Analysis of Iterative Screening with Stepwise Compound Selection Based on Novartis in-house HTS Data

Shardul Paricharak,[‡,ε,ζ] *Adriaan P. IJzerman,*[ε] *Andreas Bender,**[,‡] *and Florian Nigsch**[,ζ]

[‡] Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, United Kingdom

[ε] Division of Medicinal Chemistry, Leiden Academic Centre for Drug Research, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

[ζ] Novartis Institutes for BioMedical Research, Novartis Pharma AG, Novartis Campus, 4056 Basel, Switzerland

*To whom correspondence should be addressed. ab454@cam.ac.uk, telephone: +44 (0) 1223 762983, or florian.nigsch@novartis.com, telephone: +41 (0) 79 292 0614.

RECEIVED DATE (to be automatically inserted after your manuscript is accepted if required according to the journal that you are submitting your paper to)

ABSTRACT

With increased automation and larger compound collections, the development of high-throughput screening (HTS) started replacing previous approaches in drug discovery from around the 1980s onwards. However, even today it is not always appropriate, or even feasible, to screen large collections of compounds in a particular assay. Here, we present an efficient method for iterative screening of small subsets of compound libraries. With this method the retrieval of active compounds is optimized using their structural information and biological activity fingerprints. We validated this approach retrospectively on 34 Novartis in-house HTS assays covering a wide range of assay biology, including cell proliferation, antibacterial activity, gene expression and phosphorylation. This method was employed to retrieve subsets of compounds for screening, where selected hits from any given round of screening were used as starting points to select chemically and biologically similar compounds for the next iteration. By only screening ~1% of the full screening collection (~15,000 compounds), the method consistently retrieves diverse compounds belonging to the top 0.5% most active compounds for the HTS campaign. For most of the assays over half of the compounds selected by the method were found to be among the 5% most active compounds of the corresponding full-deck HTS. In addition, the stringency of the iterative method can be modified depending on the number of compounds one can afford to screen, making it a flexible tool to discover active compounds efficiently.

Keywords: HTS screening; Iterative screening; *in silico* heuristic compound selection

INTRODUCTION

Early drug discovery traditionally has been the result of a close collaboration between chemists, pharmacologists and clinical scientists, where knowledge from pharmacology and (medicinal) chemistry was combined to design potentially active molecules for testing.[1,2] From around the 1980s onwards rapid improvements in automation and combinatorial chemistry led to the development and increasing acceptance of high-throughput screening (HTS), which allows rapid screening of large collections of compounds using robotics and automated data processing. This ~~allowed~~ enabled HTS to be used to study relationships between compounds and putative biological targets on a very large scale, so that libraries of 1-2 million compounds are routinely screened in big pharmaceutical companies, several times per year.[2,3] Conceptually, HTS aims to screen large numbers of molecules in a brute-force approach to identify hits, and the most promising chemical entities are then selected as starting points for further investigation. It is hoped that~~The rationale behind~~ screening large numbers of molecules ~~is that it~~ increases the chance~~s~~ of finding promising chemical entities. However, the previous often iterative cycles of design–screen–refine in small interdisciplinary project teams were somewhat lost.

Over the last few decades, HTS has hence become increasingly popular and has ~~increased~~ augmented in capacity from being able to screen tens of thousands of compounds a day to over 100,000 compounds a day, and has become – besides many other techniques – of crucial importance for early drug discovery.[4–6] However, HTS also has some significant drawbacks. Cell-free HTS campaigns, such as biochemical target-based assays, are not adequately predictive of compounds' ADMET (absorption, distribution, metabolism, excretion and toxicity) properties which are important pharmacokinetic parameters for drug development.[7] For cell-based phenotypic HTS assays, which can be more predictive of certain ADMET properties such as

bioavailability and cytotoxicity, target deconvolution is an important challenge.[8] Additionally, HTS campaigns sometimes cannot be performed at scale for complex biological systems that cannot be mass-produced (*e.g.* organoids).[9] Finally, and of most relevance for the current study, HTS remains a resource-intensive endeavor with a large fraction of the compounds screened being inactive or uninteresting. The latter renders the identification of smaller screening sets which lead to a significant fraction of active chemical matter detected very relevant.[4]

The mentioned drawbacks prompted efforts to optimize various aspects of HTS campaigns, such as compound library design (for example, based on chemical diversity, where libraries are chosen on the basis of chemical knowledge),[10–14] post-HTS data analysis for triaging active compounds (in order to select subsets for further validation)[15,16] and selecting novel compounds similar to active compounds detected in the assay for further investigation.[17–21] Given the recent perceived ineffectiveness of target-based HTS,[22] a shift to phenotypic HTS has occurred,[8] hence increasing the need for target identification methods. In this regard, a high-throughput screening fingerprint (HTS-FP) capturing past performance of compounds across a number of screens was developed by Petrone *et al*. at Novartis,[23] which allows the comparison of compounds according to their bioactivity across a range of HTS assays. This approach was used for both similarity searching and various machine learning methods for target identification of hits from phenotypic screens. Later, a public version of the same fingerprint was developed and analyzed by Dančík *et al.*,[24] who also reported its usefulness in the elucidation of compound mode of action. However, despite these computational advances in post-screen analysis, HTS campaigns remain an expensive endeavor.

In this study, we aim to address efficient ways of screening subsets of compound libraries, instead of screening entire compound libraries, while at the same time optimizing the retrieval of

active compounds. We developed and retrospectively validated an iterative screening method on Novartis in-house HTS data, in which selected hits from any given round of screening were used as starting points to select chemically and biologically similar compounds for the next iteration. This approach was developed with the explicit aim to select much smaller subsets of compounds with enriched activity, by harnessing the bioactivity information of compounds in the previous iteration. While briefly mentioned by Mayr *et al.* as an idea,[5] and used on a small scale by Keenan *et al.* for the design of plasmodial kinase inhibitors,[25] the concept of iterative screening has not been explored systematically in the published literature. A related concept has been previously described by Schneider *et al.* in the context of iterative virtual synthesis and testing of individual molecules, where molecules are designed automatically using evolutionary algorithms and particle swarm optimization.[26] However, our approach differs considerably, because we iteratively generate sets of molecules instead of individual molecules, hence investigating the concept on a much larger scale.

METHODS

HTS data

Novartis proprietary HTS assays comprising at least 1,300,000 compounds with an inhibitory assay readout were used, resulting in a total of 34 assays, of which 11 were cell-based assays and 23 were cell-free (biochemical) assays. These assays covered a wide number of biological events, including cell proliferation, antibacterial activity, gene expression and phosphorylation (Supplementary Table S1).

Starting set for initial screening round

We used a starting set of well-studied and manually curated compounds, many with tested clinical relevance, known to cover a large amount of druggable bioactivity space and of which the mechanism of action (MoA) is known. This set (the MoABox) comprised 2,757 compounds and is used as a starting point for many phenotypic screening projects at Novartis due to the high-quality annotations of each compound. The physicochemical properties and the chemical and biological diversity of the MoABox were calculated using RDKit[27] (Supplementary Figures S1 and S2). The design of the MoABox inherently entails that most compounds have properties favorable for cell-based screening. Owing to operational turnover of the compound archive, not every full-deck HTS contains every compound of the MoABox. Therefore, the starting set for each specific assay was the MoABox compounds present in it at the time it was performed. The smallest starting set comprised 2,050 compounds, whereas the largest comprised 2,692 compounds.

In order to determine the importance of the starting set for good performance, we repeated our analysis with 10 randomly chosen starting sets and the results were compared to those obtained

with the MoABox as starting set. These sets were obtained by repeatedly selecting a random subset from the entire screening deck of equal size to that of the MoABox present in the corresponding assay, minus any MoABox compounds that might have been coincidentally selected.

Iterative screening algorithm (ISA)

For any given set of compounds we are able to look up its activities in a past assay with ~1.3M compounds. This *in silico* screening allows not only a relative ranking (according to activities within the subset) but also an absolute ranking (according to the 1.3M compounds). Our aim was to iteratively optimize the absolute ranking of subsets of compounds, thereby efficiently selecting highly active compounds and steering the screening process towards success with much smaller compound sets. Therefore, the method developed in this study consists of three iterative procedures (see Figure 1): (1) ranking of compounds based on retrospective activity data, (2) selection/triaging of hits (3) expanding from hits to close analogs based on chemical and biological similarity metrics. Since this study is a retrospective analysis on HTS data, the ranks of the compounds selected correspond to the ranks of the same compounds had they been screened in a full-deck screen. Our method is fundamentally different from a basic similarity search using active probes, because we perform a similarity search iteratively based on active compound information at every round of screening, rather than only once. Circular fingerprints[28] (SciTegic ECFP4-like) were used as features for determining chemical similarity and HTS-fingerprints (HTS-FP)[23] were used as features for determining biological similarity.

Metrics used for performance assessment

We used two criteria for evaluating compound sets at each iteration: (1) the rank distribution based on compound activity and (2) the cumulative coverage of Murcko scaffolds[29] found in the top 0.5% of compounds ranked by activity. In conjunction, these criteria assess the retrieval of not only active, but also structurally diverse sets of compounds. In the below, a median rank cutoff of 65,000 is sometimes used to assess performance; this corresponds to 5% of a total screening collection of 1.3 million compounds.

Systematic exploration of parameters

The number of compounds triaged per iteration as well as the number and types of expansions affect the size and diversity of the compound sets selected. First, the number of top-performing compounds triaged can be varied. Second, expansions can be adjusted (chemical and/or biological similarity), as well as the corresponding Tanimoto[30] similarity cut-off and maximum number of expansions per compound. Moreover, the maximum number of compounds originating from the same parent compound can be adjusted in order to limit the number of closely related analogs. We systematically explored the influence of these parameters in a number of *in silico* experiments (see Table 1), where the influence of each parameter was analyzed individually. Experiment 1 was considered as a realistic reference experiment that balances performance and the number of compounds screened over 10 iterations (~1% of entire collection, ~15,000 compounds). All other experiments varied one parameter, therefore allowing an assessment of its influence with respect to the reference experiment. For example, a comparison of experiment 3 with experiment 1 shows the effect of doubling the number of compounds triaged per iteration from 100 to 200.

Data analysis

The workflow comprised Python and Perl scripts for data analysis, the Indigo toolkit[31] and RDKit[27] for cheminformatics calculations. Spotfire[32] was used for data exploration and R[33] and Cytoscape[34] were used for the visualization of results.

RESULTS AND DISCUSSION

Here we present in detail the results belonging to the reference experiment, followed by a comparison to other experiments. Experiments 4, 6 and 7 showed the same results as the reference experiment and are therefore not discussed separately; these experiments highlight, however, that more than 50 expansions or a more stringent HTS-FP similarity cut-off do not change the results.

Iterative screening is highly effective across assay types

The median rank of the compounds selected was 36,101 (excluding the starting set) across all assay types, which corresponds to the top ~2.8% of a collection of ~1.3M compounds. In other words, half the compounds selected across all iterations (except for the starting set) are found among the top 2.8% of the corresponding 1.3M compound screen, indicating a clear enrichment in activity of the compounds selected. Of note, the performance is consistent for a large number of different assay types (median rank below 65,000, see Figure 2). However, for the types enzyme activity/cleavage assay, protein cleavage assay, protein functional assay and protein-protein binding assay the performance was reduced, as evidenced by a median rank greater than 65,000 combined with a higher standard deviation.

Interestingly, performance is better for the cell-free assays than the cell-based assays (rank distributions for both assay formats is shown in Figure 3). In order to investigate whether this difference was statistically significant, a paired *t*–test was performed for the median ranks across iterations 1 to 10. In addition, a Kolmogorov–Smirnov test was performed for every iteration on compound ranks of different assay formats. All p-values were smaller than 10^{-5}, hence indicating a statistically significant difference in distribution of rank between cell-free and cell-based

assays. This difference is likely due to the fact that in order for compounds to have an effect in cell-based assays, they have to be able to cross the cell membrane to reach the target of interest (in cases when this target is not membrane-bound). Hence, these compounds must have suitable physicochemical properties (such as permeability), in order to be effective. Since our method on purpose did not distinguish between cell-free and cell-based assays, these results are in line with expectations; however, specific compound criteria for cell-based assays (*e.g.* incorporation of logP values, past performance in cell-based assays) are likely to diminish this observed gap in performance between the two assay formats in the future. As mentioned before, the MoAbox content is geared towards hypothesis-generating cell-based phenotypic screening; as a result, this set of compounds performs equally well on cell-based and cell-free assays (Figure 3, iteration 0).

Next, median compound ranks were evaluated per assay type (Figure 4). The iterative method performs consistently well for the majority of assay types (median ranks are smaller than 100,000 for iterations 1–10 for 11 out of the 16 assay types), but there are a number of outlier assay types for which the median rank of compounds selected swiftly deteriorates after around iteration 3. These assays cover the biological events protein-protein binding, protein cleavage, protein function and enzyme activity/cleavage, and are the same ones shown to have an overall median rank above 65,000 (Figure 2). These results suggest that expansions in chemical and biological space are unable to effectively retrieve the most active compounds for these assay types after the first few iterations.

Chemical diversity analysis of iterative screening results

In addition to the rank distribution of the iteratively selected compounds, we also analyzed the percentage of highly active scaffolds cumulatively retrieved. Highly active scaffolds were

separately defined for each assay as the Murcko scaffolds[29] belonging to the top 0.5% most active molecules in the assay. While Murcko scaffolds are useful for assessing structural diversity of cyclic compounds (the definition by Bemis and Murcko[29] is based on ring systems and linkers), this measure of diversity is biased for assays where many aliphatic compounds are hits. In the absence of a more inclusive and/or appropriate definition of scaffold, the following analysis only includes chemical matter with a defined Murcko scaffold.

The average retrieval rate of highly active scaffolds after 10 iterations across all assay types is 41% (~1,600 unique scaffolds per assay, ~9 analogs per scaffold), with an average of 14,959 compounds screened across all iterations per assay. Examples of commonly retrieved scaffolds are shown in Supplementary Figure S3, where scaffold 1 is the second most commonly retrieved scaffold, corresponding to a prevalence of 1.4% in the compounds screened for all assays in iterations 1 to 10. These results indicate that our method is able to prioritize diverse chemical matter despite much smaller screening sets. In addition, it performs substantially better than a traditional similarity search as the retrieval of highly active scaffolds is only 11% in the first iteration where the similarity search would stop, compared to 41% after 10 rounds of iterative screening.

The percentage of cumulatively retrieved highly active scaffolds steadily increases with the iteration count (Figure 5), with the steepest increases occurring in the earliest iterations. Most assay types display a scaffold retrieval of ~30–45% after 10 iterations. The calcium quantification assay showed relatively poor scaffold coverage (~20% after 10 iterations), whereas the phosphorylation assay, typically used for kinase inhibitors, showed much better scaffold coverage compared to other assay types (~55% after 10 iterations). Given the presence of many series of high-quality kinase inhibitors from past drug discovery programs in the

Novartis screening archive, in combination with the promiscuity of kinase inhibitor binding,[35,36] it is likely that many active inhibitors retrieved are structurally/biologically similar. Hence, this is a possible explanation for the preferred retrieval of a higher number of active scaffolds for phosphorylation assays. Another interesting observation is that the assays for protein-protein binding, protein cleavage and enzyme activity show mediocre median ranks (>65,000), while having average scaffold retrieval rates (30-40% retrieval after 10 iterations). This suggests that while our ISA is able to retrieve many compounds present in the top 0.5% of most active compounds (to an extent comparable with the majority of other assays), many inactive compounds are retrieved as well, resulting in a higher standard deviation in rank (see Figure 2). The hypothetically best scaffold retrieval among the top 0.5% of compounds screened would be achieved by sorting the top 0.5% of compounds by activity and picking their scaffolds. The comparison between the hypothetically best scaffold retrieval and iterative scaffold retrieval rate is shown in Supplementary Figure S4. For example, after picking 5,000 compounds this best possible performance retrieves ~75% of highly active scaffolds, compared to ~10–25% of highly active scaffolds (depending on assay type) retrieved iteratively and ~0.4% that would be retrieved if selection was random. In other words, iterative screening of ~15,000 compounds recovers a third of the structural diversity of the top 5,000 compounds of a 1.3M compound screen.

The fraction of highly active scaffolds retrieved was also analyzed across all assay types. Here, we determined the fraction of highly active scaffolds for each iteration (see Figure 6). We observed that, in general, the active scaffolds which are easily identified are quickly retrieved: for the first few iterations the fraction of highly active scaffolds retrieved sharply increases from ~10% to ~30-80%, after which it slowly decreases, indicating the progressive difficulty in

finding the remaining highly active scaffolds. A possible explanation is the presence of unreachable singletons in the screening archive that are beyond the expansions we implemented thus far.

Visualization of stepwise exploration of chemical space

In order to illustrate the iterative compound selection in more detail, we showed the expansions for an inhibitory cell-free kinase assay in a network graph (see Figure 7). All compounds from the starting set (0^{th} iteration) leading to no further expansions have been omitted from the network graph, whereas those that lead to at least one further expansion are depicted on the large circle on the left part of the figure. Compounds are color-coded according to their rank (lower/better and higher/worse ranks are represented by green and red nodes, respectively) and edges are colored according to the expansion type (chemical similarity expansions are orange and biological similarity expansions are turquoise). Certain compounds from the starting set lead to very few further expansions, and hence produce very few branches. Other compounds lead to a larger number of expansions, as can be seen in the upper-right corner of the figure: all the compounds present in that subnetwork represent expansions from one single compound of the starting set. In the lower-right corner of the figure, we show an example of scaffold hopping, which is commonly observed for biological similarity (HTS-FP) expansions, enabling the method to explore chemical space that is not reachable *via* chemical similarity. In addition, the depiction of activity cliffs[37] (represented by bold and wide edges) allows the identification of scaffold hopping indicative of a relatively sharp increase in activity.

Tuning iterative screening to assay requirements

The number of compounds triaged per iteration has a large effect: as more compounds are carried forward, both the median ranks and the scaffold retrieval for compounds selected in iterations 1–10 increase (comparison of experiments 2 and 3 with reference, see Figure 8). When the number of compounds triaged was increased from 50 to 100 and from 100 to 200, median ranks of the compounds selected in iterations 1–10 increased significantly from 23,517 to 36,101 in the first case, and from 36,101 to 63,721 in the second case (paired *t*–test p-values of $1.2 \cdot 10^{-3}$ and $3.4 \cdot 10^{-4}$, respectively). Scaffold retrieval increased from 20% to 28%, and from 28% to 38% for the same comparisons, with respective paired *t*–test p-values of $1.1 \cdot 10^{-5}$ and $9.9 \cdot 10^{-6}$. Less stringent hit selection during triaging leads to more subsequent expansions and increases the total number of compounds screened. The overall net result is an increased retrieval of active scaffolds at the cost of screening more inactive compounds as evidenced by higher median ranks (the fraction of highly active scaffolds retrieved at each iteration decreases as more compounds are triaged, see Supplementary Figure S5). To illustrate the effect of varying the number of compounds triaged per iteration in more detail, we show possible expansions for an inhibitory cell-free kinase assay in a network graph (see Supplementary Figure S6). For example, the compound subnetwork and corresponding structure-activity relationships in the lower-right corner of the figure are only explored when triaging 100 or more compounds per iteration.

When investigating the dependence of scaffold coverage on fingerprint type, we found that HTS-FP-based and structure-based expansions accounted for 90% and 50%, respectively, of total highly active scaffold retrieval after 10 iterations (Supplementary Figure S7). Since HTS-FPs capture the biological profile of compounds, HTS-FP similarity leads to more structurally diverse sets of biologically similar compounds compared to structure-based expansions.

Increasing the Tanimoto[30] cut-off from 0.6 to 0.8 (comparison of experiment 5 to the reference experiment) for structure-based expansions decreased both median compound ranks from 36,101 to 16,831 (paired *t*–test p-value of $9.4 \cdot 10^{-4}$) and scaffold retrieval from 28% to 16% (paired *t*–test p-value of $2.6 \cdot 10^{-6}$). The maximum number of compounds triaged per parent compound did not have a clear effect on the diversity nor the ranks of the compounds screened. Lowering this number from 5 (reference experiment) to 2 (experiment 8) resulted in a 2% higher scaffold retrieval (paired *t*–test p-value of 0.047), whereas an increase to 10 (experiment 9) had no significant effect on either median ranks or scaffold retrieval. In summary, the number of compounds triaged was the most influential factor, which can be adjusted depending on the number of compounds one intends (or can afford) to screen.

Finally, iterative screening was repeated with 10 randomly chosen starting sets and the results were compared to those obtained with the MoABox as starting set. The latter resulted in better median ranks only until the first iteration and virtually identical median ranks from iteration two onwards, and slightly higher scaffold retrieval throughout all iterations (Supplementary Figure S8). While minor differences across starting sets can be observed, the key findings presented in this study are independent of the precise composition of the starting set. However, availability of a high-quality starting set, as the MoAbox for us, can provide biological insight early on through comprehensive compound annotations.

CONCLUSION

Even though alluded to in the literature and theoretically appealing, no comprehensive practical evaluation of iterative screening was published. In this study ~~We~~ we have performed an unequalled large-scale validation of iterative screening on 34 HTS assays comprising at least 1,300,000 compounds and showed greatly improved efficiency over conventional HTS campaigns. For most assays, half of the compounds found by iterative screening of only 1% (~15,000 compounds) of the entire collection correspond to the top 5% of the full collection screen. Put differently, screening only 1% of the collection provides ~7,500 top-quality hits for further optimization. On average, the compounds selected covered over 40% of the scaffolds belonging to the top 0.5% most active compounds for each assay, hence also ensuring structural diversity. Our method allows for exit points during the iterative screening process: performing large numbers of iterations is not necessary in order to retrieve active compounds, as they are retrieved starting from the 1st iteration already, and therefore, a large investment in resources upfront is not required. As expected, the method in its current state performs better for cell-free assays compared to cell-based assays; a future improvement can gear towards physicochemical properties more adapted to cell-based screens.

We used network graphs to visualize the compound selection process, and to highlight activity cliffs,[37] scaffold hopping and the effect of changing the number of compounds triaged (which was found to have the largest influence on compound selection). As an outlook for further refinement of our method, we propose (1) investigating activity cliffs[37] (to be able to prioritize expansion types) and (2) employing iteratively-retrained machine-learning methods[20] to rank the screening collection in parallel to the structure-based and HTS-FP-based expansions currently performed. We believe that the iterative method developed here can easily be fine-tuned for

specific assay types, provides multiple exit points and can potentially lead to considerable savings in both time and resources.

Figures section

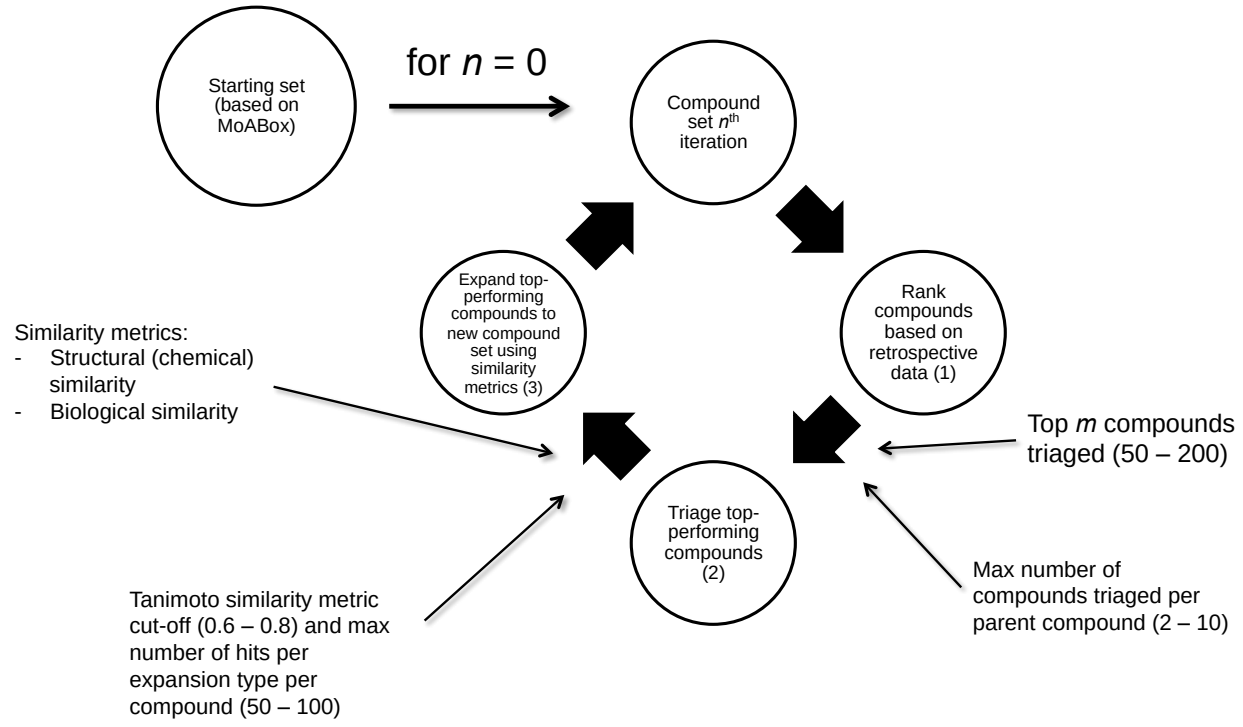


Figure 1. Iterative screening algorithm (ISA) overview. The ISA developed in this study consisted of three iteration steps: (1) ranking of compounds based on retrospective data, (2) triaging of (*i.e.* selecting) top-performing compounds and (3) expanding from top-performing compounds to close analogs based on chemical and biological similarity metrics. The starting set comprises the MoABox compounds present in the HTS assay. The ISA allows for adjustment of parameters at the triaging stage (the number of compounds carried forward, and the number of compounds originating from the same parent compound to limit large numbers of closely related analogs). At the expansion stage, the parameters used (chemical and/or biological similarity) can be adjusted, as well as the corresponding similarity cut-off and maximum number of expansions per compound.

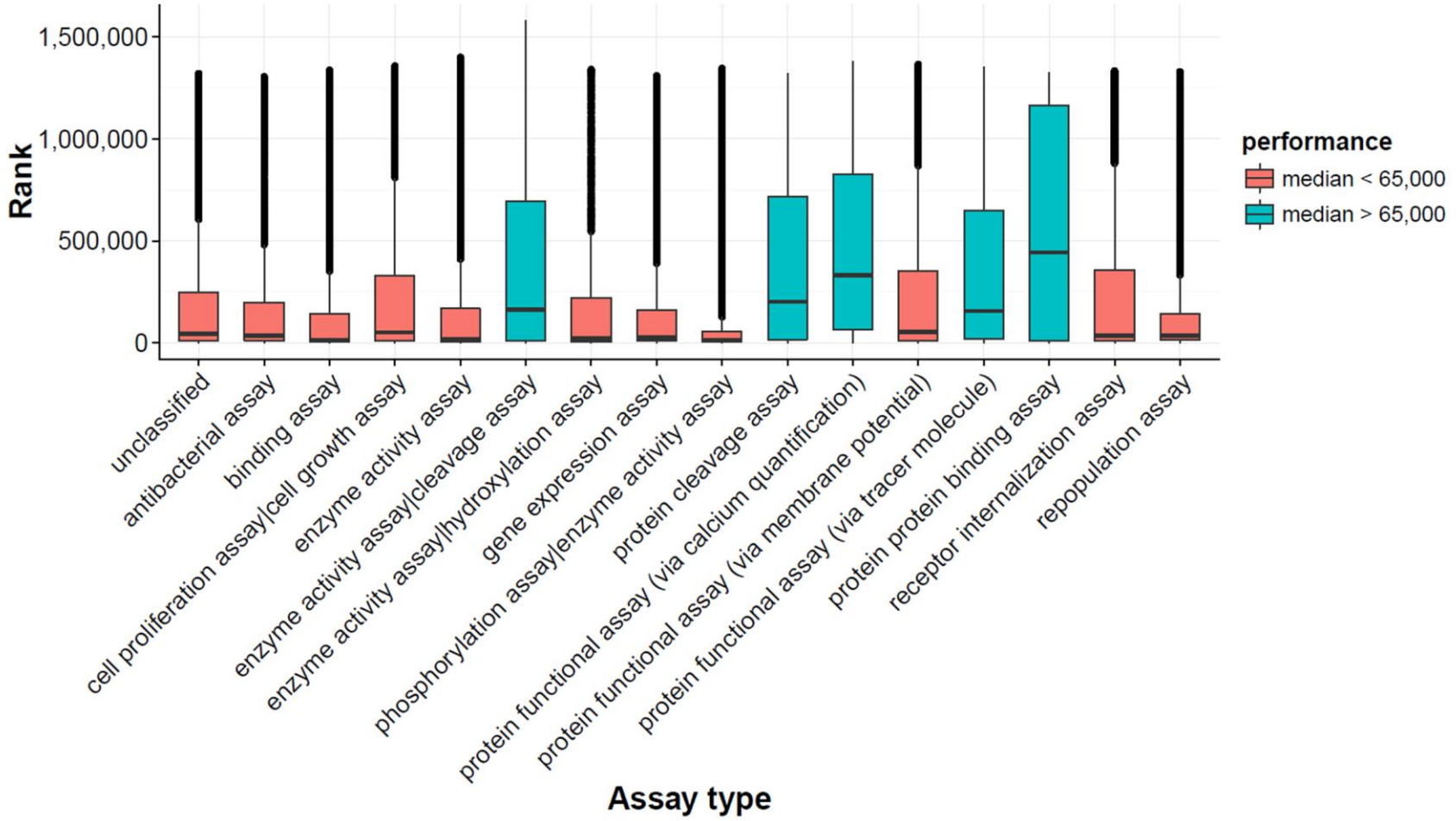


Figure 2. Ranks of compounds from iterations for all assay types. Boxplots of ranks for all compounds selected by the ISA for iterations 1–10 (excluding the starting set) are represented for each assay type. The performance for enzyme activity/cleavage assay, protein cleavage assay, protein functional assay and

protein-protein binding assay is much worse (median rank of 200,000 on average) compared to other assays, with also a broader rank distribution. Blue: median rank below 65,000; red: median rank above 65,000. The first 65,000 compounds correspond to the top ~5% of 1.3M.

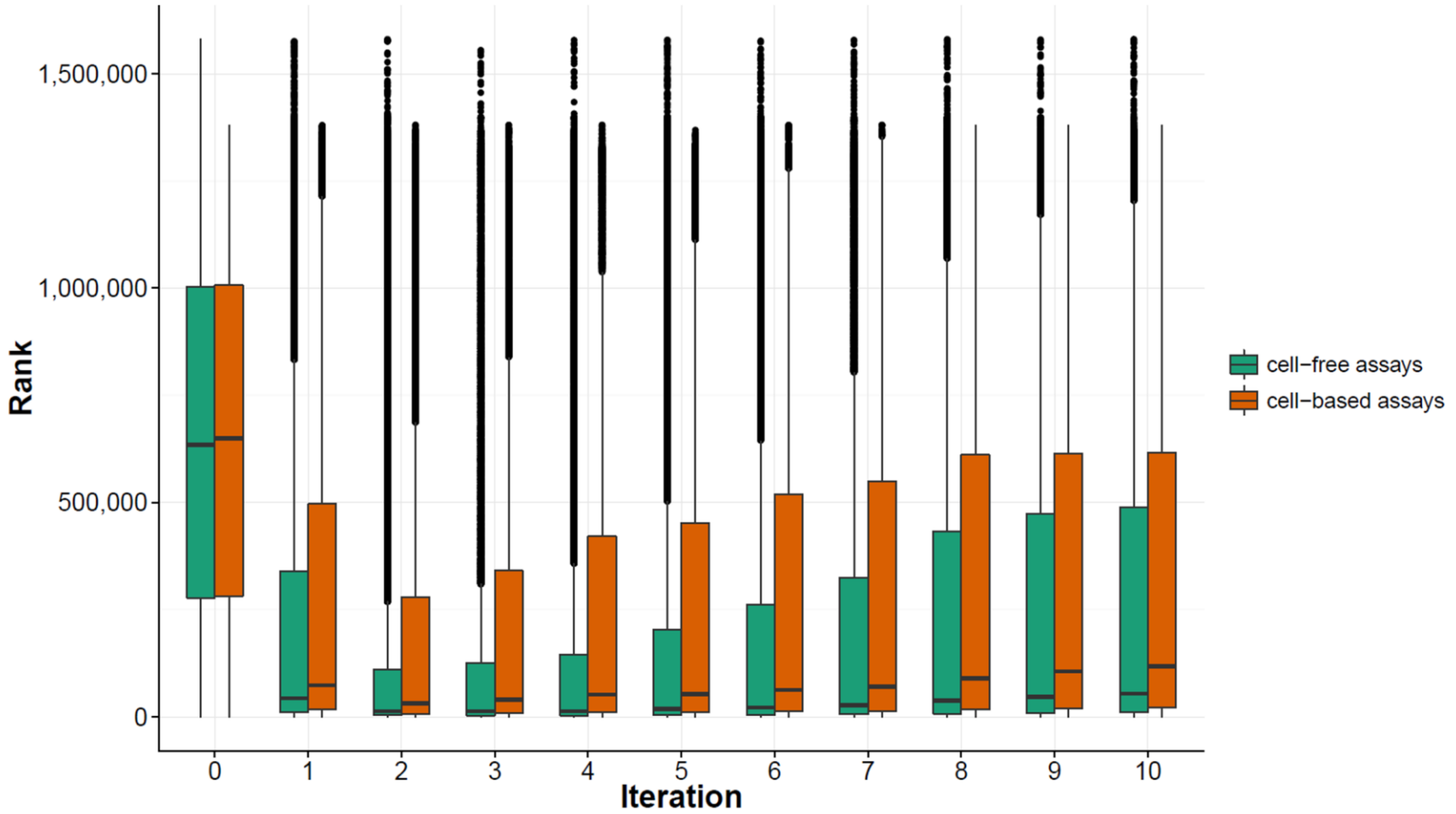


Figure 3. Ranks of iteratively selected compounds for cell-free and cell-based assays. Green: cell-free assays, orange: cell-based assays. There is a consistent difference in median rank (and interquartile range, extension of boxplot) across iterations 1 to 10 between cell-free and cell-based assays. This indicates the relative difficulty in selecting compounds that are able to satisfy cell-based screening requirements (*e.g.* cell permeability). Median ranks are significantly different (paired *t*–test, p-value < 10^{-5}), as are the rank distributions for each iteration (Kolmogorov–Smirnov test, p-value < 10^{-5}).

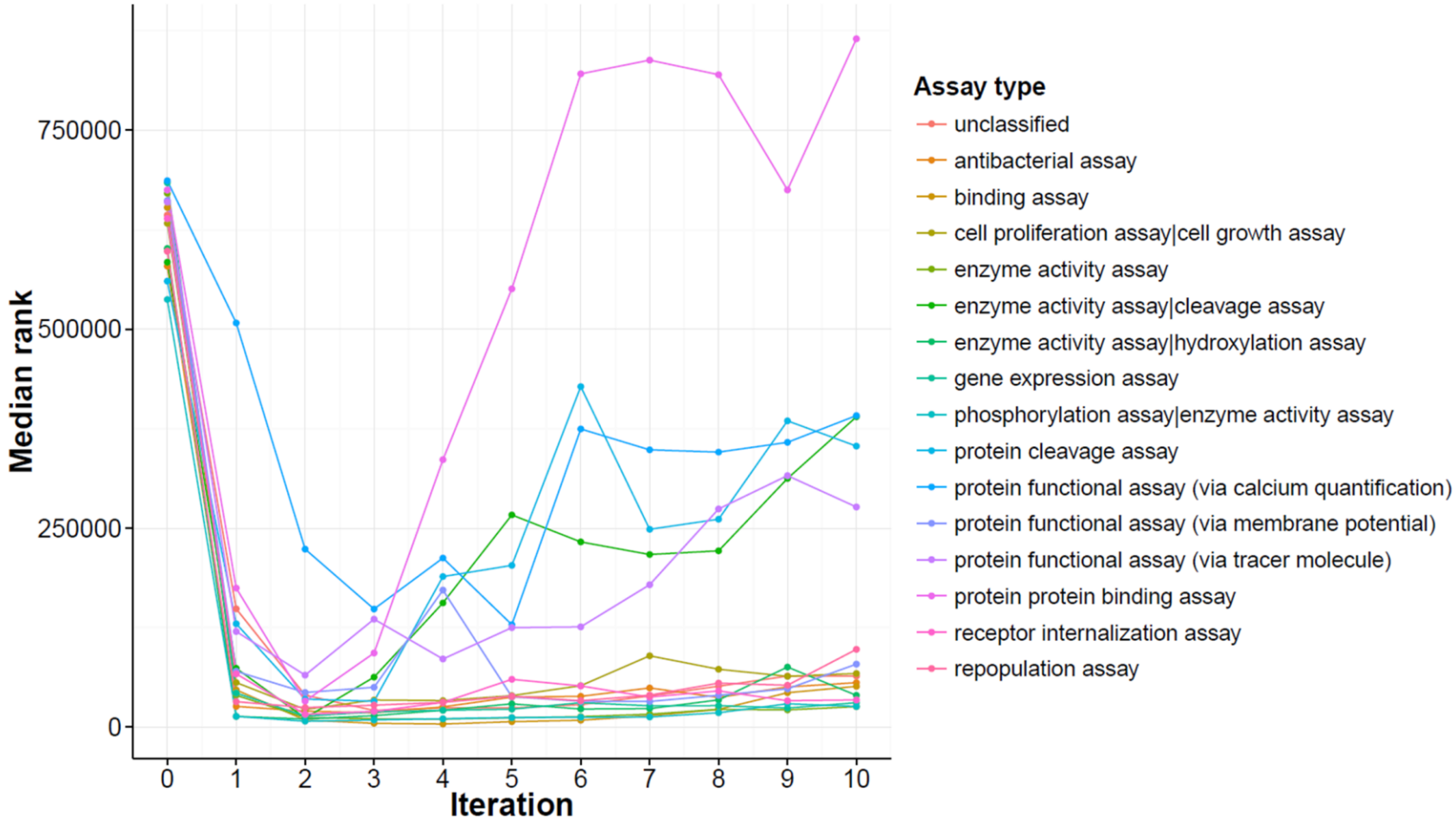


Figure 4. Median rank per iteration across assay types. The median rank of the compounds of the selected subset at each iteration is plotted versus iteration. The ISA performs consistently well for most assay types, but there are a number of assays for which the median rank of compounds selected swiftly deteriorates after around iteration 3. These assays are for protein-protein binding, protein cleavage, protein function and enzyme activity/cleavage, and are the same ones shown to have an overall median rank greater than 65,000 (Figure 2).

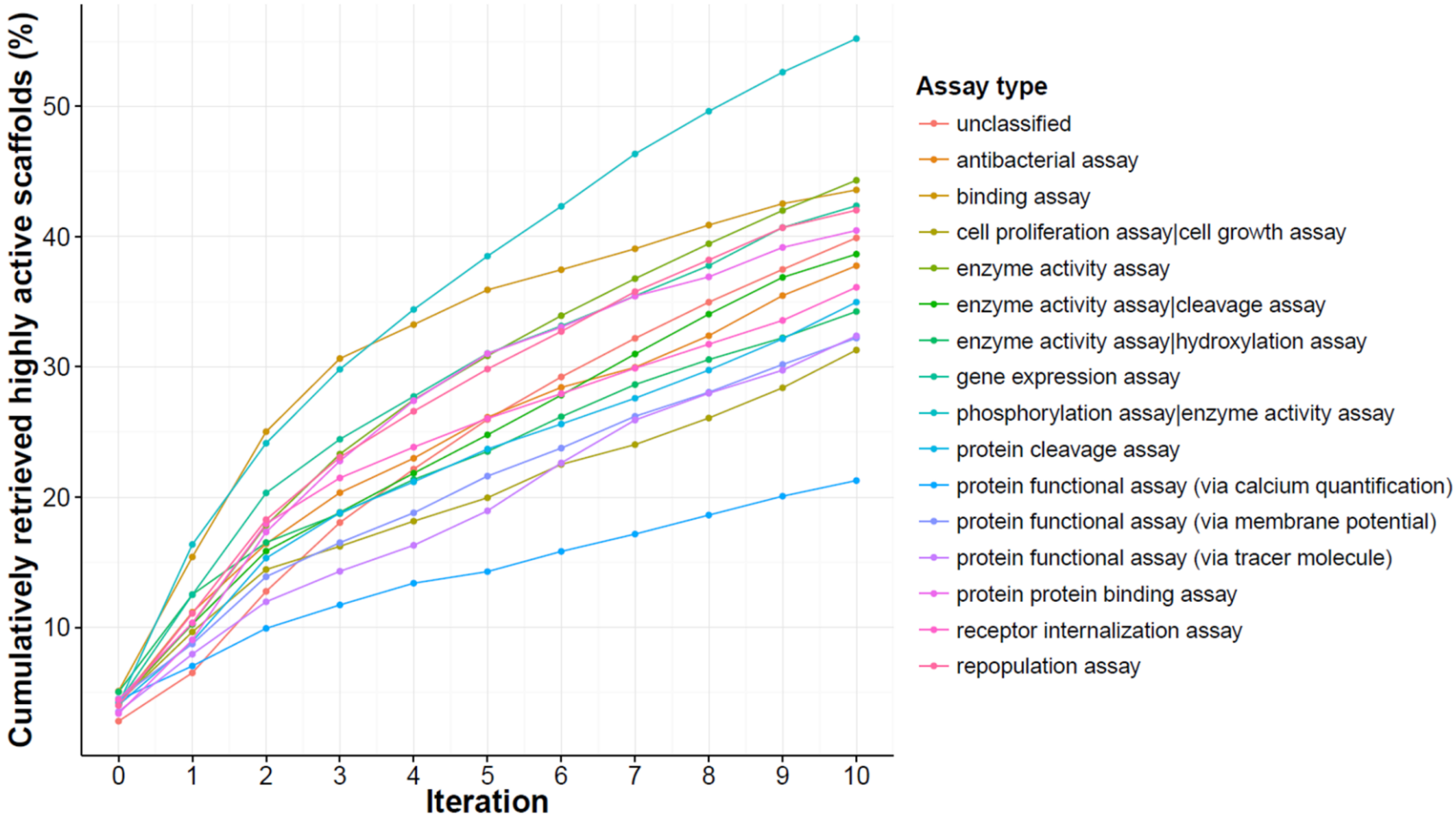


Figure 5. Cumulatively retrieved highly active scaffolds (%). For all assay types, the percentage of cumulatively retrieved highly active scaffolds (scaffolds of the 0.5% most active compounds of the full HTS) steadily increases, with the steepest increases occurring in the earliest iterations. Most assay types display a scaffold retrieval of between ~30–45% after 10 iterations. The calcium quantification assay showed relatively poor scaffold coverage (~20% after 10 iterations), whereas the phosphorylation assay showed much better scaffold coverage compared to other assay types (~55% after 10 iterations).

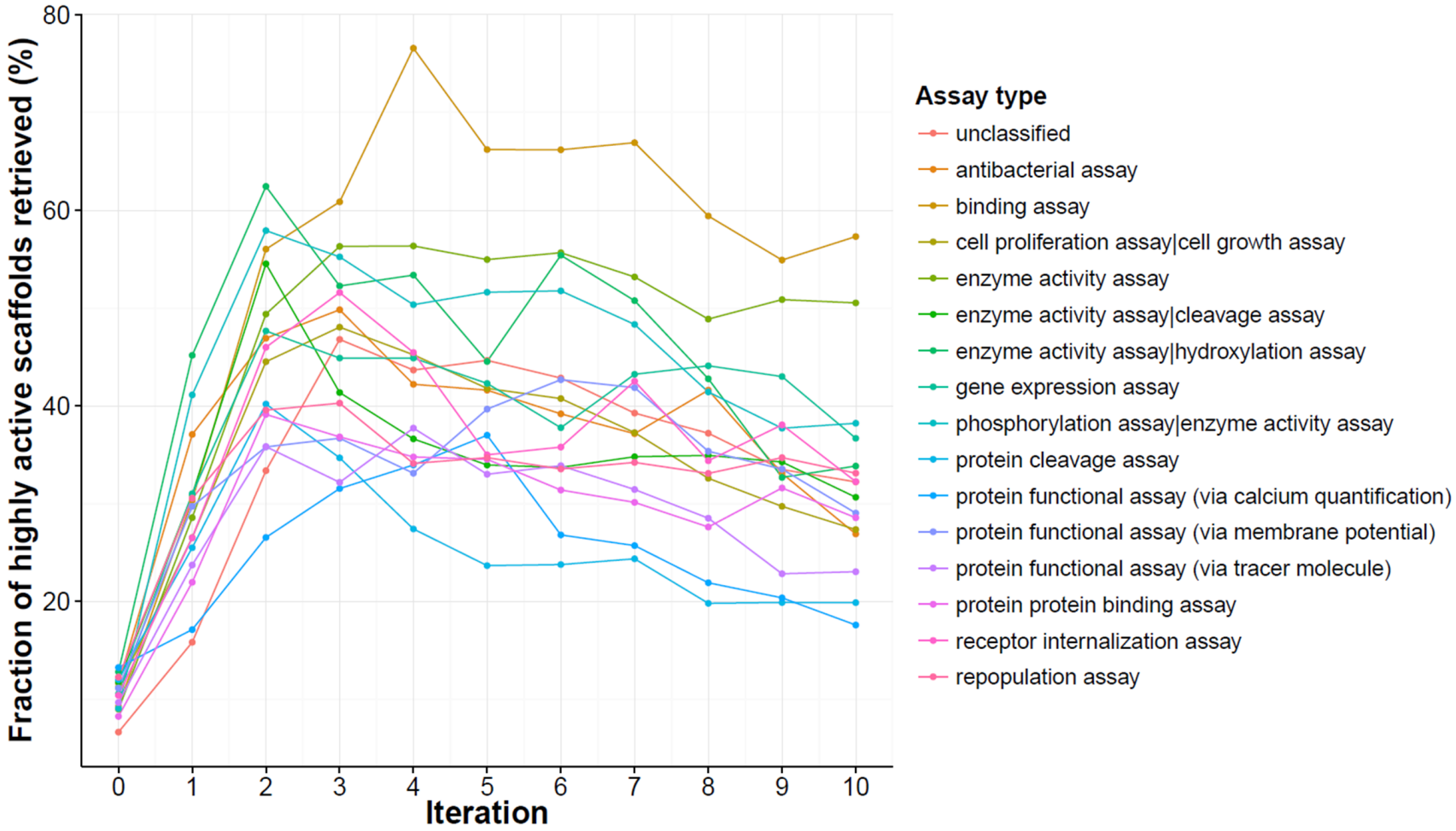


Figure 6. Fraction of highly active scaffolds retrieved (%). The ISA exhibits a general trend for all assays: for the first 2 or 3 iterations the fraction of highly active scaffolds retrieved per iteration sharply increases from ~10% to 30-80% depending on assay type (the active scaffolds which are easy to identify are quickly retrieved), after which it slowly decreases, as it becomes increasingly difficult to find the remaining highly active scaffolds. Nevertheless, active scaffolds are still retrieved at the last iterations.

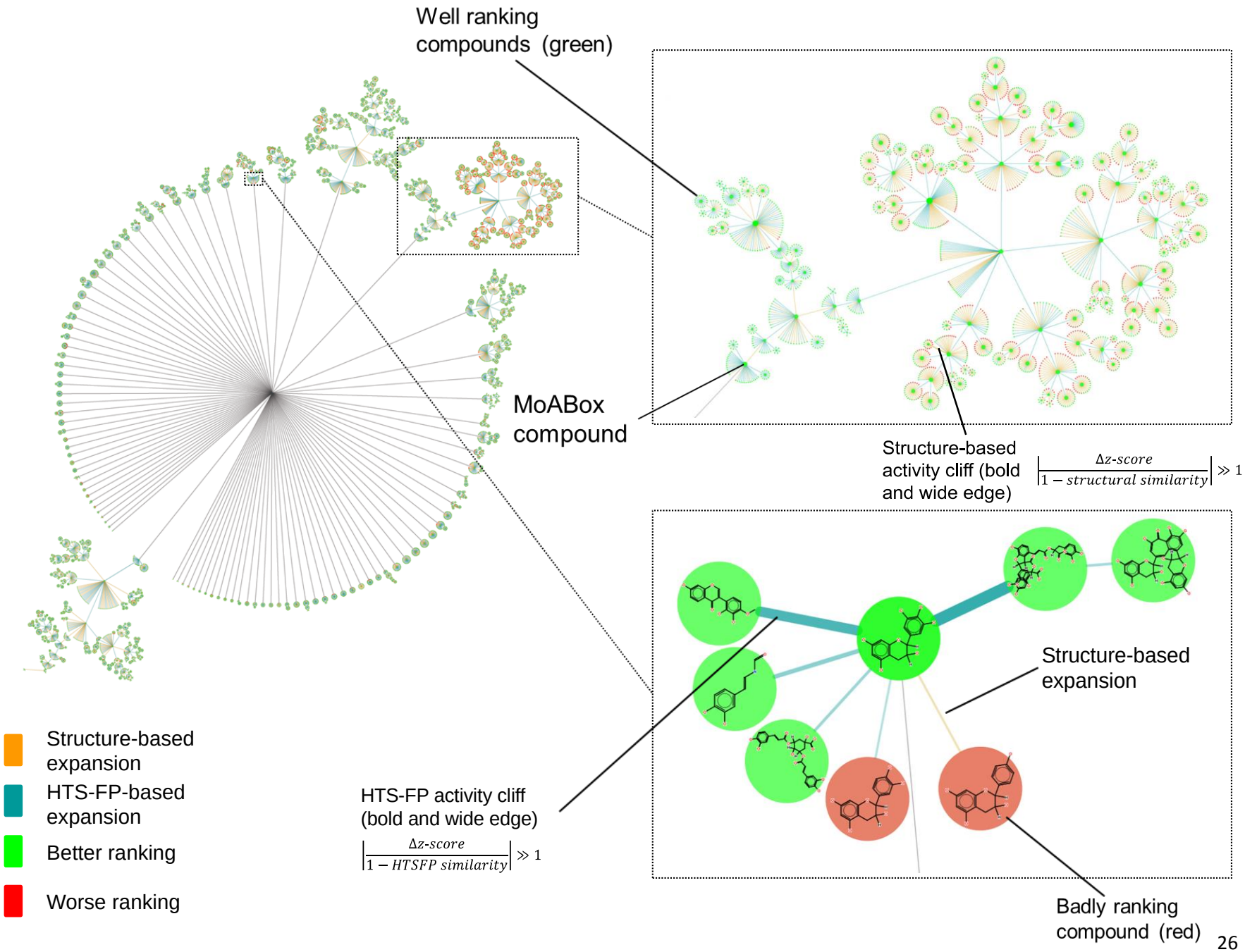

Well ranking compounds (green)
MoABox compound
Structure-based activity cliff (bold and wide edge)
$\left|\frac{\Delta z\text{-}score}{1 - structural\ similarity}\right| \gg 1$
Structure-based expansion
HTS-FP activity cliff (bold and wide edge)
$\left|\frac{\Delta z\text{-}score}{1 - HTSFP\ similarity}\right| \gg 1$
Badly ranking compound (red)
Structure-based expansion
HTS-FP-based expansion
Better ranking
Worse ranking

Figure 7. Visualization of stepwise exploration of chemical space for an inhibitory cell-free kinase assay. Expansions for an inhibitory cell-free kinase assay are shown in a network graph. All compounds from the starting set (0^{th} iteration) leading to no further expansions have been omitted from the network graph, whereas those that led to at least one further expansion are depicted on the large circle on the left part of the figure. All the compounds present in the subnetwork in the upper-right corner of the figure represent expansions from one single compound from the starting set (MoABox). In the lower-right corner of the figure, we show an example of scaffold hopping, which is commonly caused by expansions based on biological similarity (HTS-FP), enabling the method to explore chemical space that is not reachable *via* expansions based on chemical similarity. In addition, the depiction of activity cliffs[37] (represented by bold and wide edges) allows the identification of scaffold hopping leading to relatively sharp increases in activity.

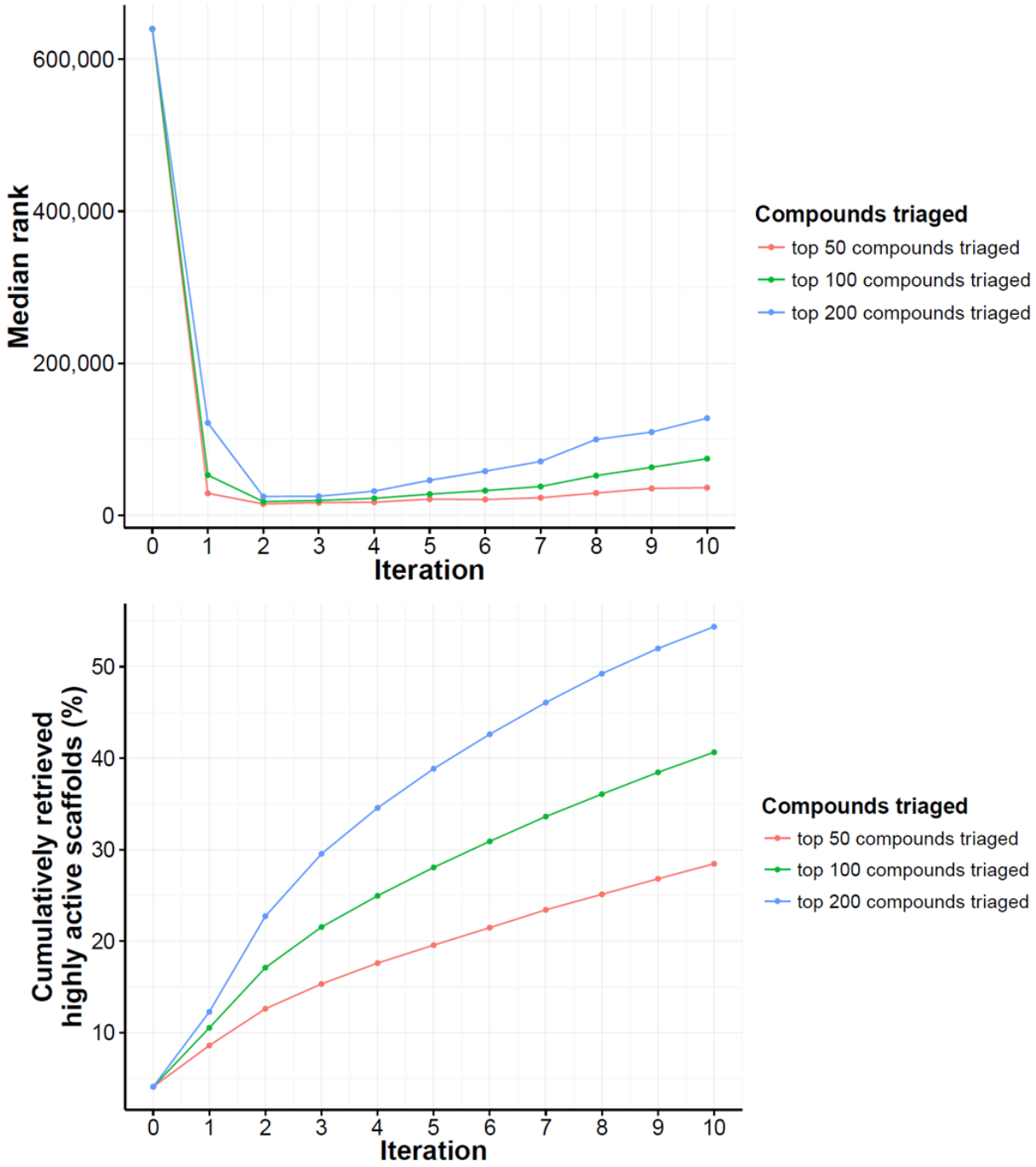


Figure 8. Effect of varying the number of compounds triaged per iteration in terms of median compound rank and percentage cumulatively retrieved highly active scaffolds. As the number of compounds triaged increases, the median ranks consistently increase for iterations 1 to 10, whereas scaffold retrieval is higher as well. These results are in accordance with our expectations: as the number of triaged compounds is increased (*i.e.* a less

stringent selection criterion is applied for compound triaging), more expansions take place and more compounds are screened overall.

Tables section

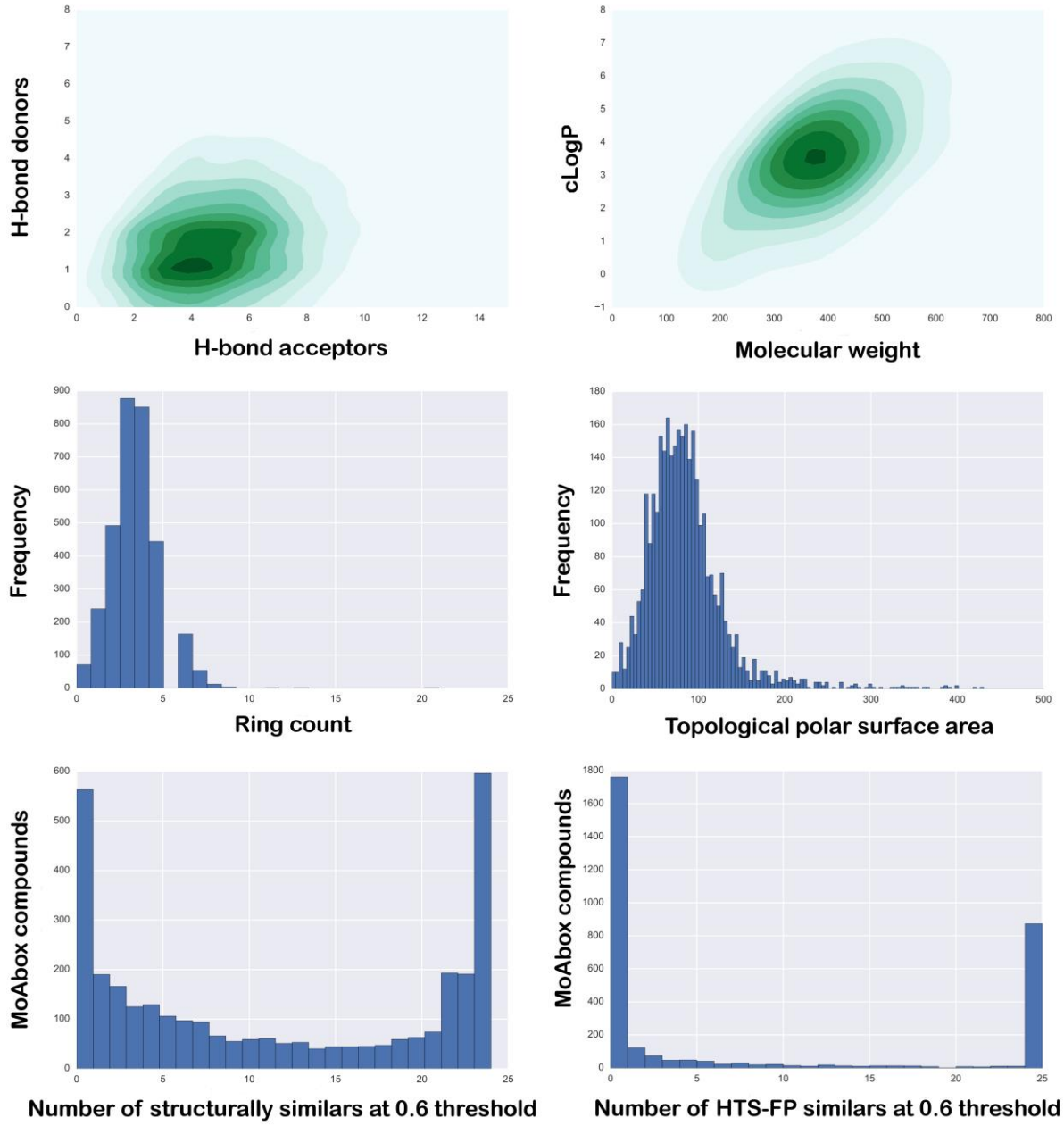
Table 1. Summary of parameters explored over 9 *in silico* experiments. Here, experiment 1 was considered as the reference experiment, which was chosen on basis of a trade-off between number of compounds screened over 10 iterations (approximately 1% of screen size) and performance. All other experiments varied one parameter, therefore allowing an assessment of its influence with respect to the reference experiment. For example, a comparison of experiment 3 with experiment 1 shows the effect of doubling (from 100 to 200) the number of compounds triaged per iteration.

Experiment number	Iteration count	Triaged number of compounds	Maximum number of expansions (structure-based)	Tanimoto cut-off (structure-based)	Maximum number of expansions (HTS-FP-based)	Tanimoto cut-off (HTS-FP-based)	Maximum number of compounds triaged per parent compound
1	10	100	50	0.6	50	0.6	5
2	10	50	50	0.6	50	0.6	5
3	10	200	50	0.6	50	0.6	5
4	10	100	100	0.6	50	0.6	5
5	10	100	50	0.8	50	0.6	5
6	10	100	50	0.6	100	0.6	5
7	10	100	50	0.6	50	0.8	5
8	10	100	50	0.6	50	0.6	2
9	10	100	50	0.6	50	0.6	10

Supporting Information

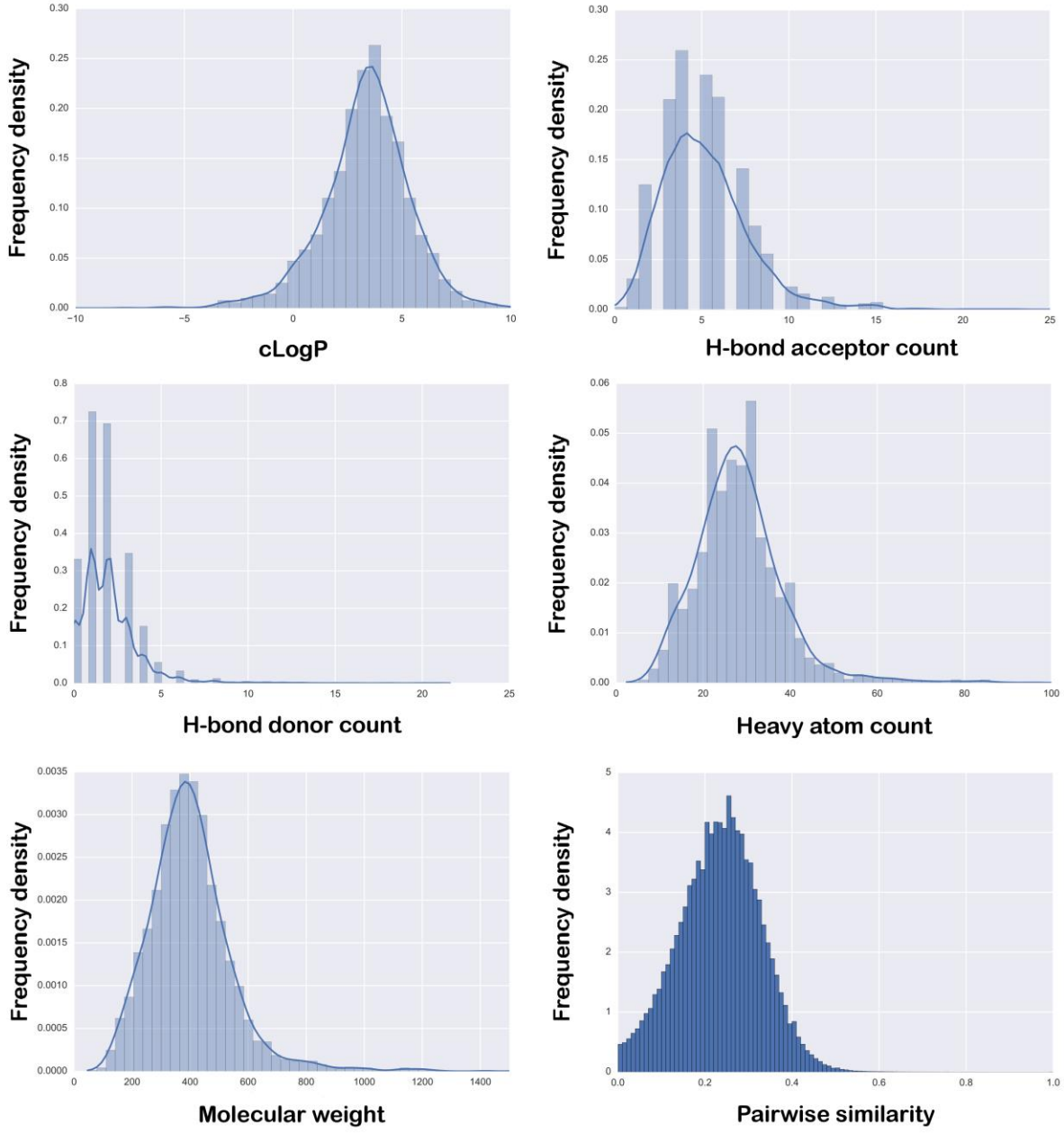
Supplementary Table S2. Assay types used for retrospective validation. All 34 assays comprise at least 1,300,000 compounds.

Biological events	Number of HTS assays
Antibacterial activity	1
Binding activity	2
Cell proliferation	1
Enzyme activity	14
Gene expression	1
Phosphorylation/enzyme activity	3
Protein cleavage	1
Protein function	5
Protein protein binding	1
Receptor internalization	1
Repopulation	1
Unclassified	3



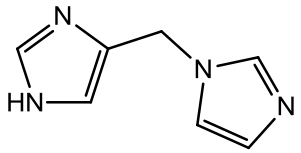
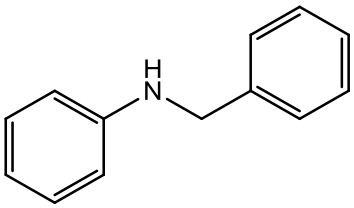
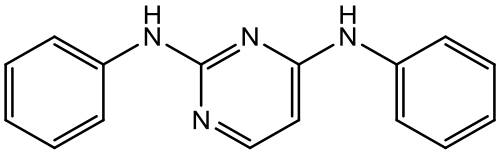
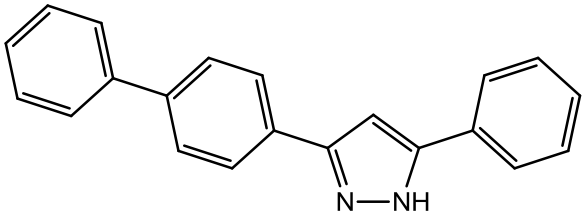
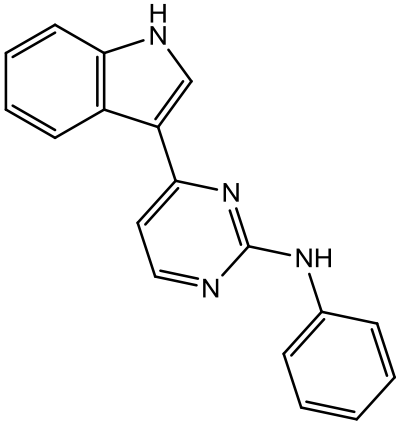
Supplementary Figure S1. Overview of physicochemical properties, and chemical and biological diversity of the MoABox. The following properties are summarized for the MoABox: H-bond donors, H-bond acceptors, molecular weight, cLogP, ring count, topological polar surface area. The bottom two figures show how many nearest neighbors can be found in the entire screening collection for how many MoABox compounds. A Tanimoto[30]

cut-off of 0.6 was used to define (structurally or biologically) similar compounds. The MoABox compounds have properties favorable for cell-based screening.



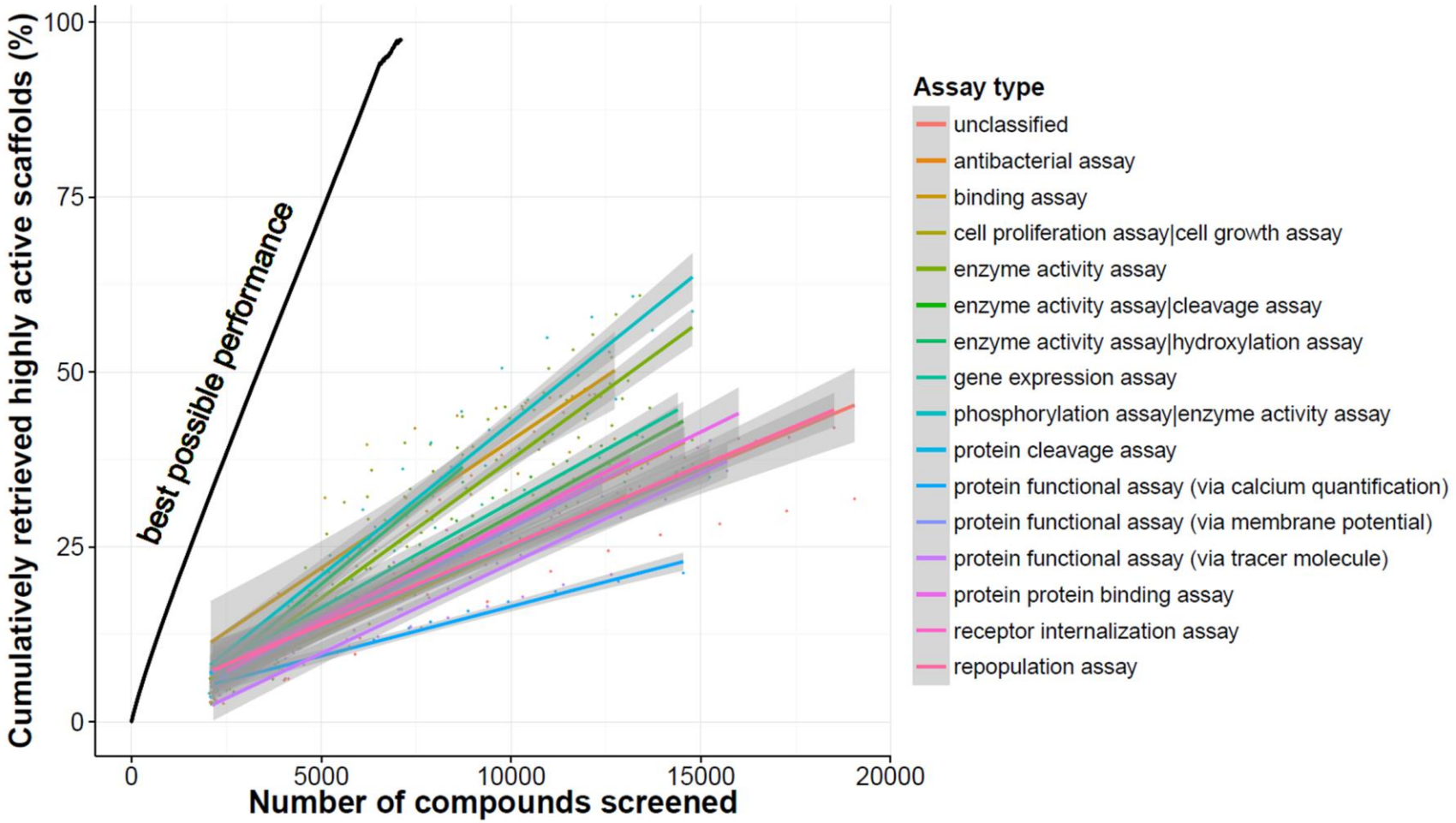
Supplementary Figure S2. Overview of frequency densities of physicochemical properties and pairwise similarity of the MoABox. The frequency densities of the following properties are shown for the MoABox: cLogP,

H-bond acceptor count, H-bond donor count, heavy atom count, molecular weight and pairwise (structural) similarity. The MoABox compounds have properties favorable for cell-based screening.

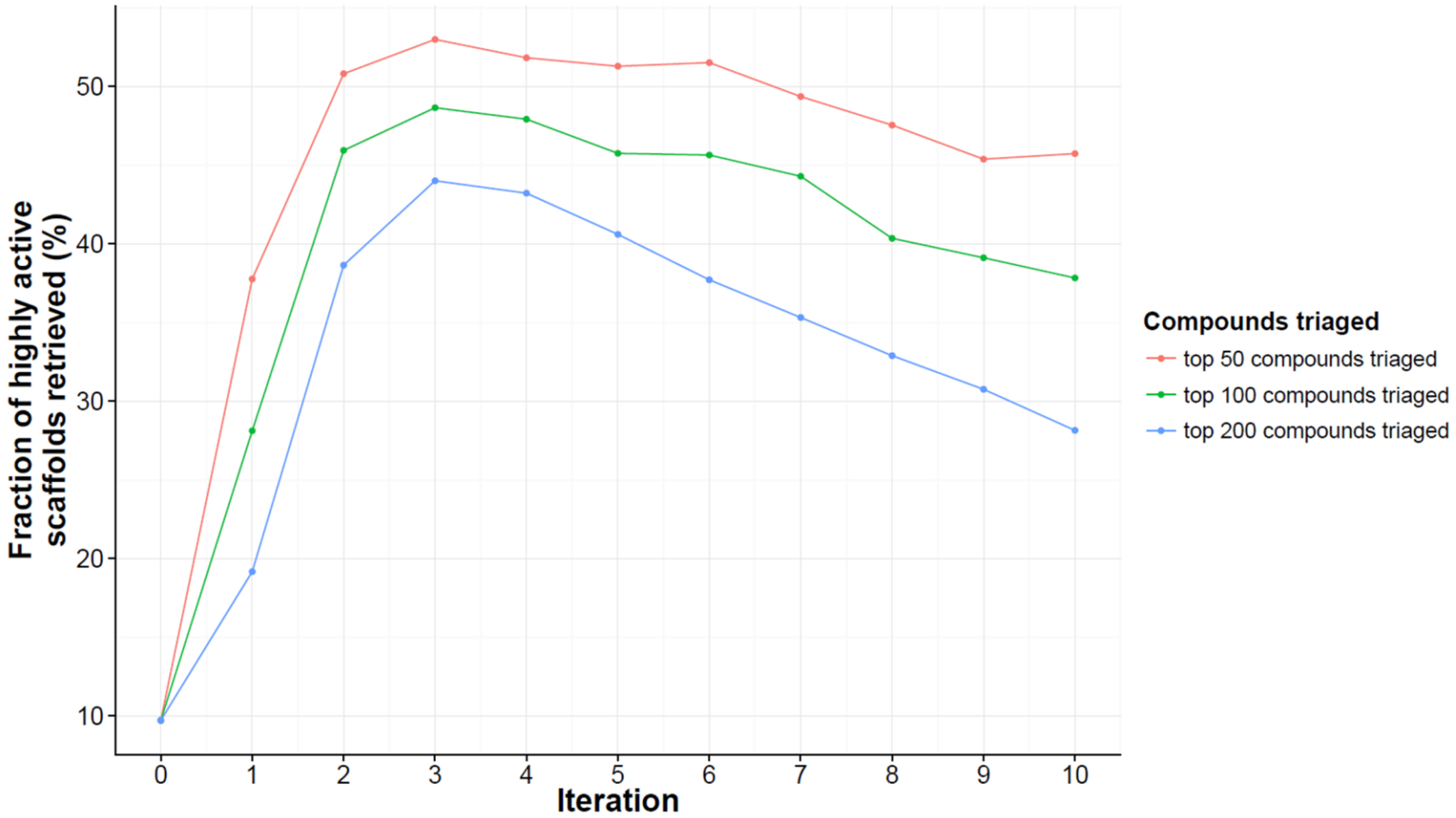
Scaffold	Prevalence
 1.	1.4%
 2.	0.93%
 3.	0.50%
 4.	0.14%
 5.	0.14%

Supplementary Figure S3. Examples of commonly retrieved scaffolds for iterations 1–10 for all assay types. While the scaffolds are not rank ordered according to their prevalence (for example, many scaffolds with a prevalence higher than 0.14% are not shown in the figure), scaffold 1 is the second most commonly retrieved

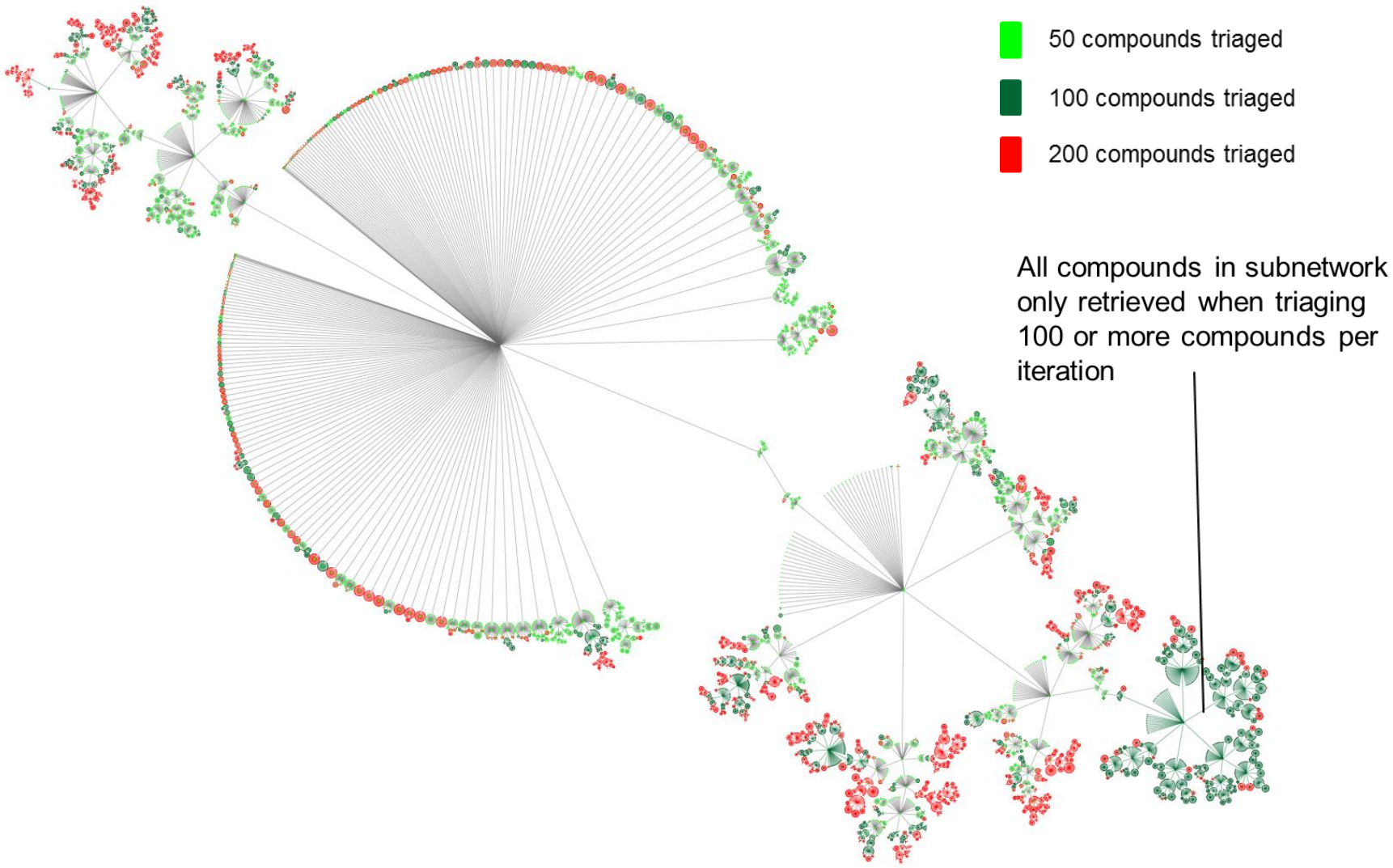
scaffold by the ISA, corresponding to a prevalence of 1.4% in the compounds screened for all assays in iterations 1 to 10. This indicates that the diversity of the compound sets selected by the ISA is large.



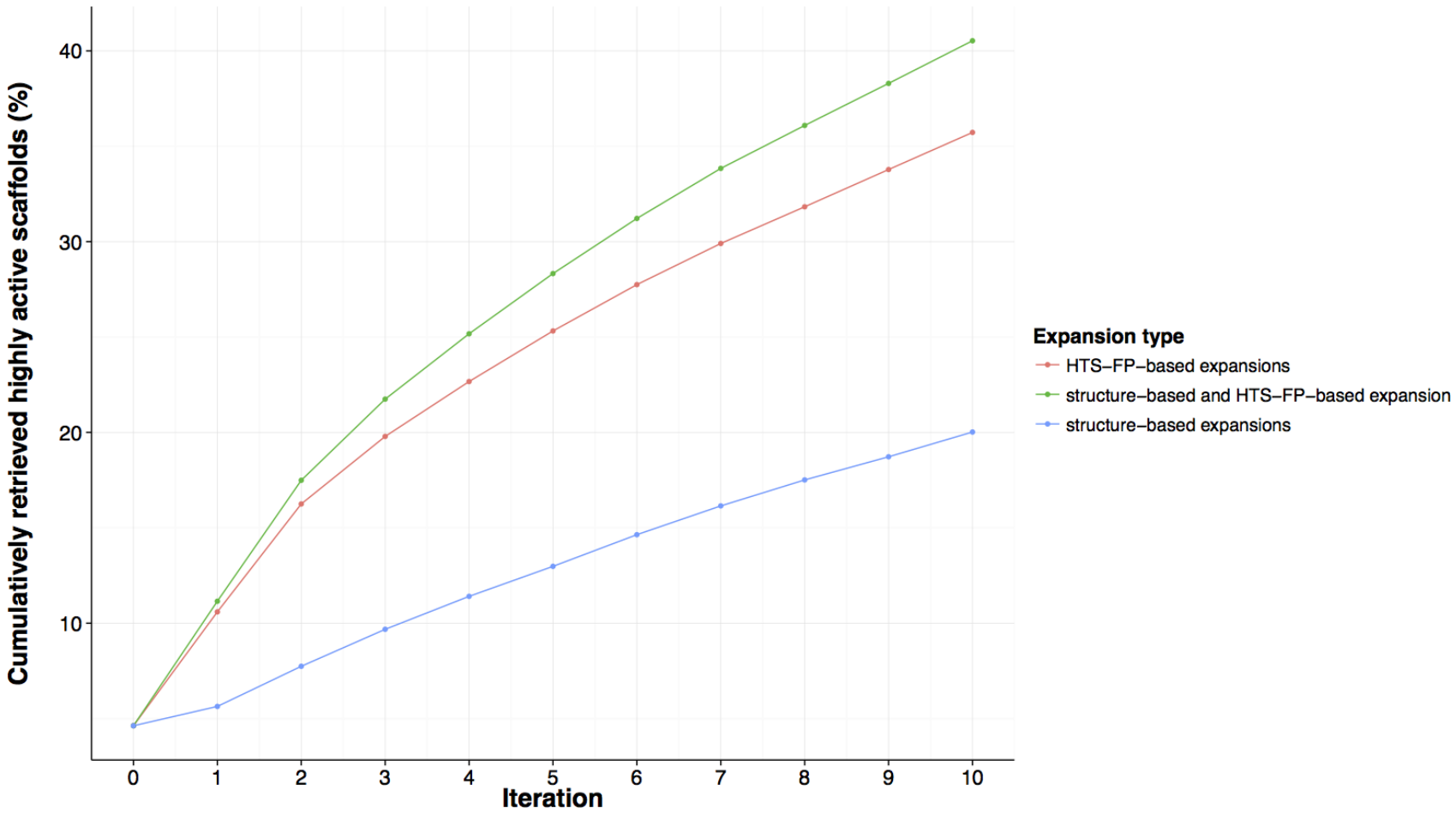
Supplementary Figure S4. Cumulatively retrieved highly active scaffolds (%) for all assay types compared to the "best possible performance". The gray area represents the 95% confidence interval for the corresponding assay type based on a linear model. When the "best possible performance" retrieves ~75% of highly active scaffolds, the ISA retrieves approximately ~10-25% of highly active scaffolds (depending on assay type).



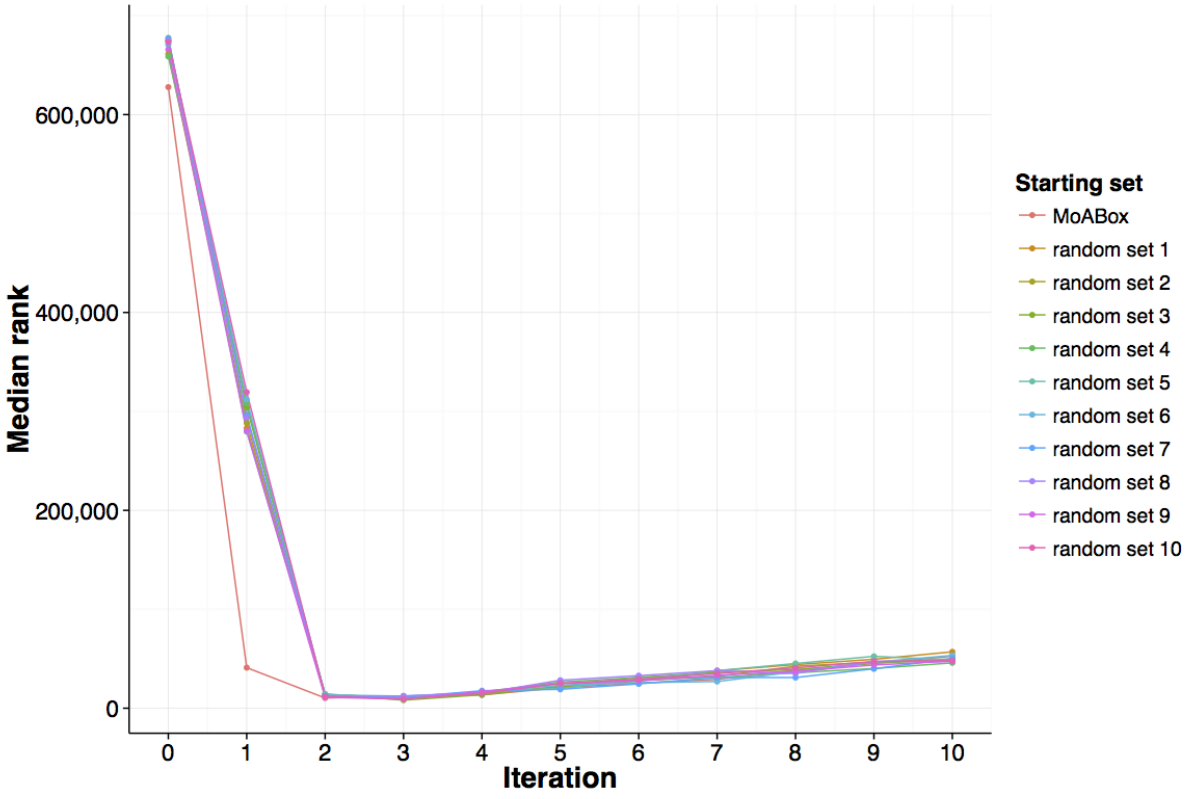
Supplementary Figure S5. Effect of varying the number of compounds triaged per iteration in terms of the fraction of highly active scaffolds retrieved (%). As more compounds are triaged, more compound expansions take place and more compounds are screened overall. As a consequence, the fraction of highly active scaffolds retrieved for every iteration decreases consistently.

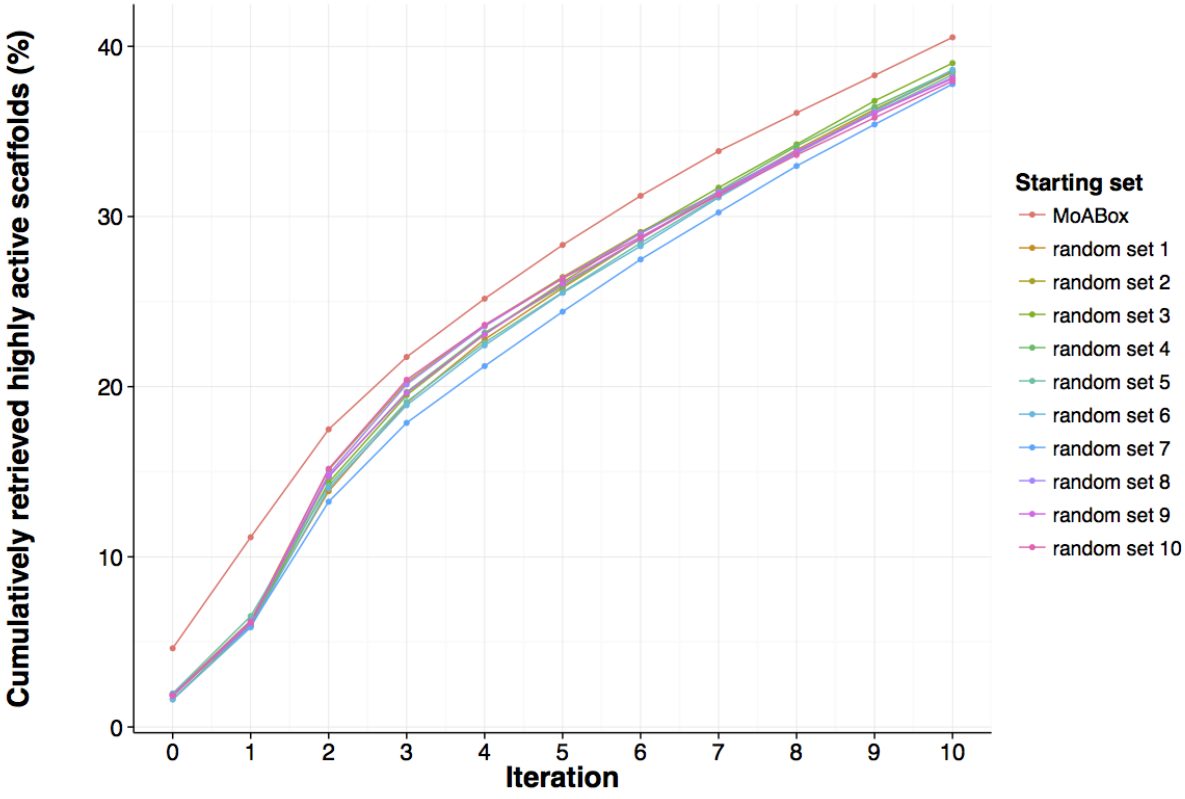

50 compounds triaged
100 compounds triaged
200 compounds triaged
All compounds in subnetwork only retrieved when triaging 100 or more compounds per iteration

Supplementary Figure S6. Illustration of the effect of varying the number of compounds triaged per iteration using networks for an inhibitory cell-free kinase assay. As more compounds are triaged, more compound expansions take place and more compounds are screened overall. For example, the compounds belonging to the subnetwork in the lower-right corner of the figure can be found by the algorithm by triaging 100 or more compounds per iteration (these compounds are not found in case only 50 compounds are triaged per iteration).



Supplementary Figure S7. The effect of expansion type on highly active scaffold retrieval. HTS-FP-based expansions accounted for 90% of total highly active scaffold retrieval, whereas structure-based expansions only accounted for 50%. This can be explained by the fact that HTS-FP capture the biological similarity between compounds and therefore, HTS-FP-based expansions lead to more structurally diverse sets of biologically similar compounds, compared to structure-based expansions.


Median rank
600,000
400,000
200,000
0
0
1
2
3
4
5
6
7
8
9
10
Iteration
Starting set
MoABox
random set 1
random set 2
random set 3
random set 4
random set 5
random set 6
random set 7
random set 8
random set 9
random set 10

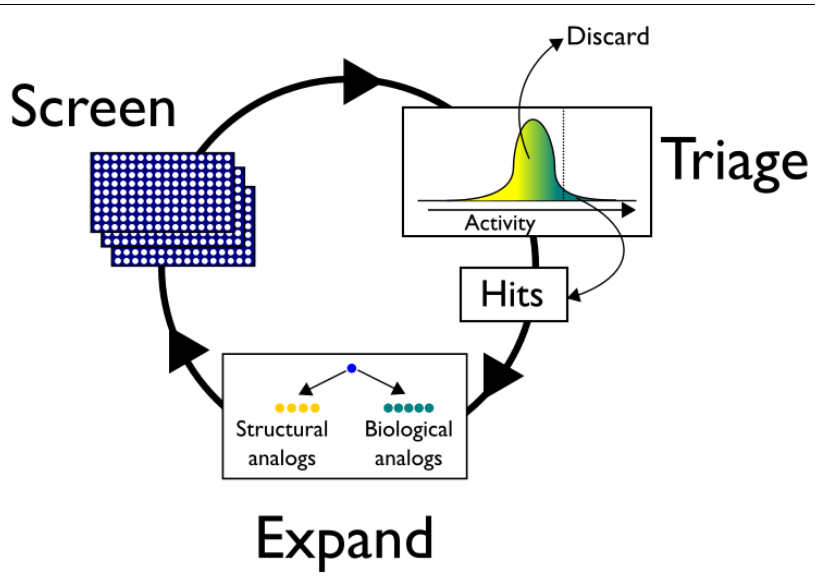

Cumulatively retrieved highly active scaffolds (%)
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
Iteration
Starting set
MoABox
random set 1
random set 2
random set 3
random set 4
random set 5
random set 6
random set 7
random set 8
random set 9
random set 10

Supplementary Figure S8. The effect of using randomly selected starting sets on median ranks and highly active scaffold retrieval. Using the MoABox as starting set resulted in better median ranks until the first iteration and virtually identical median ranks for iteration two onwards, and slightly higher scaffold retrieval throughout all iterations. While minor differences across starting sets can be observed, the key findings presented in this study are independent of the precise composition of the starting set.

Acknowledgements

S. Paricharak thanks the Netherlands Organisation for Scientific Research (NWO, grant number NWO-017.009-065), Novartis Institutes for BioMedical Research (NIBR) and the Prins Bernhard Cultuurfonds for funding and Christian Parker, Mathias Frederiksen, Gregory Landrum and Nikolas Fechner for insightful discussions.

For use in Table of Contents

Analysis of Iterative Screening with Stepwise Compound Selection Based on Novartis in-house HTS Data Shardul Paricharak, Adriaan P. IJzerman, Andreas Bender and Florian Nigsch	

References

(1) Drews, J. (2000) Drug Discovery : A Historical Perspective. *Science (80-.). 287*, 1960–1965.

(2) Macarron, R. (2006) Critical review of the role of HTS in drug discovery. *Drug Discov. Today 11*, 277–279.

(3) Mayr, L. M., and Fuerst, P. (2008) The future of high-throughput screening. *J. Biomol. Screen. 13*, 443–448.

(4) Phatak, S. S., Stephan, C. C., and Cavasotto, C. N. (2009) High-throughput and in silico screenings in drug discovery. *Expert. Opin. Drug Discov. 4*, 947–959.

(5) Mayr, L. M., and Bojanic, D. (2009) Novel trends in high-throughput screening. *Curr. Opin. Pharmacol. 9*, 443–448.

(6) Valler, M. J., and Green, D. (2000) Diversity screening versus focussed screening in drug discovery. *Drug Discov. Today 5*, 286–293.

(7) Fox, S., Farr-Jones, S., Sopchak, L., Boggs, A., Nicely, H. W., Khoury, R., and Biros, M. (2006) High-Throughput Screening: Update on Practices and Success. *J. Biomol. Screen. 11*, 864–869.

(8) Terstappen, G. C., Schlüpen, C., Raggiaschi, R., and Gaviraghi, G. (2007) Target deconvolution strategies in drug discovery. *Nat. Rev. Drug Discov. 6*, 891–903.

(9) Astashkina, A., Mann, B., and Grainger, D. W. (2012) A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. *Pharmacol. Ther. 134*, 82–106.

(10) Huggins, D. J., Venkitaraman, A. R., and Spring, D. R. (2011) Rational Methods for the Selection of Diverse Screening Compounds. *ACS Chem. Biol. 6*, 208–217.

(11) Perez, J. J. (2005) Managing molecular diversity. *Chem. Soc. Rev. 34*, 143–152.

(12) Petrone, P. M., Wassermann, A. M., Lounkine, E., Kutchukian, P., Simms, B., Jenkins, J., Selzer, P., and Glick, M. (2013) Biodiversity of small molecules - a new perspective in screening set selection. *Drug Discov. Today 18*, 674–680.

(13) Willett, P. (1999) Dissimilarity-Based Algorithms for Selecting Structurally Diverse Sets of Compounds. *J. Comput. Biol. 6*, 447–457.

(14) Koutsoukas, A., Paricharak, S., Galloway, W. R. J. D., Spring, D. R., Ijzerman, A. P., Glen, R. C., Marcus, D., and Bender, A. (2013) How Diverse Are Diversity Assessment Methods? A Comparative Analysis and Benchmarking of Molecular Descriptor Space. *J. Chem. Inf. Model. 54*, 230–242.

(15) Baell, J., and Walters, M. A. (2014) Chemical con artists foil drug discovery. *Nature 513*, 481–483.

(16) Che, J., King, F. J., Zhou, B., and Zhou, Y. (2012) Chemical and Biological Properties of Frequent Screening Hits. *J. Chem. Inf. Model. 52*, 913–926.

(17) Stanton, D. T., Morris, T. W., Roychoudhury, S., and Parker, C. N. (1999) Application of Nearest-Neighbor and Cluster Analyses in Pharmaceutical Lead Discovery. *J. Chem. Inf. Comput. Sci. 39*, 21–27.

(18) Crisman, T. J., Jenkins, J. L., Parker, C. N., Hill, W. A. G., Bender, A., Deng, Z., Nettles, J.

H., Davies, J. W., and Glick, M. (2007) “Plate cherry picking”: a novel semi-sequential screening paradigm for cheaper, faster, information-rich compound selection. *J. Biomol. Screen. 12*, 320–327.

(19) Boyle, N. M. O., Bostro, J., Sayle, R. A., and Gill, A. (2014) Using Matched Molecular Series as a Predictive Tool To Optimize Biological Activity. *J. Med. Chem. 57*, 2704–2713.

(20) Riniker, S., Wang, Y., Jenkins, J. L., and Landrum, G. A. (2014) Using Information from Historical High-Throughput Screens to Predict Active Compounds. *J. Chem. Inf. Model. 54*, 1880–1891.

(21) Wassermann, A. M., Lounkine, E., and Glick, M. (2013) Bioturbo Similarity Searching: Combining Chemical and Biological Similarity To Discover Structurally Diverse Bioactive Molecules. *J. Chem. Inf. Model. 53*, 692–703.

(22) Sams-Dodd, F. (2005) Target-based drug discovery: Is something wrong? *Drug Discov. Today 10*, 139–147.

(23) Petrone, P. M., Simms, B., Nigsch, F., Lounkine, E., Kutchukian, P., Cornett, A., Deng, Z., Davies, J. W., Jenkins, J. L., and Glick, M. (2012) Rethinking Molecular Similarity: Comparing Compounds on the Basis of Biological Activity. *ACS Chem. Biol. 7*, 1399–1409.

(24) Dančík, V., Carrel, H., Bodycombe, N. E., Seiler, K. P., Fomina-Yadlin, D., Kubicek, S. T., Hartwell, K., Shamji, A. F., Wagner, B. K., and Clemons, P. a. (2014) Connecting Small Molecules with Similar Assay Performance Profiles Leads to New Biological Hypotheses. *J. Biomol. Screen. 19*, 771–781.

(25) Keenan, S. M., Geyer, J. A., Welsh, W. J., Prigge, S. T., and Waters, N. C. (2005) Rational inhibitor design and iterative screening in the identification of selective plasmodial cyclin dependent kinase inhibitors. *Comb. Chem. High Throughput Screen. 8*, 27–38.

(26) Schneider, G., Hartenfeller, M., Reutlinger, M., Tanrikulu, Y., Proschak, E., and Schneider, P. (2009) Voyages to the (un)known: adaptive design of bioactive compounds. *Trends Biotechnol. 27*, 18–26.

(27) RDKit: cheminformatics and machine learning software (http://www.rdkit.org/); 2013.

(28) Rogers, D., and Hahn, M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model. 50*, 742–754.

(29) Bemis, G. W., and Murcko, M. A. (1996) The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem. 39*, 2887–2893.

(30) Willett, P., Barnard, J. M., and Downs, G. M. (1998) Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci. 38*, 983–996.

(31) Indigo toolkit, version 1.1.12; GGA Software Services, 2013.

(32) TIBCO Spotfire, version 4.0.4.38; TIBCO Software Inc., 2014.

(33) Dessau, R. B., and Pipper, C. B. (2008) “R”--project for statistical computing. *Ugeskr. Laeger. 170*, 328–330.

(34) Saito, R., Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., Lotia, S., Pico, A. R., Bader, G. D., and Ideker, T. (2012) A travel guide to Cytoscape plugins. *Nat. Methods 9*, 1069–1076.

(35) Paricharak, S., Klenka, T., Augustin, M., Patel, U. A., and Bender, A. (2013) Are

phylogenetic trees suitable for chemogenomics analyses of bioactivity data sets: the importance of shared active compounds and choosing a suitable data embedding method, as exemplified on Kinases. *J. Cheminform. 5*, 49–68.

(36) Hanks, S. K., and Hunter, T. (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J. 9*, 576–596.

(37) Guha, R., and Van Drie, J. H. (2008) Structure-Activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model. 48*, 646–658.