASS.git>).

Quantifying 3D-Likeness. The 3D-likeness of compounds was characterized with the help of a plane of best fit (PBF)¹³ measure implemented in the RDKit Python Application Programming Interface (API) (version 2017.09.1).³⁷ To generate PBF values, a low energy conformer needs to be generated for each compound. Low energy conformers were generated by a KNIME workflow.^{36–39}

Embedding Compounds into 2D Chemical Space. We defined a physicochemical property space using molecular weight (MW), H-bond donor count, H-bond acceptor count, XLogP,¹² PBF,¹³ and flexibility.¹⁴ We then employed tSNE (as implemented in the Rtsne package⁴⁴ for R) to perform dimension reduction to 2D,¹⁷ using the Euclidean metric for distance computations. Apart from setting $\theta = 0.1$, all other parameters were set at their default values. For the case of fingerprints, we first computed a pairwise similarity matrix using the Tanimoto metric^{45,46} and converted it to a distance matrix (using $D = 1 - \text{similarity}$) for input to the tSNE algorithm, using the same parameters as used for the physicochemical property space. The 2D data sets were visualized using R 3.4.4. All workflow files, Python source code, computed descriptor, and fingerprint data are available at <https://spotlite.nih.gov/gzahoranszky/CANVASS.git>.

Novelty Assessment of Canvass Compound Structures. The novelty of the Canvass compounds was assessed on the basis of the Bemis–Murcko scaffold (BMS)⁴² of compounds. First, the BMSs of compounds were determined in all molecular libraries. Next, a unique set of BMSs were derived library-wise. Finally, the overlap of pairs of unique BMS sets was quantified relative to the number of unique BMSs in each set. Of note, the rest of the analyses in this study involved only a random subset of the LC50K library whereas this assessment was performed on the entire LC50K library (Table S1).

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.8b00747.

Additional figures depicting more detailed analyses of the assay data (including assay QC measurement, additional heatmaps and clustering analyses, additional dose–response

curves), a more in depth description of the novelty of the canvass compounds, additional information and structures for promiscuous compounds in the Canvass library, and the assay protocols utilized in our study (PDF)

AUTHOR INFORMATION

Corresponding Authors

*Phone: 301-480-9928. Fax: 301-217-5736. E-mail: hallma@mail.nih.gov.

*Phone: 301-319-9272. Fax: 301-319-9449. E-mail: jason.m.rohde3.civ@mail.mil.

ORCID

Kanny K. Wan: 0000-0002-7246-343X

Rodrigo B. Andrade: 0000-0002-4554-8323

M. Kevin Brown: 0000-0002-4993-0917

Noah Z. Burns: 0000-0003-1064-4507

Jin K. Cha: 0000-0003-4038-3213

Jon Clardy: 0000-0003-0213-8356

Neil K. Garg: 0000-0002-7793-2629

Robert B. Grossman: 0000-0001-8866-3754

Chambers C. Hughes: 0000-0002-7000-6438

Jeffrey N. Johnston: 0000-0002-0885-636X

A. Douglas Kinghorn: 0000-0002-6647-8707

David G.I. Kingston: 0000-0001-8944-246X

Ohyun Kwon: 0000-0002-5405-3971

Thomas J. Maimone: 0000-0001-5823-692X

Brian T. Murphy: 0000-0002-1372-3887

Pavel Nagorny: 0000-0002-7043-984X

David E. Olson: 0000-0002-4517-0543

Larry E. Overman: 0000-0001-9462-0195

Lauren E. Brown: 0000-0001-9489-484X

John A. Porco, Jr.: 0000-0002-2991-5680

Samir A. Ross: 0000-0002-3807-1299

Richmond Sarpong: 0000-0002-0028-6323

Indrajeet Sharma: 0000-0002-0707-0621

Jared T. Shaw: 0000-0001-5190-493X

Ben Shen: 0000-0002-9750-5982

Wei Shi: 0000-0001-5453-1753

Corey R.J. Stephenson: 0000-0002-2443-5514

Derek S. Tan: 0000-0002-7956-9659

Yi Tang: 0000-0003-1597-0141

Regan J. Thomson: 0000-0001-5546-4038

David A. Vosburg: 0000-0003-3424-5471

Jimmy Wu: 0000-0002-0938-8019

Armen Zakarian: 0000-0002-9120-1232

Zhong Zuo: 0000-0002-6976-6157

James Inglese: 0000-0002-7332-5717

Matthew D. Hall: 0000-0002-5073-442X

Jason M. Rohde: 0000-0003-3187-3739

Present Addresses

^UNexus Discovery Advisors, 7820B Wormans Mill Road, Suite 208, Frederick, MD 21701, USA

^VVertex Pharmaceuticals, 50 Northern Avenue, Boston, MA 02210, USA

^WWalter Reed Army Institute of Research, 503 Robert Grant Avenue, BLDG 503, RM 2A20, Silver Spring, MD 20910, USA

Author Contributions

[†]S.E.K. and G.Z.-K. contributed equally.

Notes

The authors declare no competing financial interest.

Safety statement: no unexpected or unusually high safety hazards were encountered in this work.

ACKNOWLEDGMENTS

NCATS thanks Dingyin Tao for assistance with compound characterization. This research was supported by the Intramural Research Program of the National Center for Advancing Translational Sciences, National Institutes of Health (NIH). R.B.A. acknowledges support from NSF (CHE-1665145) and NIH (GM126221). M.K.B. acknowledges support from NIH (5R01GM110131). N.Z.B. thanks support from NIGMS, NIH (R01GM114061). J.K.C. acknowledges support from NSF (CHE-1665331). J.C. acknowledges support from the Fogarty International Center, NIH (TW009872). P.A.C. acknowledges support from the National Cancer Institute (NCI), NIH (R01CA158275), and the NIH/National Institute of Aging (P01AG012411). N.K.G. acknowledges support from NSF (CHE-1464898). B.C.G. thanks the support of NSF (RUI: 213569), the Camille and Henry Dreyfus Foundation, and the Arnold and Mabel Beckman Foundation. C.C.H. thanks the start-up funds from the Scripps Institution of Oceanography for support. J.N.J. acknowledges support from NIH (GM 063557, GM 084333). A.D.K. thanks the support from NCI, NIH (P01CA125066). D.G.I.K. acknowledges support from the National Center for Complementary and Integrative Health (1 R01 AT008088) and the Fogarty International Center, NIH (U01 TW00313), and gratefully acknowledges courtesies extended by the Government of Madagascar (Ministère des Eaux et Forêts). O.K. thanks NIH (R01GM071779) for financial support. T.J.M. acknowledges support from NIH (GM116952). S.M. acknowledges support from NIH (DA045884-01, DA046487-01, AA026949-01), the Office of the Assistant Secretary of Defense for Health Affairs through the Peer Reviewed Medical Research Program (W81XWH-17-1-0256), and NCI, NIH, through a Cancer Center Support Grant (P30 CA008748). K.N.M. thanks the California Department of Food and Agriculture Pierce's Disease and Glassy Winged Sharpshooter Board for support. B.T.M. thanks Michael Mallowney for his contribution in the isolation, elucidation, and submission of the compounds in this work. P.N. acknowledges support from NIH (R01 GM111476). L.E.O. acknowledges support from NIH (R01-HL25854, R01-GM-30859, R01-NS-12389). L.E.B., J.K.S., and J.A.P. thank the NIH (R35 GM-118173, R24 GM-111625) for research support. F.R. thanks the American Lebanese Syrian Associated Charities (ALSAC) for financial support. I.S. thanks the University of Oklahoma Startup funds for support. J.T.S. acknowledges support from ACS PRF (53767-ND1) and NSF (CHE-1414298), and thanks Drs. Kellan N. Lamb and Michael J. Di Maso for their synthetic contribution. B.S. acknowledges support from NIH (CA78747, CA106150, GM114353, GM115575). W.S. acknowledges support from NIGMS, NIH (R15GM116032, P30 GM103450), and thanks the University of Arkansas for startup funds and the Arkansas Biosciences Institute (ABI) for seed money. C.R.J.S. acknowledges support from NIH (R01GM121656). D.S.T. thanks the support of NIH (T32 CA062948-Gudas) and PhRMA Foundation to A.L.V., NIH (P41 GM076267) to D.S.T., and CCSG NIH (P30 CA008748) to C. B. Thompson. R.E.T. acknowledges support from NIGMS, NIH (GM129465). R.J.T. thanks the American Cancer Society (RSG-12-253-01-CDD) and NSF (CHE1361173) for support. D.A.V. thanks the Camille and

Henry Dreyfus Foundation, the National Science Foundation (CHE-0353662, CHE-1005253, and CHE-1725142), the Beckman Foundation, the Sherman Fairchild Foundation, the John Stauffer Charitable Trust, and the Christian Scholars Foundation for support. J.W. acknowledges support from the American Cancer Society through the Research Scholar Grant (RSG-13-011-01-CDD). W.M.W. acknowledges support from NIGMS, NIH (GM119426), and NSF (CHE1755698). A.Z. acknowledges support from NSF (CHE-1463819).

REFERENCES

- (1) Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629–61.
- (2) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discovery* **2015**, *14* (2), 111–29.
- (3) Cheng, K. C.; Cao, S.; Raveh, A.; MacArthur, R.; Dranchak, P.; Chlipala, G.; Okoneski, M. T.; Guha, R.; Eastman, R. T.; Yuan, J.; Schultz, P. J.; Su, X. Z.; Tamayo-Castillo, G.; Matainaho, T.; Clardy, J.; Sherman, D. H.; Inglesse, J. Actinoramide a identified as a potent antimalarial from titration-based screening of marine natural product extracts. *J. Nat. Prod.* **2015**, *78* (10), 2411–22.
- (4) Ziemert, N.; Alanjary, M.; Weber, T. The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016**, *33* (8), 988–1005.
- (5) Hifan—center for high-throughput functional annotation of natural products. <http://www.hifan.swmed.edu> (accessed Mar 6, 2018).
- (6) Combating antibiotic-resistant bacteria (carb). <https://obamaadministration.archives.performance.gov/content/combating-antibiotic-resistant-bacteria-carb.html> (accessed Oct 1, 2018).
- (7) Inglesse, J.; Auld, D. S.; Jadhav, A.; Johnson, R. L.; Simeonov, A.; Yasgar, A.; Zheng, W.; Austin, C. P. Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (31), 11473–8.
- (8) Yasgar, A.; Shinn, P.; Jadhav, A.; Auld, D.; Michael, S.; Zheng, W.; Austin, C. P.; Inglesse, J.; Simeonov, A. Compound management for quantitative high-throughput screening. *JALA* **2008**, *13* (2), 79–89.
- (9) Natural products were categorized with varying levels of specificity/granularity to reasonably distribute compounds into classes.
- (10) Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B.; Zhou, Y.; Wishart, D. S. Drugbank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res.* **2014**, *42* (D1), D1091–D1097.
- (11) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Kruger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL bioactivity database: An update. *Nucleic Acids Res.* **2014**, *42* (D1), D1083–D1090.
- (12) Wang, R. X.; Fu, Y.; Lai, L. H. A new atom-additive method for calculating partition coefficients. *J. Chem. Inf Comp Sci.* **1997**, *37* (3), 615–621.
- (13) Firth, N. C.; Brown, N.; Blagg, J. Plane of best fit: A novel method to characterize the three-dimensionality of molecules. *J. Chem. Inf. Model.* **2012**, *52* (10), 2516–25.
- (14) Dragon. http://www.taletе.mi.it/products/dragon_description.htm (accessed Aug 14, 2018).
- (15) Meyers, J.; Carter, M.; Mok, N. Y.; Brown, N. On the origins of three-dimensionality in drug-like molecules. *Future Med. Chem.* **2016**, *8* (14), 1753–67.
- (16) Lovering, F.; Bikker, J.; Humblet, C. Escape from flatland: Increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, *52* (21), 6752–6756.
- (17) van der Maaten, L.; Hinton, G. Visualizing data using t-sne. *J. Mach Learn Res.* **2008**, *9*, 2579–2605.
- (18) Thornton, C. Separability is a learner's best friend. *Persp Neural Comp* **1998**, 40–46.
- (19) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–54.
- (20) Willighagen, E. L.; Mayfield, J. W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliazkova, N.; Kuhn, S.; Pluskal, T.; Rojas-Cherto, M.; Spjuth, O.; Torrance, G.; Evelo, C. T.; Guha, R.; Steinbeck, C. The chemistry development kit (cdk) v2.0: Atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminf.* **2017**, *9* (1), 33.
- (21) Henderson, M. J.; Wires, E. S.; Trychta, K. A.; Richie, C. T.; Harvey, B. K. Sercamp: A carboxy-terminal protein modification that enables monitoring of er calcium homeostasis. *Mol. Biol. Cell* **2014**, *25* (18), 2828–39.
- (22) Iversen, P. W.; Eastwood, B. J.; Sittampalam, G. S.; Cox, K. L. A comparison of assay performance measures in screening assays: Signal window, z' factor, and assay variability ratio. *J. Biomol. Screening* **2006**, *11* (3), 247–52.
- (23) Zhang, J. H.; Chung, T. D.; Oldenburg, K. R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screening* **1999**, *4* (2), 67–73.
- (24) Rishton, G. M. Reactive compounds and in vitro false positives in hts. *Drug Discovery Today* **1997**, *2* (9), 382–384.
- (25) Baell, J.; Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* **2014**, *513* (7519), 481–3.
- (26) Walters, W. P.; Ajay, Murcko, M. A. Recognizing molecules with drug-like properties. *Curr. Opin. Chem. Biol.* **1999**, *3* (4), 384–387.
- (27) Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E. M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G. Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **2002**, *45* (1), 137–42.
- (28) Olefsky, J. M. Nuclear receptor minireview series. *J. Biol. Chem.* **2001**, *276* (40), 36863–4.
- (29) Lynch, C.; Pan, Y.; Li, L.; Ferguson, S. S.; Xia, M.; Swaan, P. W.; Wang, H. Identification of novel activators of constitutive androstane receptor from fda-approved drugs by integrated computational and biological approaches. *Pharm. Res.* **2013**, *30* (2), 489–501.
- (30) Lynch, C.; Zhao, J.; Huang, R.; Xiao, J.; Li, L.; Heyward, S.; Xia, M.; Wang, H. Quantitative high-throughput identification of drugs as modulators of human constitutive androstane receptor. *Sci. Rep.* **2015**, *5*, 10405.
- (31) Li, L.; Chen, T.; Stanton, J. D.; Sueyoshi, T.; Negishi, M.; Wang, H. The peripheral benzodiazepine receptor ligand 1-(2-chlorophenylmethylpropyl)-3-isoquinoline-carboxamide is a novel antagonist of human constitutive androstane receptor. *Mol. Pharmacol.* **2008**, *74* (2), 443–53.
- (32) Maglich, J. M.; Parks, D. J.; Moore, L. B.; Collins, J. L.; Goodwin, B.; Billin, A. N.; Stoltz, C. A.; Kliewer, S. A.; Lambert, M. H.; Willson, T. M.; Moore, J. T. Identification of a novel human constitutive androstane receptor (car) agonist and its use in the identification of car target genes. *J. Biol. Chem.* **2003**, *278* (19), 17277–83.
- (33) Mekahli, D.; Bultynck, G.; Parys, J. B.; De Smedt, H.; Missiaen, L. Endoplasmic-reticulum calcium depletion and disease. *Cold Spring Harbor Perspect. Biol.* **2011**, *3* (6), a004317.
- (34) Moreno, J.; Picazo, E.; Morrill, L. A.; Smith, J. M.; Garg, N. K. Enantioselective total syntheses of akuammiline alkaloids (+)-strictamine, (–)-2(s)-cathafoline, and (–)-aspidophylline a. *J. Am. Chem. Soc.* **2016**, *138* (4), 1162–5.
- (35) Smith, C. G.; Vane, J. R. The discovery of captopril. *FASEB J.* **2003**, *17* (8), 788–9.
- (36) Open for innovation knime—rdkit. <https://www.knime.com/nodeguide/community/rdkit> (accessed Oct 27, 2018).
- (37) Epam, life sciences open source—nodes for knime. <http://lifescience.opensource.epam.com/indigo/knime.html> (accessed Oct 27, 2018).
- (38) Rdkit: Open-source cheminformatics software. <http://www.rdkit.org> (accessed Oct 27, 2018).

(39) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. Knime: The Konstanz information miner. In *Studies in classification, data analysis, and knowledge organization (gkl 2007)*; Springer, 2007; pp 319–326.

(40) Seethala, R.; Zhang, L. *Handbook of drug screening*, 2nd ed.; Informa Healthcare: New York, 2009; pp xiii, 489.

(41) Wang, Y.; Jadhav, A.; Southal, N.; Huang, R.; Nguyen, D. T. A grid algorithm for high throughput fitting of dose-response curve data. *Curr. Chem. Genomics* **2010**, *4*, 57–66.

(42) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–93.

(43) Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43* (20), 3714–7.

(44) Krijthe, J.; van der Maaten, L.; Rtsne, L. T-distributed stochastic neighbor embedding using a Barnes-hut implementation. <https://cran.r-project.org/web/packages/Rtsne/index.html> (accessed Aug 14, 2018).

(45) Jaccard, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* **1901**, *37* (142), 547–579.

(46) Tanimoto, T. T. *Ibm internal report november 17th*; IBM, 1957.