Enhanced HTS Hit Selection via a Local Hit Rate Analysis

Bruce A. Posner,[†] Hualin Xi,[‡] and James E. J. Mills*[,§]

PGRD Groton Laboratories, Pfizer Inc., Groton, Connecticut 06340, PGRD Computational Sciences CoE, Pfizer Inc., Cambridge, Massachusetts 02139, and PGRD Sandwich Laboratories, Pfizer Inc., 32 Ramsgate Road, Sandwich, Kent CT13 9NJ, Great Britain

Received March 29, 2009

The postprocessing of high-throughput screening (HTS) results is complicated by the occurrence of false positives (inactive compounds misidentified as active by the primary screen) and false negatives (active compounds misidentified as inactive by the primary screen). An activity cutoff is frequently used to select "active" compounds from HTS data; however, this approach is insensitive to both false positives and false negatives. An alternative method that can minimize the occurrence of these artifacts will increase the efficiency of hit selection and therefore lead discovery. In this work, rather than merely using the activity of a given compound, we look at the presence and absence of activity among all compounds in its "chemical space neighborhood" to give a degree of confidence in its activity. We demonstrate that this local hit rate (LHR) analysis method outperforms hit selection based on ranking by primary screen activity values across ten diverse high throughput screens, spanning both cell-based and biochemical assay formats of varying biology and robustness. On average, the local hit rate analysis method was ~2.3-fold and ~1.3-fold more effective in identifying active compounds and active chemical series, respectively, than selection based on primary activity alone. Moreover, when applied to finding false negatives, this method was 2.3-fold better than ranking by primary activity alone. In most cases, novel hit series were identified that would have otherwise been missed. Additional uses of and observations regarding this HTS analysis approach are also discussed.

INTRODUCTION

High-throughput screening (HTS) is a commonly used approach in modern small-molecule drug discovery programs. With advances in the automation of plate-based screening (typically a 384-well microtiter plate), high throughput biology, and combinatorial chemistry, many pharmaceutical and biotech companies can now routinely screen millions of compounds in their inventory against targets of therapeutic interest. HTS campaigns sometimes result in thousands of primary "hits", compounds in the collection that display an activity in the HTS assay that is greater than an activity threshold. Several factors can influence the activity threshold used to define a primary hit, including statistical measures of assay performance, costs associated with running the assays (reagents and consumables), project goals, and practical limitations of compound ordering systems.[1] The initial hits are then retested to "confirm" their activity in a single-dosed format and/or a dose−response format (IC50/EC50) of the HTS assay. It is common that a significant percentage of the primary HTS hits are false positives, compounds that demonstrate activity in the initial assay but have no specific interaction with the target. False positives are often identified and discarded during this confirmation process as they either fail to recapitulate their original activity or orthogonal assays rule them out due to lack of specificity for the intended drug target. Many factors could contribute to the occurrence of false positives, including most notably the purity and solubility of compounds on screening plates,[2] compound aggregation,[3] the presence of "frequent hitters" (e.g., compounds that act downstream of the target and affect mechanisms that influence the assay readout), compression artifacts,[4] errors from HTS instruments, the quality of the assays,[5,6] and random fluctuation of biological signals. Genuine actives can be missed in the initial HTS assay for similar reasons and are often referred to as "false negatives".

The impact of false positives and false negatives can be significant, with the former being more prevalent and the latter having generally greater consequences. The presence of a large number of false positives in the initial selection of hits for confirmation studies constitutes a waste of screening resources and a potential source of delay in the progress of a project. However, a large number of false negatives (missed actives) in the primary HTS assay could have more severe consequences, since each missed active compound could have provided a viable starting point for lead development and ultimately drug discovery. Thus, the method used to select primary hits from a given HTS plays a very important role in ensuring that the impacts of false negatives and false positives on project progression are minimized.

Several computational and statistical methods have been developed in recent years to aid in defining assay quality and selecting primary hits for confirmation studies. The most commonly cited measure of assay quality for HTS is the *Z*-factor,[7] an index developed by Zhang and Oldenburg that uses the activity distribution of positive and negative controls. The *B* score, proposed by Brideau,[8] tries to correct column

* Corresponding author e-mail: james.e.mills@pfizer.com.
[†] PGRD Groton Laboratories.
[‡] PGRD Computational Sciences CoE.
[§] PGRD Sandwich Laboratories.

10.1021/ci900113d CCC: $40.75
Published on Web 10/01/2009

and row biases in assay plates (e.g., edge effects, liquid addition artifacts, etc.). A well-position-based bias correction method was also reported recently to remove positional bias frequently caused by mis-calibration of HTS instruments.[9] Orthogonal to these statistical methods, chemistry-knowledge-based methods have also been proposed. For example, a naïve Bayesian model for predicting compound aggregation was recently developed by Feng and Shoichet using data from aggregation screens.[3] Yan and co-workers proposed a knowledge-based scoring system, called the ontology-based pattern identification (OPI) score,[10] to improve hit selection by considering both the activity ranking and the number of structurally similar compounds in the screening collection. Others used clustering techniques to first group all compounds tested into structural classes and then score compound classes rather than individual compounds.[11,12]

In this study, we propose a novel, knowledge-based, hit selection method called local hit rate (LHR) analysis. LHR analysis is based on the observation that if a compound is genuinely active, there is a higher probability that its structural analogs in the neighboring chemical space are also active.[13,14] This observation is supported by the fact that closely related analogs often share the same shape and pharmacophoric interactions required for target recognition; therefore, these congeners of the true active are more likely to be active. Here we define the local hit rate for a given compound as the number of active compounds in its "chemical neighborhood" (as defined by Tanimoto similarity) divided by the total number of compounds in this neighborhood. We envisaged that compounds could be selected with greater confidence for retesting if the local hit rate was used as a criterion in addition to an activity value threshold. We reasoned that false positives might be more readily identified as compounds with a low local hit rate and that apparently weak actives (compounds with activity values below the assay threshold) with high local hit rates are likely to be false negatives. To verify this, we developed a statistical scoring system based on the probability of observing a LHR that is statistically different from the global hit rate (the hit rate for the whole screen, defined as the number of actives in the primary screen divided by the number of compounds tested). We applied this scoring method retrospectively to ten HTSs conducted recently within Pfizer, covering different assay formats and biological targets. We have demonstrated that, on average, the LHR methodology out-performed activity-based filtering in terms of improving the retest confirmation rate i.e. reducing the number of false positives. Furthermore, we were able to show that the scoring system can also be used to identify false negatives.

METHODS

1. Calculation of Local Hit Rate Score. The local hit rate of a compound is defined as the proportion of its near neighbors that are active in the primary assay. A compound with a high local hit rate would be expected to be a genuinely active compound. The number of near neighbors will influence the confidence of this prediction; so, for example, a compound having 30 actives among 100 near neighbors generates a higher confidence of activity than a compound with 3 actives among 10 near neighbors. Furthermore, a compound with 0 actives among 100 near neighbors generates a higher confidence of inactivity than a compound with 0 actives among 10 near neighbors. In this work, the chi-squared statistic[15] was used as a simple method to differentiate such cases, comparing the number of observed hits with the number of expected hits thus

$$\chi^2 = \frac{(Observed - Expected)^2}{Expected}$$

The chi-squared value was converted to a probability value using standard lookup tables. We also made use of other statistical methods, including calculating probability values based on the binomial distribution, as was used by Krumrine et al. in a similar method for prioritizing virtual screening hits.[16] Notably, these other approaches did not change the nature of the results we obtained.

This method is illustrated using an example high throughput screen of 1 million compounds, of which 5000 compounds achieved primary activity values above a 40% inhibition threshold. For each of the 5000 apparent actives, the near neighbors (here defined as compounds that share a Daylight fingerprint Tanimoto similarity value of at least 0.5) are determined. The proportion of these neighbors that show an activity value greater than 40% is determined and compared with the background hit rate (in this case 5000/1MM* 100% or 0.5%) using the chi-squared statistic. In this case, a compound with 200 neighbors would be expected to yield one active neighbor based on the background hit rate. If there were 3 active neighbors (local hit rate=1.5%), this would translate to a chi-squared value of 4.0, with a corresponding P-value of about 0.045 (i.e., there is a 4.5% chance that this LHR value was obtained purely by chance). The LHR score is defined as the negative logarithm of this P-value. If the observed value is less than the expected value, the LHR score is converted to its negative value to indicate a local hit rate lower than expected. The chi-squared test cannot be applied to compounds with no near neighbors. These compounds are given a LHR score of 0.0 i.e. we cannot say whether the local hit rate is above or below the expected LHR. Such compounds are therefore unlikely to appear in the top ranking compounds for selection or the lower-ranking compounds for deselection. If prioritization of this (usually small) subset of compounds is required, a method other than LHR should be applied (e.g., ranking by activity value) combined with visual inspection of the structures, since singletons tend quite often to be outliers in chemical space and therefore unattractive.

2. Performance Metrics. Two metrics were used to assess the performance of the local hit rate analysis in differentiating true from false positives among primary HTS hits. The assessment of success can be defined as either the number of genuinely active compounds (i.e., as identified by a retest confirmation assay) or chemotypes retrieved. For the latter case, the number of active chemotypes was defined as the number of clusters represented by the active compounds retrieved. In this work, the confirmed active compounds were clustered using a simple, agglomerative clustering method: all pairwise similarity values were calculated, and any pair of compounds with a Daylight fingerprint Tanimoto similarity value above 0.7 was assigned to the same cluster. Other similarity cutoffs and clustering methods have been applied, but they do not change the nature of the results obtained in

this work. A cluster was defined as being retrieved if any compound within the cluster was retrieved. The performance of LHR was compared with other hit selection methods using enrichment plots,[17] in which the fraction of genuinely active compounds or clusters retrieved was plotted against the rank order of compounds as determined by LHR, activity value from the primary HTS, "perfect", and random selection. Methods with better performance show a larger area under the curve (AUC) value on enrichment plots. The random model on the enrichment plots retrieves genuinely active compounds at the same rate as the overall confirmation rate. The perfect model, in contrast, retrieves all genuinely active compounds first before picking up false positives. The chemotype retrieval rate for the random selection is not linear because the number of clusters retrieved is only incremented when the first member of a cluster is encountered. Therefore, we assessed this retrieval rate by taking the average cluster retrieval rate from 100 simulation runs in which the ranks of the compounds in the input file were randomly assigned. The ability of local hit rate analysis to recover false negatives in the "activity twilight zone" (the activities just below the assay threshold set for the screen, typically 10 percentage points) was assessed similarly at both the compound and the chemotype levels. Comparisons were made between LHR and more simple similarity searching methods in the identification of false negatives.

RESULTS AND DISCUSSION

It is not always tractable or desirable to determine a dose−response relationship for every primary hit from an HTS assay. The necessity of shrinking cycle times in drug discovery makes it imperative to find more efficient and effective means for selecting HTS hits than simply using a threshold for activity.[1] More specifically, we wish to avoid missing true actives (false negatives) and spending time and resource on false positives. Some form of orthogonal information or an additional criterion is required to "triage" or prioritize this large number of putative active compounds such that the maximum benefit of the HTS campaign is realized in the shortest time frame. A simple approach is to pick the compounds that show the highest activity values in the primary screen, a "top N" approach. This can often lead to picking many compounds from fairly large chemical series while ignoring other smaller series where chemical space and biological activity overlap. Another commonly used approach is to cluster putative actives and pick representative compounds from each cluster for screening with or without a weighting for cluster size. In our experience, this approach still tends to bias selections toward densely populated regions of chemical space in the screening collection. This occurs in part due to the intrinsic inaccuracy of the HTS assay. More specifically, if the assay is 99.9% accurate in identifying active and inactive compounds, then 0.1% of the time the assay may misidentify an inactive compound as being an active. In the context of a large compound collection (~2 MM compounds), this translates into ~2000 false positives. Therefore, large clusters or libraries of compounds in the screening collection are more likely to be represented in selection lists as false positives in an assay than smaller clusters of compounds (Figure 1A).

In this paper, we have explored a new method for primary HTS hit selection, local hit rate analysis. Very simply, this

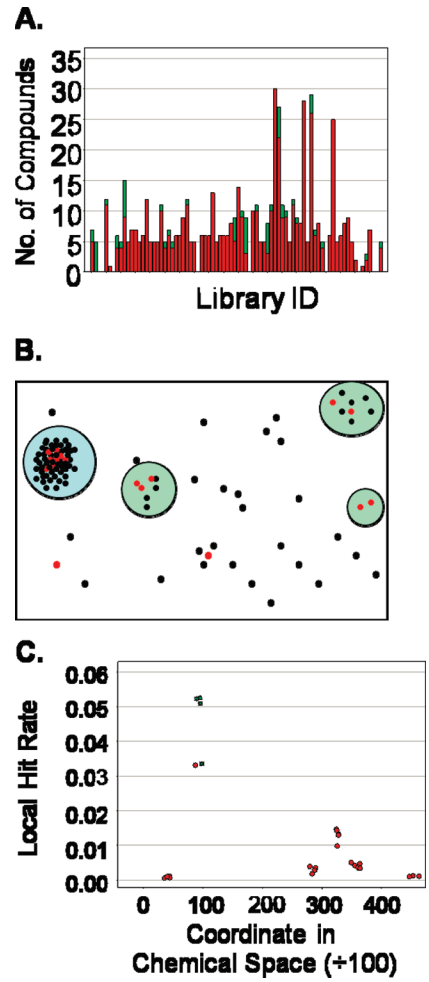

Figure 1. A. Confirmation studies of HTS actives from large combinatorial chemistry libraries. The number of HTS actives from each library (*y*-axis) is plotted against the library number (*x*-axis). The number of compounds that confirmed in follow up screening is indicated by green bars. Those that did not confirm are indicated by red bars. The total number of compounds active in the primary HTS assay from any given library is the sum of the green and red bars. **B.** Threshold-based selection of actives from primary screening. For illustration purposes, chemical space is represented in 2 dimensions with closely related compounds displaying similarities in both dimensions. Inactives are included in the plot, and it can be seen that the large cluster of active compounds is part of a much larger group of generally inactive compounds (highlighted by a blue background), so it becomes less attractive than the smaller clusters of compounds with a higher local hit rate (highlighted by a green background). Threshold-based selection of HTS hits is often biased toward large clusters of active compounds (highlighted with a blue background). **C.** Local hit rate (*y*-axis) plotted against chemical space (*x*-axis) for the largest 6 clusters found in the screen G7332. Compounds subsequently confirmed as being active are shown as green stars. Simply picking compounds from the largest clusters would have only yielded 4 actives from one of these clusters. Taking account of the local hit rate would have provided a more efficient method for identifying true actives.

method looks at the activity of a compound in the context of the activities of its nearest neighbors in the entire HTS campaign, determines a local hit rate (number of active nearest neighbors divided by the total number of nearest neighbors), and calculates a score based on the chi-squared P-value (which answers the question "Is this local hit rate significantly different from the hit rate for the entire screen?"). Unlike structure-based clustering approaches, which look at the activity of clusters, this method looks

Table 1. High-Throughput Screens Considered in Evaluating the Local Hit Rate Selection Method

screen no.	target class	mode of action (assay type)[a]	Z′	hit threshold (%)	library (millions)[b]	hit rate[c]	confirm rate[d]
G7345PC	non-GPCR receptor	agonist (C, BL)	0.7	25	2.2	0.86	9.9
G7344PC	ion channel	antagonist (C, Ca)	0.5	50	2.2	0.17	8.2
S8513P	GPCR	agonist (C, BL)	0.6	20	1.75	0.51	12.2
G7184A	non-GPCR receptor	agonist (C, BL)	0.6	32	2.0	0.27	43.2
G7664PC	GPCR	antagonist (C, L)	0.7	50	3.0	0.29	16.4
S8569PC	protein−protein	inhibition (B, F)	0.7	40	2.2	0.89	16.1
S8614PC	kinase	inhibition (B, L)	0.81	40	2.1	1.70	0.4
G7391PC	enzyme	inhibition (B, F)	0.84	60	2.3	0.60	16.5
G7332PC	enzyme	inhibition (B, F)	0.5	50	2.1	0.45	26.9
G7378PC	enzyme	inhibition (B, A)	0.86	30	2.3	0.53	14.1

[a] C, cell-based assay; B, biochemical assay; BL, beta lactamase; Ca, Ca^{2+} fluorescence; L, luminescence; F, fluorescence; A, absorbance. [b] Compound libraries were screened in a compressed format (multiple compounds per well). [c] Expressed as a percentage: (no. of hits/no. of compounds tested in the primary screen) * 100%. [d] Expressed as a percentage: (no. of confirmed hits/no. of compounds tested in confirmation) * 100%. In the confirmation format, each compound was tested multiple times at the dose used for the primary HTS screen (typically, 10 μM). A confirmed hit is defined as a compound with a median activity greater than the hit threshold.

Table 2. Percentage of Active Compounds and Series Recovered by Local Hit Rate Analysis and Activity Ranking When the Top 10% Ranked Compounds Are Screened

	percentage active compounds recovered[a]		percentage active series (inc. singletons) recovered[b]		percentage active series (exc. singletons) recovered[c]	
screen no.	rank by activity	rank by LHR	rank by activity	rank by LHR	rank by activity	rank by LHR
G7345PC	17	41	42	47	53	66
G7344PC	22	33	38	33	67	70
S8513P	12	40	30	40	46	71
G7184A	14	49	19	38	39	69
G7664PC	12	30	40	41	62	73
S8569PC	11	38	37	51	68	72
S8614PC	18	49	41	52	68	72
G7391PC	15	24	30	34	53	70
G7332PC	20	38	30	41	55	81
G7378PC	19	28	37	36	70	84

[a] Compounds were ranked by either primary activity or LHR. The percentage of actives recovered was calculated by dividing the number of true actives identified among the top 10% ranked by each method by the total number of true active compounds and multiplying this result by 100. True actives were defined as those compounds that recapitulated activity in the confirmation studies. [b] True active compounds were clustered into series using an agglomerative clustering algorithm (see Methods for details). A cluster was considered to be recovered if any compound in the cluster was retrieved among the top 10% ranked compounds, with the constraint that no more than 3 compounds could be selected from any one cluster. Percentage active series recovered was calculated by dividing the number of active series identified in the top 10% ranked by the method by the total number of active series and multiplying this result by 100%. [c] Similar to 2 with the exception that clusters containing only one compound were removed from the analysis.

at each compound individually and compares it with its neighbors in calculating the local hit rate and associated score (see Figure 1B). This ensures that there are no built-in biases from a clustering method and that each compound is evaluated in the context of both active and inactive neighbors (Figure 1C).

1. Data Sets. The LHR method was retrospectively applied to ten high-throughput screens recently run at Pfizer (Table 1). In selecting these screens, we tried to represent both cell-based and biochemical formats as well as a variety of target classes, modes of action, and assay readouts. Each screen tested in excess of 1.5 million compounds in a compressed compound format (multiple compounds per well) in the primary assay and tested over 1000 compounds in confirmation (single compound per well) studies. The local hit rate method was applied to each screen in three modes: hit selection, hit deselection, and false-negative searching. In each modality, we evaluated the method for its ability to retrieve genuinely active compounds and chemotypes from the primary screening results, as determined by results from confirmation screens. For hit selection, we used the activity threshold criteria for each screen (Table 1), and either selected the top 10% ranked by the local hit rate method or deselected compounds with LHR scores below a number of thresholds. For the false-negative searching approach, we artificially set the activity threshold 10% higher than that originally set for the screen and then calculated a local hit rate score for those compounds that were in the 10% interval below this new threshold. As for the hit selection approach above, we selected the top 10% ranked by the LHR score. For example, in the case of screen G7332PC, we defined an "active" or "hit" in the HTS assay as a compound giving rise to an activity value greater than 50% and then assigned local hit rates to compounds with primary HTS activity values in the range 50−60% based on their relationships to active (activity $\geq$50%) and inactive (activity <50%) neighbors. In each modality, the results from the LHR method were compared with two alternative approaches: 1) selecting a random subset of compounds and 2) ranking compounds on the basis of their activity values from the primary HTS. We performed this analysis for all ten high-throughput screens listed in Table 1.

2. Retrieval of Active Compounds. To illustrate this process, we describe our findings with the screen G7332PC. For the hit selection approach, we used the LHR method to identify true active compounds from a set of 8782 com-

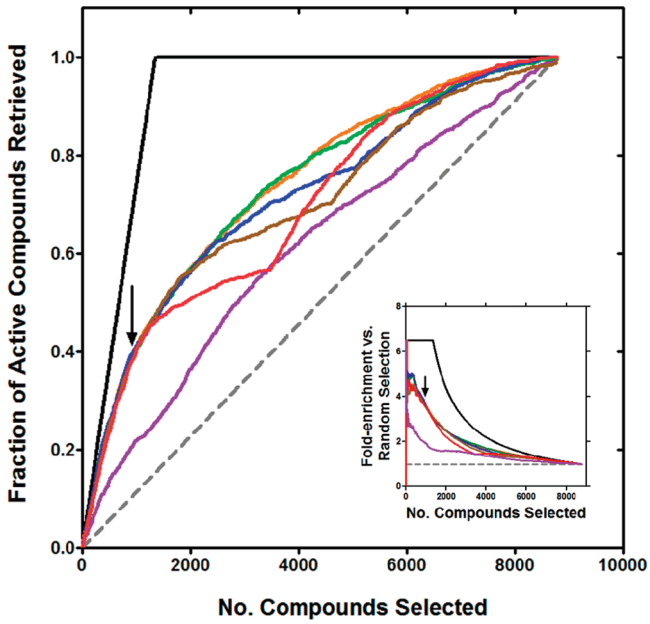

Figure 2. Retrieval of confirmed active compounds from an HTS for all compounds that generated an activity value of greater than 50% in the initial assay. The proportion of retrieved actives is plotted against the cumulative number of compounds selected. The black and gray lines show the results expected from perfect and random selections, respectively. Compounds were selected and ranked on the basis of their activity value in the initial assay (purple) or their local hit rate score calculated with the 0.4 (orange), 0.5 (green), 0.6 (blue), 0.7 (brown), and 0.8 (red) Tanimoto similarity cutoffs. Inset: The fold enrichment relative to random selection is plotted versus the number of compounds ranked. The arrow indicates the fold-enrichment achieved when the top 10% of the ranked compounds is considered.

pounds that gave a positive readout in a primary HTS assay. As shown in the enrichment plot in Figure 2, selecting compounds using the LHR method is more successful (larger AUC, statistically significant at the 0.01% level with the Wilcoxon test[18]) than making selections based on their activity in the primary HTS assay or selecting compounds at random. In this plot, the fraction of true actives retrieved (*y*-axis) is plotted as a function of the rank of compounds according to each approach. Again, the true actives were those compounds that reproducibly demonstrated activity in the confirmation studies (single-dose studies, multiple replicates per compound). Compared with random selection, we observed that selecting the top 10% ranked by local hit rate resulted in a useful balance between the fraction of actives recovered and fold-enrichment (Figure 2, inset).

The major parameter that can be varied during the LHR calculation is the similarity cutoff used to define the chemical neighborhood within which the local hit rate is calculated (See Figure 2). Higher similarity values will lead to sampling of lower number of close-in neighbors, and an increased number of compounds that are regarded as singletons (0 active neighbors out of 0 near neighbors and a LHR score of 0). This gives rise to the biphasic nature of some of these curves. The reasons for the local rise in slope following this juncture are not completely clear but are likely to be consequences of the nature of our compound collection. One explanation is that many commerical compounds in the collection are singletons that are known or suspected to be active in certain target classes. Enrichment increases significantly when similarity values are reduced toward 0.5

Table 3. Impact of Varying the Tanimoto Similarity Cutoff for Nearest Neighbors in the Local Hit Rate Method[a]

Tanimoto similarity cutoff for near neighbors	# active/# screened	LHR	LHR score
0.8	6/6	1.000	26.688
0.7	23/38	0.605	54.894
0.6	55/129	0.426	80.668
0.5	120/490	0.245	68.462

[a] Effects of varying the Tanimoto similarity cutoff on the local hit rate around an active compound and the log of the local hit rate probability value. As the similarity threshold is relaxed, more neighbors come into play. This causes the local hit rate to drop because more compounds are dissimilar to the active compound. However, the magnitude of the P-value decreases and the LHR score increases because statistically the ratio is more likely to be different to the background hit rate (which in this case is 1%). However, this difference in P-values is not in itself likely to be relevant because a P-value of 10^{-81} and a P-value of 10^{-55} are in essence the same. This observation cautions against overinterpreting the magnitude of the P-value. Both the LHR and P-values are worth consideration.

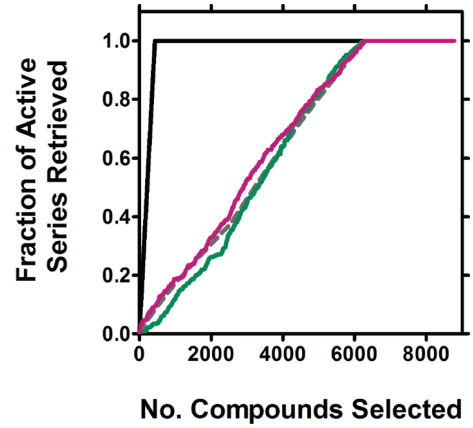

Figure 3. Retrieval of confirmed active chemotypes from an HTS for all compounds that generated an activity value greater than 50% in the initial assay. The proportion of retrieved chemotypes is plotted against the cumulative number of compounds tested. The black and gray lines show the results expected from perfect and random selections, respectively. Compounds were selected and ranked on the basis of their activity value in the initial assay (purple) or their local hit rate score using a value of 0.5 for the Tanimoto similarity cutoff (green).

but beyond that does not change much. Therefore, for the remainder of the work presented in this paper, a Tanimoto cutoff of 0.5 is used to define near neighbors for LHR analysis.

Interestingly, when looking at individual compounds, the local hit rate was often observed to decrease as the similarity threshold was reduced from 0.8 to 0.5, in line with work recently published.[19] However, the LHR score was observed to increase over this range of similarity values due to the increase in neighborhood size. As the size of the neighborhood increases, the likelihood of a difference between the LHR from the background hit rate being statistically significant increases, even though the magnitude of the difference may decrease (Table 3).

3. Selection of Active Chemotypes. It is often the case that a more useful measure than the number of active compounds retrieved from a HTS is the number of active chemotypes identified. The performance of LHR in this regard is shown in Figure 3. In this work, a chemotype is

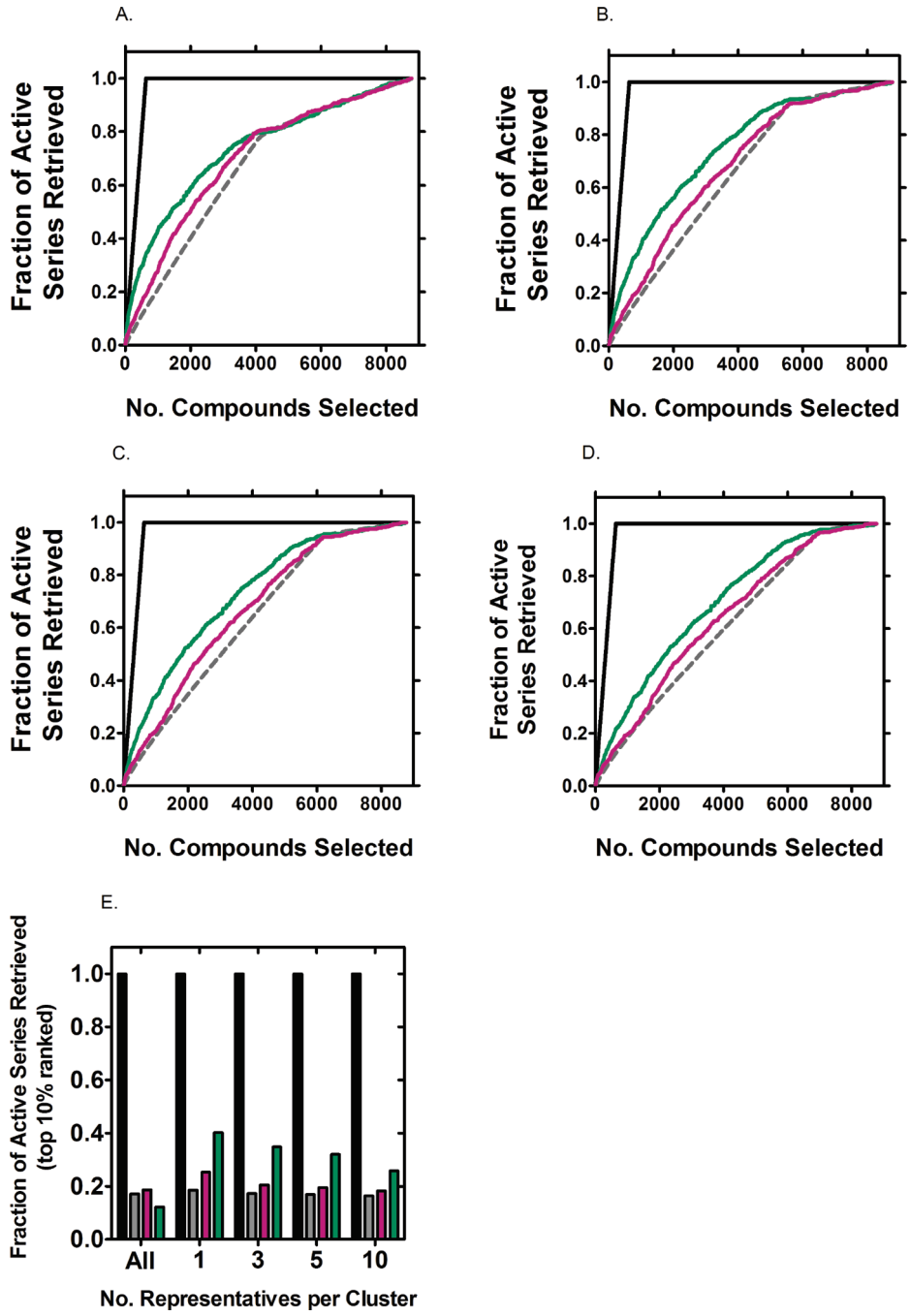

Figure 4. Retrieval of confirmed active chemotypes from an HTS (G7332A). Compounds with inhibition values greater than 50% relative to the assay controls were scored as "active". The proportion of chemotypes retrieved is plotted against the cumulative number of compounds selected. The number of compounds that were selected from any one chemotype was restricted to no more than 1 (**A.**), 3 (**B.**), 5 (**C.**), and 10 (**D.**). The black and gray lines show the results expected from perfect and random selections, respectively. Compounds were ranked and selected on the basis of their activity value in the initial assay (purple) or their local hit rate score using a value of 0.5 for the Tanimoto similarity cutoff (green). **E.** Summary of the results for active series retrieval when All, 1, 3, 5, or 10 members of each series are selected based on ranking the top 10% using HTS activity (purple) and LHR with a Tanimoto similarity of 0.5 (green). Results for perfect and random rankings are shown in black and gray, respectively.

defined as a cluster of compounds sharing a similarity of 0.7. The true active compounds were clustered using a simple agglomerative clustering algorithm. If a method retrieved any active compound from a cluster, the cluster was said to be retrieved. Ranking compounds on the basis of their local hit rate score gave poorer recovery of chemotypes than ranking on the basis of their HTS primary activity value. This is not surprising given that larger clusters of active compounds would be expected to populate the top few compounds ranked by the LHR method, given the larger numbers and better statistics (see Table 3). This illustrates that using the LHR value in isolation is likely to restrict the number of chemotypes identified.

If the goal of an HTS is to identify as many active chemotypes as possible, one modification to this simple LHR implementation is to restrict the number of compounds selected per cluster. In this modified method, the HTS hits are first grouped into clusters using a scaleable clustering algorithm (in this case agglomerative clustering). Only the highest ranking hits from each cluster are selected for

confirmation testing. The lower ranked compounds in each cluster are appended to the bottom of the rankings. This approach ensures the selection of the highly ranked compounds in all clusters and avoids oversampling of the large clusters. As a result, more chemotypes are represented in the top-ranking compounds, which should lead to an increase in chemotype retrieval success rates. To test this approach, we compared the chemotype recovery rate of ranking using LHR to that obtained by ranking using HTS activity. In this comparison, a uniform restriction on the number of compounds per cluster was applied to both methods. The results of several comparisons (under different levels of restriction) are shown in Figure 4. Under each of these conditions, LHR ranking gave a better recovery of chemotypes than ranking by the HTS primary activity value (all enrichment curves statistically significantly different to both random- and activity-based ranking at the 0.01% significance level with the Wilcoxon test[18]). Picking only one compound per cluster gave rise to the most efficient retrieval of chemotypes (Figure 4A). The efficiency of retrieval reduced as the number of prepresentatives per cluster was increased. This can be seen in Figure 4E, in that the convergence between the LHR and random curves occurs further to the right as the number of representative compounds is increased. However, picking fewer representatives per cluster runs the risk of missing chemotypes because inactive representatives could be picked. For the remainder of the work in this paper, we chose to select up to 3 representatives per cluster as a compromise between retrieval rate and the asymptotic fraction of series retrieved.

4. Deselection of False Positives. False positives usually represent wasted resource. Any method that can reduce the number of compounds screened to achieve the same retrieval of active chemotypes would be welcome. Table 4 shows the results from applying three different LHR score cutoffs to remove apparently poorly performing compounds. In the most conservative approach, removal of around 5% of the lowest ranking actives from the first round of screening gives rise to no loss in chemotype retrieval for 9 out of the 10 screens. A consistent saving of 5% of screening resource for likely no loss in information could represent a worthwhile use of the local hit rate methodology. Removal of all compounds with a local hit rate below that expected at random (LHR score <0) leads to around 60% of compounds being screened and around 85% of actives from around 95% of chemotypes being identified. The small loss in information in exchange for a large saving in screening resource could be a price worth paying, particularly for a hit-rich screen.

5. Retrieval of False Negatives. False negatives can represent a costly oversight, i.e. the failure to identify genuinely active compounds or chemotypes. Because this was a retrospective analysis, our "false negative searching" was carried out by testing the ability of different methods to identify hits and series from the "activity twilight zone" (compounds with activity values just below the assay threshold set for the screen, typically 10 percentage points). More specifically, if a cutoff of 50% inhibition was used to identify primary hits, the ability to pull back confirmed active compounds from the 50−60% cohort of primary hits was tested. In addition to comparing the local hit rate methodology with screening on the basis of primary activity value, a comparison was made with simple similarity searching. Two

Table 4. Percentage of Actives Recovered by Removing Compounds on the Basis of Their Low Local Hit Rates

	LHR score <0.0			LHR score < −1.0			LHR score < −5.0			
screen	% cpds screened	% actives recovered	% clusters (inc. singletons) recovered	% clusters (exc. singletons) recovered	% cpds screened	% actives recovered	% clusters (inc. singletons) recovered	% clusters (exc. singletons) recovered	% cpds screened	% actives recovered	% clusters (inc. singletons) recovered	% clusters (exc. singletons) recovered
G7345PC	62	87	99	99	82	94	99	100	92	98	100	100
G7344PC	59	79	97	98	84	96	99	100	95	99	100	100
S8513P	53	79	95	97	74	91	99	100	89	96	100	100
G7184A	58	91	92	91	84	98	98	100	96	99	100	100
G7664PC	58	84	95	94	84	95	98	98	97	99	100	100
S8569PC	58	82	97	100	79	94	99	100	90	98	100	100
S8614PC	63	93	97	96	80	98	99	99	90	99	100	100
G7391PC	65	89	98	98	81	96	99	99	90	98	100	100
G7332PC	59	85	95	96	86	97	99	100	96	99	99	100
G7378PC	61	85	97	100	86	97	99	100	97	99	100	100

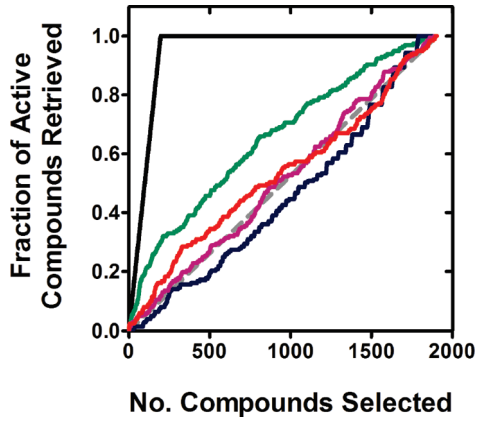

Figure 5. Retrieval of false negative compounds from an HTS. In this case, the cutoff used by the HTS was 60%. Hits were sought among compounds giving initial activity values in the 50−60% range. The black and gray curves show results expected from perfect and random selections, respectively. Compounds were selected and ranked on the basis of their activity value in the initial assay (purple), their local hit rate score using a value of 0.5 for the Tanimoto similarity cutoff (green), the number of active neighbors with a Tanimoto similarity greater than 0.5 (magenta), or the Tanimoto similarity value for the closest active neighbor (red).

similarity-based metrics were used for ranking for each compound, namely the number of active neighbors and the similarity of the closest active neighbor, both calculated using Daylight Tanimoto similarity. For consistency with the LHR method, we used a similarity threshold of 0.5 to define the number of near neighbors for a compound. As can be seen in Figure 5, picking compounds on the basis of their primary activity values retrieves true actives barely more efficiently than random selection (0.11/0.1 = 1.1-fold enrichment for the top 10%), but LHR (0.29/0.1 = 2.9-fold enrichment) outperforms both this method and the two similarity methods (which give 1.6- and 0.7-fold enrichment). Furthermore, of the top 10% of compounds selected by LHR (amounting to 193 compounds), 6 true actives were not contained in clusters derived from the original active cohort (primary activity ≥60%). These represent series that would have been missed by an HTS run using our previous paradigms.

The results for the remaining screens are shown in Table 5. Here we show the percentage of genuinely active compounds and series identified if the top 10% of compounds in the "activity twilight zone" are selected on the basis of either activity value or local hit rate score. In the majority of cases, the local hit rate method achieves better results than ranking on the basis of activity, in terms of both compounds and series retrieved. However, ranking on the basis of initially observed activity value consistently retrieves more of the novel series, a result we could expect because the local hit rate methodology is more likely to identify actives that are similar to pre-existing actives.

6. Observations and Conclusions. In considering the utility of the local hit rate analysis method in selecting potential actives and false negatives, we chose to look at 10 high throughput screens that encompassed different targets classes, assay formats, levels of robustness (e.g., Z′), and modes of action. Using the top 10% ranked in our evaluations, this approach recovered on average ∼37% of the active compounds (standard deviation ∼ 8.4%) which is ∼2.3-fold better than using activity ranking alone to make selections (average ∼16% of actives; standard deviation ∼4%). In selecting active series, the local hit rate method fared slightly better than ranking by activity (LHR ranking: average ∼41% of active series, standard deviation ∼7%; activity ranking: average ∼34%, standard deviation ∼7%). However, if the active compounds are clustered and cluster representatives are selected, then the average percentage of active series selected by the LHR method increases 1.8-fold to ∼73% (standard deviation ∼5.5%) while that achieved using activity changes by ∼1.7-fold (average percentage of active series recovered is ∼58%, standard deviation ∼11%). The latter observation underscores the propensity of the LHR method to favor large clusters of compounds if used in isolation.

Selecting the top 10% of the primary HTS hits may have limited use for screening groups with resources available to assay thousands of compounds in confirmation studies. For example, an HTS assay with a high hit rate could yield 40,000 hits from a screening collection of 1 million compounds (a 4% hit rate), and it might be necessary to select 10 or 20% of the hits for confirmation studies to expedite lead discovery. The local hit rate method would be an effective way of selecting these compounds. Yet, for many screens, the number of primary hits is well within the resources of many groups and selecting a subset is not necessary. In these instances, the question may be "Can the impact of false positives be reduced?" To answer this question, we set thresholds for the local hit rate score that would select the top 60, 80, and 90% and asked questions similar to those posed above for the top 10% (see Table 4). Across 10 high throughput screens, the local hit rate method identified an average of 85% of the active compounds (standard deviation: ∼5%) in an average of 96% of the active series (standard deviation: ∼2%) when the top scoring 60% of the compounds are selected. Excluding singleton com-

Table 5. Percentage of Actives Recovered by False Negative Searching When Screening the Top 10% Compounds As Ranked by Local Hit Rate or Activity Value

	% active compounds recovered		% active series recovered		novel active series	
screen no.	rank by activity	rank by LHR	rank by activity	rank by LHR	rank by activity	rank by LHR
G7345PC	18	38	31	43	3	0
G7344PC	13	27	27	27	8	3
S8513P	8	32	18	36	0	1
G7184A	20	46	28	41	4	2
G7664PC	13	31	29	47	4	1
S8569PC	11	27	23	44	3	3
S8614PC	12	48	27	48	6	3
G7391PC	11	24	24	36	11	6
G7332PC	11	29	21	35	6	2
G7378PC	13	35	22	35	4	2

pounds (compounds with no nearest neighbors) had no appreciable effect on the number of active series recovered (average: ~96%; standard deviation: ~3%). Increasing stringency by selecting the top 80% or 90% affords more complete recovery of active compounds and a limited advantage where recovery of active series is concerned (e.g., top 60%: a mean of 96% of the active series are recovered; top 80%: a mean of ~99% of the active series are recovered; top 90%: a mean of ~100% of the active series are recovered). Thus, it is reasonable to consider the local hit rate method as a means of reducing the number of potential false positives, compounds that have no active nearest neighbors and arise due to random errors during screening. However, in thinking about this use of the method, it is important to consider the average cluster size of the screening collection. For example, a very diverse screening collection may contain a significant number of compounds with no near neighbors (cluster size = 1) in which case, true singleton actives could be missed if the LHR method is used in isolation. Alternative approaches are required to deal with singletons in a screening collection. The pragmatic approach could be to screen all singletons, though prioritization could be carried out by either ranking based on activity values or LHR-type methodologies based on descriptors that do not rely on 2D similarity e.g. pharmacophores or shape.

Searching for false-negatives should be a final quality control check for any high throughput screen. It enables one to estimate whether or not all the active compounds and series were found, and, in doing so, comment on the effectiveness of the screen. In the retrospective analysis described here, the LHR method was roughly 1.6- and 2.6-fold more effective than ranking by activity in finding false negative series and compounds, respectively, with activity values in the primary screen just below the activity threshold used in the original screen. Importantly, it was gratifying to see that new series can be identified using this method, although alternative methods of false negative identification (e.g., incorporating degree of activity) are likely to be more fruitful in this regard. Preliminary studies with "live" screens (biology resource available to do testing) suggest that one can dip lower into the assay "noise" of the primary assay with the LHR method (data not shown). The extent of false-negative searching will, of course, depend on the available resource for testing and the tractability of series already in hand. Typically, false negative searching is currently carried out following confirmation of activity, either similarity or substructure searching in a compound collection around favored actives. The LHR approach allows the same ground to be covered more quickly, without the bias introduced by the manual selection of series.

This methodology does not address all sources of error in a screen. For example, it will not differentiate assay false positives, compounds interacting with the assay technology (e.g., reporter enzymes, fluorescence signals); rather it may favor such compounds. One approach to this problem is to calculate the local hit rate scores for all hits in another assay of the same format with a target protein expected to deliver different chemotypes and remove compounds that score highly in both assays.

In summary, this work demonstrates the utility of the LHR method of selecting compounds for testing based on data from primary HTS screening results. It can be successfully applied to hit selection (or deselection) or to finding false negatives that were missed in the initial primary screen. Additional applications are possible. For example, we have successfully used this approach to rank compounds that have never been tested based on their similarity to a cohort of compounds for which activities are known. This particular use of the method has allowed us to prospectively profile external and virtual compounds prior to synthesis and actual testing. In some other cases, LHR was used to quickly survey the SAR landscape around the confirmed hits. For example, genuine hits with a low LHR are likely to display steep structure−activity relations (SARs) to compounds around them and may be less amenable for hit to lead optimization. These and other uses of the LHR method are currently being evaluated across more examples.

ACKNOWLEDGMENT

We thank Richard Lyons, Savina Jaeger, Enoch Huang, Eryl Sharp, Chris Williams, and Jeremy Everett for their advice, suggestions, and support of this project.

REFERENCES AND NOTES

(1) Posner, B. A. High-throughput Screening-driven Lead Discovery: Meeting the Challenges of Finding New Therapeutics. *Curr. Opin. Drug Discovery Dev.* **2005**, *8* (4), 487–494.
(2) Di, L.; Kerns, E. H. Biological assay challenges from compound solubility: strategies for bioassay optimization. *Drug Discovery Today* **2006**, *11* (9−10), 446–451.
(3) Feng, B. Y.; Shelat, A.; Doman, T. N.; Guy, R. K.; Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* **2005**, *1* (3), 146–148.
(4) Hann, M.; Hudson, B.; Lewell, X.; Lifely, R.; Miller, L.; Ramsden, N. Strategic pooling of compounds for high-throughput screening. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 897–902.
(5) Diller, D. J.; Hobbs, D. W. Deriving Knowledge Through Data Mining High-throughput Screening Data. *J. Med. Chem.* **2004**, *47*, 6373–6383.
(6) Spencer, R. W. High-Throughput Screening of Historic Collections: Observations on File Size, Biological Targets, and File Diversity. *Biotechnol. Bioeng.* **1998**, *61* (1), 61–67.
(7) 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.
(8) Brideau, C.; Gunter, B.; Pikounis, B.; Liaw, A. Improved statistical methods for hit selection in high-throughput screening. *J. Biomol. Screening* **2003**, *8* (6), 634–647.
(9) Makarenkov, V.; Zentilli, P.; Kevorkov, D.; Gagarin, A.; Malo, N.; Nadon, R. An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics* **2007**, *23* (13), 1648–1657.
(10) Yan, S. F.; Asatryan, H.; Li, J.; Zhou, Y. Novel statistical approach for primary high-throughput screening hit selection. *J. Chem. Inf. Model.* **2005**, *45*, 1784–1790.
(11) Klekota, J.; Brauner, E.; Schreiber, S. L. Identifying biologically active compound classes using phenotypic screening data and sampling statistics. *J. Chem. Inf. Model.* **2005**, *45*, 1824–1836.
(12) Nilakantan, R.; Nunn, D. S. A fresh look at pharmaceutical screening library design. *Drug Discovery Today* **2003**, *8* (15), 668–672.
(13) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.
(14) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity. *J. Med. Chem.* **2002**, *45*, 4350–4358.
(15) Greenwood, P. E.; Nikulin, M. S. *A guide to chi-squared testing*; Wiley: New York, 1996.
(16) Krumrine, J. R.; Maynard, A. T.; Lerman, C. L. Statistical tools for virtual screening. *J. Med. Chem.* **2005**, *48*, 7477–7481.
(17) Hassan, M.; Brown, R. D.; Varma-O'Brien, S.; Rogers, D. Cheminformatics analysis and learning in a data pipelining environment. *Mol. Diversity* **2006**, *10* (3), 283–299.
(18) Bauer, D. F. Constructing confidence sets using rank statistics. *J. Am. Stat. Assoc.* **1972**, *67*, 687–690.
(19) Muchmore, S. W.; Debe, D. A.; Metz, J. T.; Brown, S. P.; Martin, Y. C.; Hajduk, P. J. Application of Belief Theory to Similarity Data Fusion for Use in Analog Searching and Lead Hopping. *J. Chem. Inf. Model.* **2008**, *48*, 941–948.

CI900113D