Published in final edited form as:
J Biomol Screen. 2014 March ; 19(3): 344–353. doi:10.1177/1087057113505325.

Using Weighted Entropy to Rank Chemicals in Quantitative High Throughput Screening Experiments

Keith R. Shockley*

Abstract

Quantitative high throughput screening (qHTS) experiments can simultaneously produce concentration-response profiles for thousands of chemicals. In a typical qHTS study, a large chemical library is subjected to a primary screen in order to identify candidate hits for secondary screening, validation studies or prediction modeling. Different algorithms, usually based on the Hill equation logistic model, have been used to classify compounds as active or inactive (or inconclusive). However, observed concentration-response activity relationships may not adequately fit a sigmoidal curve. Furthermore, it is unclear how to prioritize chemicals for follow-up studies given the large uncertainties that often accompany parameter estimates from nonlinear models. Weighted Shannon entropy can address these concerns by ranking compounds according to profile-specific statistics derived from estimates of the probability mass distribution of response at the tested concentration levels. This strategy can be used to rank all tested chemicals in the absence of a pre-specified model structure or the approach can complement existing activity call algorithms by ranking the returned candidate hits. The weighted entropy approach was evaluated here using data simulated from the Hill equation model. The procedure was then applied to a chemical genomics profiling data set interrogating compounds for androgen receptor agonist activity.

Keywords

quantitative high throughput screening; information theory; concentration response; Tox21

Introduction

High-throughput screening (HTS) experiments have been used extensively in drug discovery initiatives, but they also have been applied to explore alternative targets and diseases with less commercial interest.[1] Remarkably, recent advances in robotics and miniaturization of biological assays have led to an increased volume and quality of HTS data. For example, the

*Address correspondence to: Keith R. Shockley, PhD, Biostatistics Branch, The National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709. Phone: 919-541-3033; Fax: 919-541-4311; shockleykr@niehs.nih.gov..

Supplemental Material

Supplemental Figure S1 provides a heat map of the top 50 chemicals (ranked by *WES*) in the *Ar* assay. Supplemental Figure S2 gives contour plots comparing the range of variance and entropy values across typical Hill equation parameter space found in Tox21 for data simulated with 25% error. Supplemental Figure S3 shows diverse profiles from the *Ar* data set and compares *WES* to AC_{50} model parameter estimates for these examples. R code implementing the weighted entropy approach described here is available at www.niehs.nih.gov/research/atniehs/labs/assets/docs/q_z/wes_example_code.zip.

multi-agency Tox21 partnership between the U.S. Environmental Protection Agency (EPA), the U.S. Food and Drug Administration (FDA), the National Center for Advancing Translational Sciences (NCATS) and the National Toxicology Program (NTP) now employs quantitative high throughput screening (qHTS) to predict the toxicities of drugs, pesticides, suspected carcinogens and other environmental chemicals.[2] In phase I of Tox21, more than 2,800 substances were tested in over 50 assays, including those related to nuclear receptor transactivation and stress response. Data will soon be available for phase II of Tox21, which will test more than 10,000 compounds in a more targeted set of assays. Nevertheless, advancements in data analysis methods are needed to accommodate the technological progress in data generation and fulfill the potential of HTS in compound discovery and testing efforts.

At present it is not clear how to rank candidate hits from qHTS experiments for secondary screening, confirmation studies or prediction modeling. Classification of chemical activity has been based on heuristics[3] and clustering by pattern dissimilarity,[4] but neither strategy relies on statistical parameter estimation or produces a single quantifiable ranking metric. Other approaches to identify candidate hits have been based on the four parameter logistic Hill equation model.[5-8] However, activity calls resulting from such procedures consist of descriptive categorizations (e.g., active, inactive, inconclusive) instead of ranking statistics. In addition, parameter estimates from nonlinear regression models tend to be unreliable due to large standard errors,[9] making them unsuitable ranking measures. To complicate things further, the Hill equation[10] is not appropriate for fitting non-sigmoidal patterns, such as bell shaped curves or more complex profiles, which may nevertheless reflect true concentration-response profiles.

The strategic implementation of a concept from information theory is proposed here to meet these challenges. The average uncertainty in a random variable can be described mathematically using a statistical quantity termed Shannon entropy.[11] Shannon entropy has been used to investigate the complexity of biological sequence data,[12] find differentially methylated regions in the genome[13] and identify non-uniform gene expression patterns in microarray data.[14-16] This same measure can also be used to determine and compare molecular descriptors for different compound classes[17]. Shannon entropy treats all observations as equally reliable, but responses below an empirically derived assay detection limit are not as meaningful as observations lying above this threshold. Equal weighting of all observations in the calculation thus obscures the interpretation of entropy in the qHTS context. We propose a weighted entropy score[18] to characterize chemical profiles, where entropy is calculated as a function of the observed response vector and weights derived from the reliability of each response measurement. Weighted entropy scores (*WES*) can be used to quantify the average activity level of each chemical in a tested library. Chemicals can be ranked according to *WES* as a data driven approach without regard to any pre-specified model structure, or *WES* can be used to rank order hits identified with an existing activity call algorithm.

Here, we describe the concept of entropy and explain the utility of *WES* as a ranking measure for qHTS studies. The usefulness of *WES* is explored within the context of sigmoidal profiles based on the Hill equation logistic model and compared with Shannon

entropy. The performance of the *WES* based ranking procedure is evaluated using a previously simulated data set.[7] Finally, the approach is applied to an experimental qHTS data set generated from phase I of Tox21 that assayed for androgen receptor agonist activity.[6]

Materials and Methods

In this section the application of classical Shannon entropy and a weighted version of Shannon entropy will be described for qHTS experiments. Data sets simulated previously according to the Hill equation[7] will be used to evaluate the performance of these entropy scores across a range of parameter space typical of qHTS experiments. Compounds will be ranked from largest to smallest average activity based on entropy and receiver operating characteristic (ROC) curves[19] will be generated based on these rankings. The area under the curve of ROC curves (AUROC) will evaluate the performance of each approach. To conclude, the weighted entropy approach will be applied to an experimental chemical genomics data set generated within phase I of Tox21.[6]

Description of simulated data

Concentration-response data sets were previously simulated using the Hill equation model,

$$\mathrm{R_i}=\mathrm{R0}+\frac{RMAX-\mathrm{R0}}{\left(1+\left(\frac{2^{log_2 AC50}}{2^{\mathrm{C_i}}}\right)^{\mathrm{SLOPE}}\right)}+\mathrm{error} \quad (1)$$

for fourteen point concentration-response profiles.[7] R_i is expressed as the percentage activity compared to positive control values and represents the normalized response at C_i (test concentration i, expressed in $\log_2$ units). The *error* term is residual error of the model. As shown in Figure 1 for an activator chemical, *RMAX* is the maximal response defining the upper asymptote of the sigmoidal curve, *R0* is the minimal response defining the lower asymptote, AC_{50} is the concentration yielding 50% of the maximal response and *SLOPE* affects the shape of the curve. The detection limit was set to 25% of the positive control activity. The concentrations (C_i) in μM units were based on nuclear receptor activity assay data[6] and consequently set to (4.90×10^{-4}, 2.45×10^{-3}, 1.23×10^{-2}, 2.74×10^{-2}, 6.13×10^{-2}, 1.38×10^{-1}, 3.07×10^{-1}, 6.85×10^{-1}, 1.53, 3.43, 7.66, 17.13, 38.31, 76.63 μM) before $\log_2$ transformation. The values of *RMAX* and AC_{50} were set to (25, 50, 100) and (10^{-3}, 10^{-1}, 10 μM), respectively, which span the range of concentrations (μM) and responses (% positive control) generally observed in qHTS data within Tox21. The *R0* parameter was set to "0" and *SLOPE* was set to "1". Residual errors were modeled as *error* ~ $N(0,\sigma^2)$ for σ = 25%, where σ is expressed as percent of positive control activity.

There were a total of 10,000 simulated substances in each data set, including 2,000 simulated "actives" (*RMAX* = 25%, 50%, or 100% of positive control activity) and 8,000 simulated "inactives" (*RMAX* = 0%). These simulated data sets were used to evaluate the performance of the entropy measures (see "Shannon entropy" and "Weighted entropy" sections below). First, for a given ranked list size, simulated profiles were ranked by entropy score (from highest entropy to lowest entropy). Then, the fraction of the simulated actives

that were correctly identified was compared to the fraction of simulated actives that were falsely classified. Ranked list sizes ranged from 1 to 10,000 (the total number of simulated profiles in a simulated data set).

Description of nuclear receptor agonist data sets

Normalized chemical genomics data for NTP compounds evaluated using androgen receptor (*Ar*) and estrogen receptor (*Esr1*) agonist assays were obtained from a previously published study.[6] As described more extensively in the Results section, the *Esr1* assay data was used to adjust *Ar* entropy scores for *Esr1* activity. A total of 1,408 compounds were assayed in 14 concentrations ranging from 4.90×10^{-4} μM to 76.63 μM for each experimental assay. Raw plate reads for each concentration were normalized using the positive and negative control wells (positive values for activation and negative values for inhibition). This data was then corrected for row, column, and plate effects by a pattern correction algorithm based on linear interpolation.[3] The final normalized response measures R_i for test concentration C_i can be regarded as expressing the percentage activity relative to the change generated by the positive control compared to negative controls. Activity calls and Hill equation parameter estimates were obtained for each compound as described previously.[7]

Shannon entropy

Each chemical substance will produce a set of outcomes described by a response vector **R** = (R_1, R_2, …, R_N) for N concentrations, where R_i corresponds to the observed response at the *i*th concentration, C_i. The relative response at C_i can be defined as

$$p_i = |R_i| / \Sigma_{i=1}^{N} |R_i| \quad (2)$$

where $p_i \geq 0, \Sigma_{i=1}^{N} \; p_i = 1$, and $|R_i|$ stands for the absolute value of response R_i. A similar expression has been used for DNA microarray data to calculate the relative expression of a gene from a vector of expression levels.[15] Relative response values **p** = (p_1, .., p_N) represent a probability mass distribution based on the extent of observed responses across the N tested concentrations where the sign of each R_i may be positive or negative for activation or inhibition, respectively. Because $p_i \geq 0$ for all *i*, Eq. (2) describes the extent of response at each concentration but does not necessarily describe the complexity of the concentration-response pattern as determined by the sign of R_i values in **R**.

Entropy is a concept from information theory that can be used to quantify the average amount of information (or uncertainty) in **R** with probabilities p_1, …, pN.[11] The entropy of R_i, or surprisal of the *ith* event, is defined as

$$h(R_i) = -log_b p_i \quad (3)$$

where the base of the logarithm determines the units of information. A base of 2 is used here, so that the units are in bits. The function $h(R_i)$ increases when p_i moves toward zero and goes toward zero when p_i approaches one. The average Shannon entropy over the measured concentration range is denoted by *H* and given by the expression,

$$H = -\sum_{i=1}^{N} p_i log_2 p_i \quad (4)$$

where the convention of $0\log_2 0 = 0$ is used, since $\lim_{p \to 0} p_i log_2 p_i = 0$. *H* ranges from zero for chemicals with $|R_i| > 0$ at only one concentration level to $\log_2(N)$ for chemicals responding equally at all concentrations ($|R_1| = |R_2| = \ldots = |R_N|$). Smaller values of *H* indicate that a greater mass of the probability distribution is limited to fewer concentrations, while larger values of *H* values imply that the response distribution is more uniform across concentrations levels.

Weighted entropy

Eq. (4) does not take into account uncertainties in response measurements. As a result, Shannon entropy scores can be large for profiles in which all of the observed response values fall below the assay detection limit (see Table 1 as explained in the Results section). However, observed responses below the detection limit are less reliable than observations exceeding the threshold of detection. Weighted entropy measures can be formulated to take into account the extent of R_i relative to the detection limit of the assay. We use a weighting scheme that reduces the associated component of entropy in direct proportion to the reliability of the observed response. Here, the weighted entropy score (*WES*) of a substance across N concentration levels is given by the expression

$$WES = -\sum_{i=1}^{N} w_i p_i log_2 p_i \quad (5)$$

where *WES* is always greater than or equal to zero and $0\log_2 0 = 0$ as described earlier for Eq. (4). The weights w_i will be defined as

$$w_i = \begin{cases} 1, & \text{for} |R_i| \geq DetLim \\ |R_i|/DetLim, & \text{for} |R_i| < DetLim \end{cases} \quad (6)$$

where *DetLim* represents the assay detection limit. Detection limits equal to about 25% of the positive control activity are typical within Tox21 efforts and provide reasonable estimates of the true assay detection limit in most studies. Smaller values of *WES* indicate a greater response density of detectable response observations at fewer concentrations or more uniform (but unreliable) response measurements. Larger values of *WES* indicate that the response distribution is comprised of a greater proportion of detectable responses across concentrations levels. Based on the weighted scheme shown in (6) above, *WES* will always be less than or equal to *H* for any given profile.

Results

Calculating entropy for illustrative profiles

Shannon entropy quantifies the average surprisal (or comparative likelihood of a response) across tested concentration levels, regardless of whether the concentration intervals with the

largest responses are contiguous (see Eq. (4)). The weighted entropy procedure used here also describes the average probability mass of observed responses, but *WES* scores adjust each entropy component according to the reliability of the underlying response measurements (compare Eq. (4) and Eq. (5)). In both cases, chemicals with larger entropy scores will be ranked higher in an ordered list of tested chemicals. However, larger Shannon entropy implies increased variation in response across concentrations, while larger weighted entropy implies greater average activity across concentrations. Differences between these two ranking measures are described below in greater detail by considering the six example profiles in Table 1.

Each of the illustrative profiles in Table 1 consists of observed responses at four concentration levels (N=4) from a hypothetical assay with a detection limit of 25%. *Chemical-1* exhibits 75% of the positive control response at each tested concentration so that $R_1=R_2=R_3=R_4=75$. In this case, the relative response p_i does not change across concentration levels (i.e., $p_1=p_2=p_3=p_4$) and the surprisal h_i remains the same for all tested concentrations (i.e., $h_1=h_2=h_3=h_4$). Because R_i is always greater than the assay detection limit, the Shannon entropy and *WES* scores for this profile are both equivalent to the maximal entropy ($H_1 = WES_1 = \log_2 N = 2.00$). *Chemical-2* also displays equal responses across the four concentrations, but in this case R_i is always below the assay detection limit. The Shannon entropy for *Chemical-2* is identical to the Shannon entropy for *Chemical-1*, because Shannon entropy does not consider the detection limit. In contrast, the weighted entropy score for *Chemical-2* ($WES_2 = 0.27$) is considerably lower than the calculated Shannon entropy ($H_2 = 2.00$).

Chemical-3 and *Chemical-4* have the same relative response levels at each concentration, which exceed the limit of detection. However, while *Chemical-3* is an activator, with increasing response for increasing concentration, the profile produces the same entropy as the oscillatory response profile represented by *Chemical-4* ($H_3 = H_4 = WES_3 = WES_4 = 1.81$). This comparison illustrates that entropy is not linked to the complexity (or sign) of the R_i values comprising the response profile, since the concentration-response pattern of *Chemical-4* is more complex (exhibiting sign changes) than the response pattern of *Chemical-3* (no sign changes). *Chemical-5* has a single response in the detectable region ($R_4 = 85\%$), such that most of the probability mass is concentrated at the last concentration. The weighted entropy of this profile ($WES_5 = 0.25$) is lower than the Shannon entropy of the profile ($H_5 = 0.85$), and both quantities are considerably smaller than the maximal entropy of $\log_2 N$ for *Chemical-1*. Finally, *Chemical-6* is a condition of minimal entropy, in which only one response ($R_4 = 100\%$) is detected. By definition, $H_6 = 0$ and $WES_6 = 0$, since every surprisal h_i will be zero in this case according to Eq. (2).

Variance across response measurements can also be used to characterize profiles.[14] Here, variance represents a distance between response values, whereas entropy is a measure of probability mass across a response profile. Variances are greatest when $|RMAX|$ is relatively large and AC_{50} lies between the highest and lowest tested concentrations. In contrast, entropy scores are greatest when $|RMAX|$ is large and AC_{50} values are small (see Supplemental Figure S1). In Table 1, *Chemical-1* and *Chemical-2*, with uniform response

distributions, have zero variance (i.e., $\sigma_1^2=\sigma_2^2=0.00$) compared to maximal Shannon entropy ($H_1 = H_2 = 2.00$) and disparate *WES* scores ($WES_1 = 2.00$, $WES_2 = 0.27$). The oscillating response pattern of *Chemical-4* leads to much greater variance across R_is $\left(\sigma_4^2=5,491.67\right)$ than the strictly increasing response pattern of *Chemical-3* ($\sigma_3^2=1,158.33$) even though *H* and *WES* are equivalent for these profiles. While *Chemical-5* has a smaller unweighted variance than weighted variance in this case ($\sigma_{5,unweighted}^2<\sigma_{5,weighted}^2$), the profile has higher Shannon entropy than produced by the *WES* value ($H_5 > WES_5$). *Chemical-6* only responds at the last tested concentration, but is ranked 2 out of 6 based on unweighted variance compared to 6 out of 6 based on Shannon entropy. While the weighted variance cannot be calculated (see Table 1), *Chemical-6* also ranks last (6 out of 6) based on weighted entropy.

Evaluating entropy scores based on simulated Hill model data

To explore entropy across typical AC_{50} and $|RMAX|$ parameter space, we generated fourteen point concentration-response curve data from Eq. (1) using the statistical software R.[20] The values of AC_{50} ranged from 10^{-5} to 10^{2} μM and were incremented by 0.1 units on the $\log_2$ transformed scale, while $|RMAX|$ was varied from 0% to 150% and incremented by 1% units. The detection limit (*DetLim*) was set to 25%. *H* and *WES* values were calculated for each simulated profile according to Eq. (4) and Eq. (5), respectively. The results are displayed as response surface diagrams for *H* (Figure 2A) and *WES* (Figure 2B). Visual inspection of these plots illustrates that *H* is constant across $|RMAX|$ for a fixed AC_{50} while $|RMAX| > 0$. In contrast, *WES* is very small, but greater than zero, when $0 < |RMAX| \leq DetLim$ for all AC_{50} and increases as a function of $|RMAX|$ for a fixed AC_{50} when $|RMAX| > DetLim$.

The performance of *H* and *WES* entropy scores as ranking statistics was evaluated using previously simulated data.[7] Briefly, simulated profiles were produced according to the Hill equation model with *R0*=0, *SLOPE*=1 and residual errors modeled as $\varepsilon \sim N(\mu=0, \sigma_i^2=25\%)$ for an assay detection limit of 25%. Nine Hill equation parameter configurations were used to span the scope of observations in typical qHTS studies, where AC_{50} was set to (10^{-3}, 10^{-1} and 10 μM) and $|RMAX|$ was set to (25%, 50% and 100%). These profiles were ranked by the Hill Equation estimate for AC_{50}, *H* or *WES* and the fraction of simulated actives that were correctly identified was compared to the fraction of simulated actives that were falsely classified for different ranked list sizes (see Materials and Methods). The performance of AC_{50} estimates, *H* and *WES* scores to predict true activity was assessed using the area under the ROC curve (AUROC) for different numbers of profile sample sizes (N) (Table 2). AUROC ranges from 0 to 1, where values of 0.5, 0.75 and 0.9 correspond to random, good and excellent performance, respectively. *WES* outperforms *H* and estimated AC_{50} in every case examined here. AC_{50} performed poorly for every tested scenario. The ability of *H* to discriminate between active and inactive substances was poor when $|RMAX|$ = 25% (at the assay detection limit) or when AC_{50} = 10 μM (activity only at high tested concentrations). In contrast, *WES* starts to become problematic only as $|RMAX|$ approaches the assay detection limit (25%). More tested concentrations generally produced greater AUROC values for both entropy measures. AUROC for *WES* was good or excellent in most cases, but performance

was reduced when AC_{50} = 10 μM and $|RMAX|$ = 25%. Conversely, AUROC for Shannon entropy was always low when $|RMAX|$ = 25% or when AC_{50} = 10 μM.

Ranking profiles from androgen receptor agonist assay data

The *Ar* is a steroid hormone receptor and member of the nuclear receptor superfamily of transcription factors.[21] We examined *Ar* agonist data expressed as the ratio of 460- to 530-nm emission fluorescence intensities from GeneBLAzer® β-lactamase HEK 293T cell lines (Carlsbad, CA) as described by Huang et al.[6] Weighted entropy scores were computed for this data and compounds were ranked by their *WES* score. These rankings are based strictly on the data, apart from any pre-specified concentration response model form, including directionality of response. Many of the highest ranking *Ar* chemicals were activators (increasing response with increasing concentration), but some compounds exhibited decreasing activity with increasing concentration, possibly due to cell toxicity (see Supplemental Figure S2).

The top 25 most informative substances based on *WES* (*WES* ≥ 2 bits) are shown in Table 3. Progesterone was duplicated in the compound library and appears twice in Table 3, but most substances were present only once in the experiment. A total of 20 of these 25 compounds show an “activator” response, while three compounds were classified as “inhibitors” (decreasing activity with increasing concentration) and two compounds were classified as “potent inhibitors” (activity at the lowest tested concentration). The activators include the steroids Progesterone (represented twice in the assay), Fluoxymestrone, Prednisone, corticosterone, medroxyprogesteroneacetate, androstenedione, 4-androstenedione, dexamethasone, beta testosterone, methyl testosterone and 17beta-estradiol. Progesterone and corticosterone are known *Ar* agonists in this compound set, as described previously.[6] The activators also include benzo(k)fluoranthene and Croton oil, which has been found to cause ligand independent activation of the androgen receptor through the action of the phorbol ester component.[22] The compound DDD (6-hydroxyl-2-naphthyl disulfide) exhibits inhibitory activity at lowest concentrations but shows agonist activity at the two highest concentrations. Conversely, daunomycin HCL and adriamycin hydrochloride exhibit activator responses before inhibition at higher concentrations (see Supplemental Figure S2).

Substances evaluated in qHTS experiments are sometimes ranked by AC_{50} as a measure of potency. However, the standard error of the log_2AC50 estimates from Eq. (1) was large in most cases (data not shown). As shown in Table 3, substances ranked in the top 25 (out of all 1408 tested substances) based on *WES* were ranked from 6 to 1394 based on AC_{50}. Supplemental Figure S3 shows diverse response profiles from the *Ar* data set; some substances are ranked highly based on *WES* and AC_{50} estimates while other substances show very different rankings based on these two measures. Only 9 compounds were ranked in the top 250 compounds by *WES* and AC_{50} (data not shown) and there was no correlation between ranks when comparing all 1408 compounds ranked by *WES* and AC_{50} (Spearman rho = 0.0046, p ~ 0.86).

Compounds eliciting an agonistic *Ar* response may also stimulate Esr1.[23-25] A marginally significant correlation was discovered when comparing the ranks based on WES_{Ar} to ranks based on WES_{Esr1} (Spearman rho = 0.045, p~0.09). To find informative *Ar* profiles with

limited activity in the *Esr1* assay, a test statistic was constructed based on the difference in weighted entropy ($\Delta_{WES} = WES_{Ar} - WES_{Esr1}$). Substances producing greater activity in the *Ar* assay compared to the *Esr1* assay will have larger Δ_{WES} scores. Figure 3 presents the response profiles for the top 8 most informative substances based on Δ_{WES}. Compounds not represented in Figure 3 with large Δ_{WES} scores ($\Delta_{WES} \geq 2$ bits) include medroxylprogesteroneacetate, androstenedione, 4-androsteinedione and DDD (6-hydroxy-2-naphthyl disulfide) with Δ_{WES} scores of 2.46, 2.33, 2.30 and 2.04 bits, respectively. The top ten compounds with strong negative Δ_{WES} scores indicating specificity for *Esr1* include ethinyl estradiol, diethylstilbetrol, zearalanol, 1-bromopropane, ethylenediamine, Alachlor, zearalenone (represented twice), bisphenol A and 17beta-estradiol (data not shown).

Discussion

The concept of Shannon entropy is foundational to information theory, a field which intersects bioinformatics, electrical engineering, computer science, statistical physics, mathematics and economics.[26, 27] There is a known relationship between Shannon entropy and thermodynamic entropy in statistical mechanics, such that entropy can be used as a measure of the molecular disorder of a system.[28] Applications of entropy in communication theory provide a way to quantify the amount of information associated with a received message.[11] The amount of information gained by the receiver depends on the probability that a message (or event) will occur. The surprise associated with the *ith* event is a function of the underlying probability of the event; more information is received for transmitted messages that are less probable. Thus, Shannon entropy describes the uncertainty associated with an ensemble of events $\times = \{x_1, x_2, \ldots, x_i\}$.

A weighted form of Shannon entropy is proposed here as a computational technique to guide the interpretation of the large volume of data generated in qHTS experiments. Weighted entropy characterizes a series of events by their probabilities of occurrence and associated weights.[18] Like Shannon entropy, weighted entropy provides a quantitative description of a response profile based on the probability mass function estimated from the responses observed across the tested concentrations. However, the weighted entropy measure described here, termed the Weighted Entropy Score (*WES*), differs from classical Shannon entropy by its ability to take into account the reliability of the underlying response measurements. The combination of these properties allows *WES* to rank profiles in order of importance from those with maximal entropy (full response at the lowest tested concentration) to less prominent patterns (a change in response across concentrations levels) to minimal entropy profiles (no observed activity at any tested concentration).

Conventional HTS assays for hit discovery are typically run at a compound concentration of 10 μM or less.[29] As shown in Table 2, classical Shannon entropy scores had a good ability to rank simulated curves based on the Hill equation when $|RMAX| \geq 50\%$. However, Shannon entropy performed poorly when AC_{50} was 10 μM or when $|RMAX|$ was 25% (when maximal response measurements are near the assay detection limit). Table 2 demonstrates that the *WES* approach outperformed Shannon entropy under every scenario considered here and was even useful when AC_{50} was 10 μM. Accordingly, chemicals with larger *WES* scores have more response measurements above the detection limit across the range of tested

concentrations, while chemicals with lower *WES* scores have reliably detectable responses at fewer concentrations.

As shown in Table 2, overall performance of the entropy scores generally improves with the number of tested concentrations for both Shannon entropy and *WES*. The weighted entropy metric performed well (AUROC ≥ 0.75) in most scenarios and outperformed Shannon entropy in every instance examined here, across all numbers of concentrations tested, N, and Hill function parameter configurations (Table 2). However, ranking substances based on AC_{50} estimates derived from fits to the Hill Equation was always resulted in poor performance. Other studies have demonstrated that the Hill Equation model parameters can be unreliable estimates due to large variations in observed responses at higher doses, irregular dose spacing or data collected over an incomplete dose range[5,8,31]. In contrast to AC_{50}, *WES* is based strictly on the observed responses and never relies on estimates of the behavior of a nonlinear function outside of the tested concentration range. The performance of weighted entropy as a ranking statistic is greater for substances with smaller AC_{50} values and larger |*RMAX*| values. Concentration response profiles with an AC_{50} of 10 μM have only one clearly defined asymptote and produce less reliable model fits than profiles with lower AC_{50} values.[7] Although performance was better than random classification, it is not surprising that *WES* scores had the most difficulty discriminating true actives from false positives when AC_{50} was 10 μM and |*RMAX*| was 25%. Such “marginally active” substances have relatively small entropy scores compared to substances with much smaller AC_{50} values and greater |*RMAX*| values (see Figure 2).

Ranking substances by *WES* scores can be used as a data driven approach to find the most prominent response patterns in a tested set without imposing a pre-determined heuristic scheme or model structure. This strategy will identify profiles that reliably fit a Hill model framework (e.g., Figure 1) as well as non-sigmoidal patterns that may reflect real, but complex, response patterns that do not fit a pre-specified model structure. Complex response patterns may be indicators of complex biological or chemical processes[30] or, alternatively, “false positives” in the presence of uncontrolled factors such as contamination, signal flare and carryover effects. Entropy based scores do not distinguish the directions of response (i.e., do not discriminate between activators versus inhibitors) and the complexity of some high entropy nonmonotonic response patterns could be difficult to interpret. Therefore, in practice it may be advantageous to apply the ranking procedure to a list of hits identified with an existing model-based activity call algorithm.

Entropy scores hold other useful properties for characterizing profiles generated from qHTS experiments. Missing data is easily accommodated into the ranking framework described here, since missing data will simply reduce the maximal possible entropy determined by $\log_2 N$. Profile ranks based on entropy may differ substantially from profile ranks based on variance across the response observations. Entropy is computed as a function of the probability mass at each concentration level rather than distances between response levels such as variance (see Table 1). Entropy and variance therefore capture different aspects of the profile signal (Figure S1). However, entropy is preferred to variance for compound ranking since only entropy scores will provide equivalent measures for profiles with the same response distributions irrespective of curve complexity. Entropy can also discriminate

between uniform profiles corresponding to detectable response and uniform profiles below the assay detection limit (see Table 1). Finally, *WES* scores provide a convenient metric that can be readily extended to compare outcomes from different experiments, as demonstrated by the ability of *WES* to identify chemicals with selective activity in *Ar* compared to *Esr1* (see Figure 3).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

I thank Dr. Raymond Tice (Biomolecular Screening Branch, NIEHS) and Dr. Grace Kissling (Biostatistics Branch, NIEHS) for reviewing the manuscript and providing helpful suggestions. The *Ar* and *Esr1* agonist data sets were obtained from the Chemical Effects in Biological Systems (CEBS) database[32] under accession numbers 013-00004-0002-000-0 and 013-00004-0003-000-1, respectively.

Funding

This work was supported [in part] by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (ZIA ES102865).

References

1. Malo N, Hanley JA, Cerquozzi S, et al. Statistical practice in high-throughput screening data analysis. Nat. Biotechnol. 2006; 24:167–175. [PubMed: 16465162]
2. Collins FS, Gray GM, Bucher JR. Toxicology. Transforming environmental health protection. Science. 2008; 319:906–907. [PubMed: 18276874]
3. Inglese J, Auld DS, Jadhav A, et al. 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; 103:11473–11478. [PubMed: 16864780]
4. Zhang X, Newsted JL, Hecker M, et al. Classification of chemicals based on concentration-dependent toxicological data using ToxClust. Enviorn. Sci. Technol. 2009; 43:3926–3932.
5. Parham F, Austin C, Southall N, et al. Dose response modeling of high-throughput screening data. J. Biomol. Screen. 2009; 14:1216–1227. [PubMed: 19828774]
6. Huang R, Xia M, Cho MH, et al. Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environ. Health Perspect. 2011; 119:1142–1148. [PubMed: 21543282]
7. Shockley KR. A three-stage algorithm to make toxicologically relevant activity calls from quantitative high throughput screening data. Environ. Health Persp. 2012; 120:1107–1115.
8. Lim C, Sen PK, Peddada SD. Robust analysis of high throughput screening (HTS) assay data. Technometrics. 2013; 55:150–160. [PubMed: 23908557]
9. Peddada SD, Haseman JK. Analysis of nonlinear regression models: a cautionary note. Dose Response. 2005; 3:342–352. [PubMed: 18648618]
10. Hill AV. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J. Physiol. 1910; 40:4–7.
11. Shannon CE. A mathematical theory of communication. Bell Syst. Techn. J. 1948; 27:1–55.
12. Machado JAT. Shannon entropy analysis of the genome code. Math. Prob. Eng. 2012 Article ID 132625.
13. Zhang Y, Liu H, Lv J, et al. QDMR: a quantitative method for identification of differentially methylated regions by entropy. Nucl. Acids Res. 2011; 39:e58. [PubMed: 21306990]
14. Fuhrman S, Cunningham MJ, Wen X, et al. The application of Shannon entropy in the identification of putative drug targets. BioSystems. 2000; 55:5–14. [PubMed: 10745103]

15. Schug J, Schuller W-P, Kappen C, et al. Promoter features related to tissue specificity as measured by Shannon entropy. Gen. Biol. 2005; 6:R33.
16. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–112. [PubMed: 19295514]
17. Wassermann WA, Nisius B, Vogt M, et al. Identification of descriptors capturing compound class-specific features by mutual information analysis. J. Chem. Inf. Model. 2010; 50:1935–1940. [PubMed: 20961115]
18. Guiasu S. Weighted entropy. Rep. Math. Physics. 1971; 2:165–179.
19. Fawcett T. An introduction to ROC analysis. Patt. Recog. Lett. 2006; 27:861–874.
20. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2012.
21. Mangelsdorf DJ, Thummel C, Beato M, et al. The nuclear receptor superfamily: The second decade. Cell. 1995; 1995; 83:835–839. [PubMed: 8521507]
22. Darne C, Veyssiere G, Jean C. Phorbol ester causes ligand-independent activation of the androgen receptor. Eur. J. Biochem. 1998; 256:541–549. [PubMed: 9780230]
23. Kousteni S, Bellido T, Plotkin LI, et al. Nongenotropic, sex-nonspecific signaling through the estrogen or androgen receptors: dissociation from transcriptional activity. Cell. 2001; 104:719–730. [PubMed: 11257226]
24. Richter CA, Taylor JA, Ruhlen RL, et al. Estradiol and Bisphenol A stimulate androgen receptor and estrogen receptor gene expression in fetal mouse prostate mesenchyme cells. Environ. Health Persp. 2007; 115:902–908.
25. Svensson J, Moverare-Skrtic S, Windahl S, et al. Stimulation of both estrogen and androgen receptors maintains skeletal muscle mass in gonadectomized male mice but mainly via different pathways. J. Mol. Endocrinol. 2010; 45:45–57. [PubMed: 20435684]
26. Cover, TM.; Thomas, JA. Elements of information theory. John Wiley & Sons; New York: 1991. p. 1-49.
27. Nalbantoglu OU, Russel DJ, Sayood K. Data compression concepts and algorithms and their applications to bioinformatics. Entropy. 2009; 12:34–52. [PubMed: 20157640]
28. Chakrabarti CG, Chakrabarty I. Boltzmann-Shannon entropy: Generalization and application. Mod. Phys. Lett. B. 2006; 20:1471–1479.
29. Hughes JP, Rees S, Kalindjian SB, et al. Principles of early drug discovery. Brit. J. Pharmacol. 2011; 162:1239–1249. [PubMed: 21091654]
30. Conolly RB, Lutz WK. Nonmonotonic dose-response relationships: mechanistic basis, kinetic modeling, and implications for risk assessment. Toxicol. Sci. 2004; 77:151–157. [PubMed: 14600281]
31. Lim C, Sen PK, Peddada SD. Accounting for uncertainty in heteroscedasticity in nonlinear regress. J. Stat. Plan. Infer. 2012; 142:1047–1062.
32. Waters M, Stasiewicz S, Merrick BA, et al. CEBS - Chemicals Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucl. Acids Res. 2008; 36:D892–D900. [PubMed: 17962311]

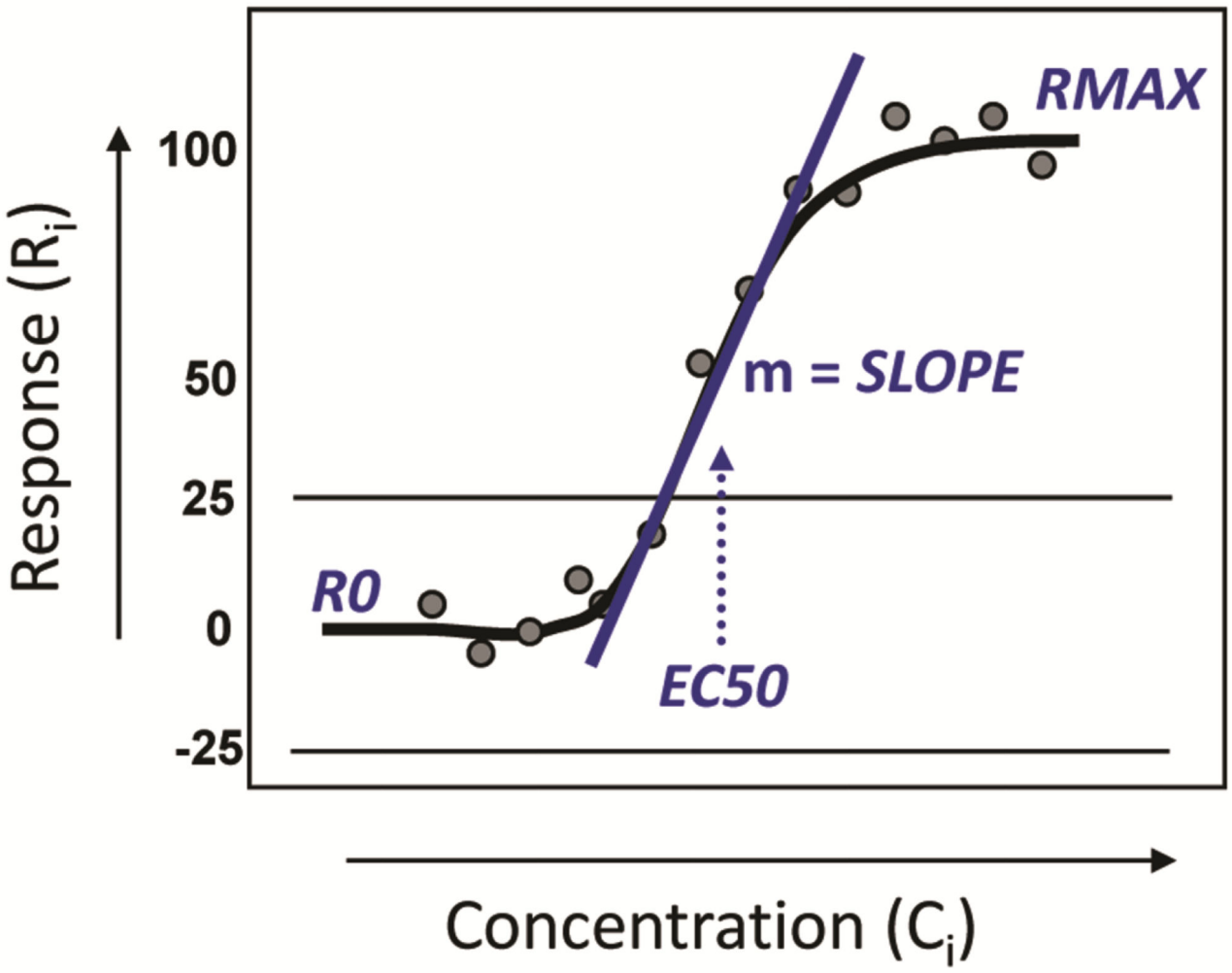

Figure 1.

A depiction of a Hill model curve for an activator. The assay detection limits are shown as horizontal lines. The model terms are described in the Materials and Methods section.

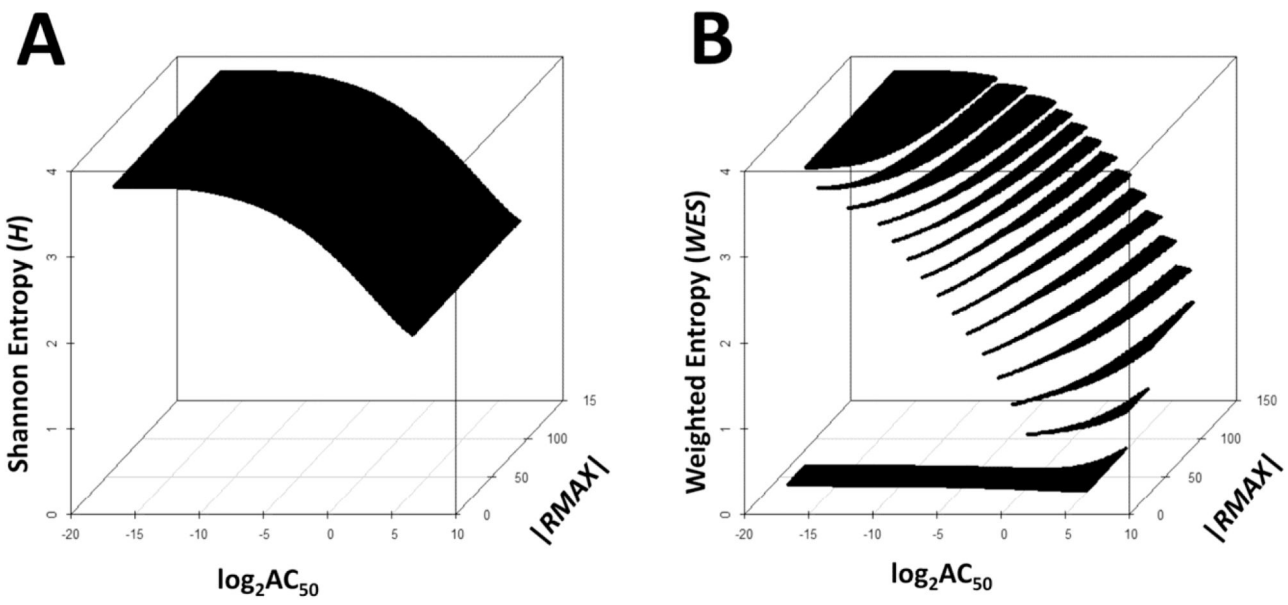

Figure 2.

Entropy of the four-parameter Hill Equation model. (A) Shannon entropy and (B) weighted entropy response surfaces for the Hill equation across a range of AC_{50} and |*RMAX*| with *R0* =0 and *SLOPE* = 1. Shannon entropy increases with decreasing AC_{50}, independent of |*RMAX*| while the weighted entropy increases as a function of decreasing AC_{50} and increasing |*RMAX*|.

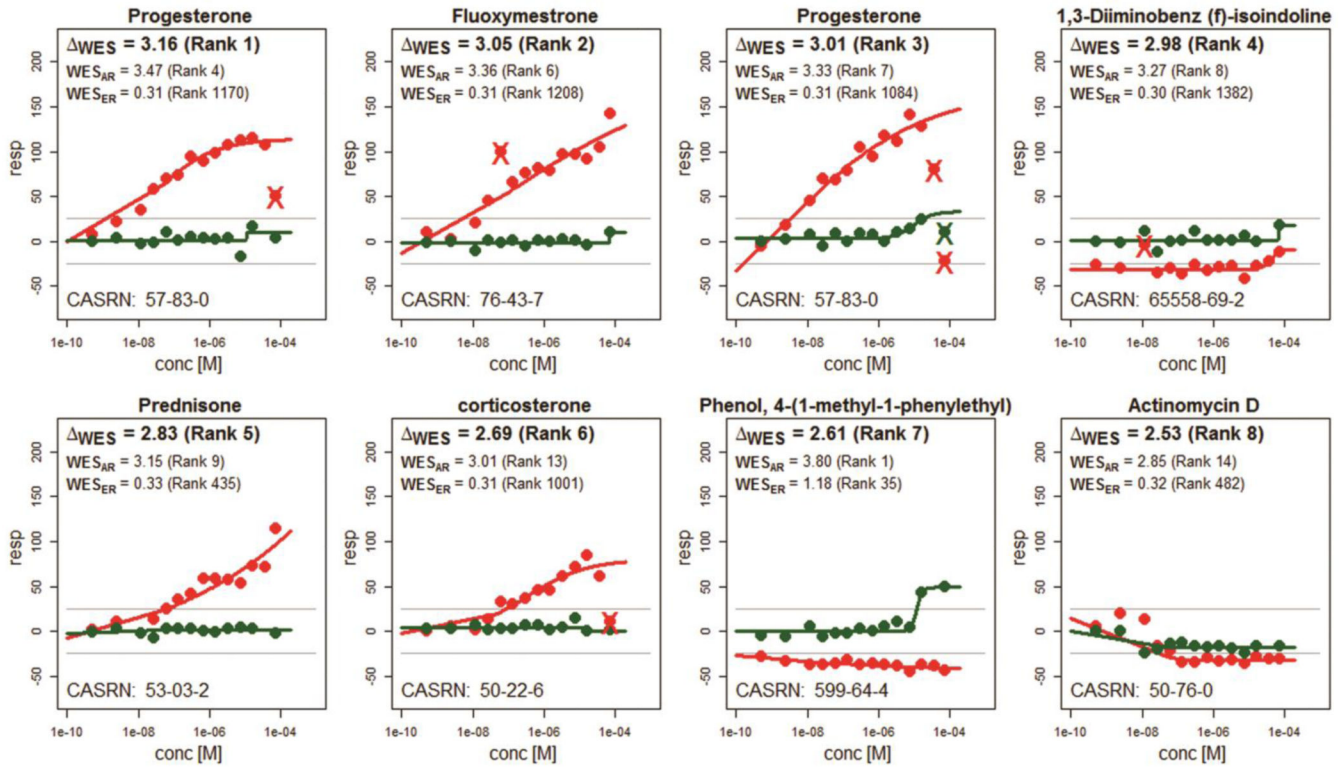

Figure 3.

Example response profiles corresponding to the top 8 chemicals from Table 2 ranked by Δ_{WES}=WES_{Ar}-WES_{Esr1}. The data for *Ar* and *Esr1* profiles are shown in red and green, respectively. *WES* scores are calculated independently for each profile. The ranking of each *WES* score out of all 1408 tested compounds is given in parentheses. The concentration-response data is used with permission from *Environmental Health Perspectives*.

Table 1

Entropy calculations for illustrative profiles with four concentration levels[a]

Profile	Responses (R_i)				Probabilities (p_i)				Surprisals[b] (h_i)				Entropy	Variance
	i=1	*i=2*	*i=3*	*i=4*	*i=1*	*i=2*	*i=3*	*i=4*	*i=1*	*i=2*	*i=3*	*i=4*		
Chemical-1	75	75	75	75	0.25	0.25	0.25	0.25	2.00	2.00	2.00	2.00	2.00	0.00
													2.00	**0.00**
Chemical-2	10	10	10	10	0.25	0.25	0.25	0.25	2.00	2.00	2.00	2.00	2.00	0.00
													0.27	**0.00**
Chemical-3	30	30	70	100	0.13	0.13	0.30	0.43	2.94	2.94	1.72	1.20	1.81	1,158.33
													1.81	**1,158.33**
Chemical-4	−30	30	−70	100	0.13	0.13	0.30	0.43	2.94	2.94	1.72	1.20	1.81	5,491.67
													1.81	**5,491.67**
Chemical-5	5	5	5	85	0.05	0.05	0.05	0.85	4.32	4.32	4.32	0.23	0.85	1,600.00
													0.25	**2,666.67**
Chemical-6	0	0	0	100	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	2,500.00
													0.00	**undefined[c]**

[a] Weighted entropies and variances are shown in bold, where the weights are based on Eq. (5) in the Methods with a detection limit of 25%.

[b] The surprisal at concentration *i* is given as $h_i = -\log_2 p_i$.

[c] In this case, both the numerator and denominator of weighted variance are zero (see the calculation of weighted variance in the Supplemental Material)

Table 2

Areas under ROC curves for different profile sizes (*N*)

True AC_{50}	True |RMAX|	$AUROC_{Estimated\ AC50}$[a]			$AUROC_{Shannon\ entropy}$			$AUROC_{WES}$		
		4	**9**	**14**	**4**	**9**	**14**	**4**	**9**	**14**
0.001	25	0.52	0.50	0.51	0.55	0.56	0.58	0.60	0.78	0.84
	50	0.53	0.50	0.53	0.60	0.78	0.87	0.72	0.98	1.00
	100	0.53	0.56	0.60	0.68	0.92	0.99	0.79	1.00	1.00
	100	0.53	0.56	0.60	0.68	0.92	0.99	0.79	1.00	1.00
0.1	25	0.53	0.50	0.51	0.55	0.57	0.56	0.58	0.72	0.77
	50	0.54	0.54	0.52	0.63	0.71	0.70	0.72	0.94	0.97
	100	0.53	0.55	0.54	0.72	0.81	0.83	0.86	1.00	1.00
10	25	0.54	0.48	0.46	0.53	0.53	0.51	0.54	0.57	0.58
	50	0.56	0.46	0.42	0.57	0.55	0.53	0.59	0.68	0.75
	100	0.55	0.38	0.36	0.62	0.52	0.41	0.70	0.83	0.92

[a] Based on the profiles (out of 10,000 simulated curves) for which an *AC50* value can be estimated according to Shockley (2012)[7]. The number of estimable *AC50* values was calculated from 5,950 to 6,307 profiles (N=4); 9,955 to 9,968 profiles (N=9); and 9,997 to 10,000 profiles (N=14).

Table 3

Top twenty-five substances ranked by weighted entropy in Tox21 Phase I *Ar* agonist assay data[a]

CASRN	NAME	WES score	AC_{50}[b]	Activity Call[c]
599-64-4	Phenol, 4−(1-methyl-1-phenylethyl)−	3.8 (1)	8.44×10^{-44} (6)	POT INH
58-22-0	beta testosterone	3.66 (2)	9.57×10^{-22} (24)	ACT
521-18-6	5alpha dihydrotestosterone	3.52 (3)	6.18×10^{-9} (149)	ACT
57-83-0	Progesterone	3.47 (4)	2.88×10^{-8} (241)	ACT
58-18-4	methyl testosterone	3.45 (5)	1.19×10^{-26} (16)	ACT
76-43-7	Fluoxymestrone	3.36 (6)	1.19×10^{-7} (277)	ACT
57-83-0	Progesterone	3.33 (7)	1.67×10^{-9} (94)	ACT
65558-69-2	1,3-Diiminobenz (f)-isoindoline	3.27 (8)	4.31×10^{-5} (912)	POT INH
53-03-2	Prednisone	3.15 (9)	5.66×10^{21} (1394)	ACT
63-05-8	androstenedione	3.08 (10)	2.28×10^{-7} (311)	ACT
63-05-8	4-androstenedione	3.04 (11)	2.58×10^{-7} (314)	ACT
434-07-1	oxymetholone	3.04 (12)	1.23×10^{-7} (278)	ACT
50-22-6	corticosterone	3.01 (13)	3.32×10^{-7} (328)	ACT
50-76-0	Actinomycin D	2.85 (14)	2.41×10^{-8} (229)	INH
71-58-9	medroxyprogesteroneacetate	2.78 (15)	1.26×10^{-4} (1173)	ACT
207-08-9	benzo(k)fluoranthene	2.52 (16)	3.8×10^{-4} (1202)	ACT
8001-28-3	Croton oil	2.5 (17)	9.17×10^{-7} (377)	ACT
6088-51-3	DDD (6-hydroxy-2-naphthyl disulfide)	2.35 (18)	3.85×10^{-5} (855)	ACT
20830-75-5	digoxin	2.31 (19)	4.63×10^{-7} (343)	INH
50-02-2	dexamethazone	2.31 (20)	3.12×10^{-7} (325)	ACT
23541-50-6	daunomycin HCL	2.21 (21)	4.74×10^{-6} (555)	ACT
25316-40-9	adriamycin, hydrochloride	2.07 (22)	2.05×10^{-6} (455)	ACT
2437-29-8	malachite green oxalate	2.05 (23)	7.26×10^{-7} (368)	INH
50-28-2	17beta-estradiol	2.01 (24)	6.73×10^{-7} (355)	ACT
52417-22-8	9-aminoacridine, monohydrochloride, monohydrate	2.00 (25)	1.00×10^{-5} (643)	ACT

[a] The rank of each statistic (*WES* or *AC50*) is shown in parentheses.

[b] The AC_{50} values are in units of μM.

[c] Activity call based on the three-stage algorithm[7] using a significance threshold of $p < 0.05$. ACT=activator; POT INH=potent inhibitor; INH=inhibitor.