‘Robust Hit Threshold’: A Simple and Objective Hit Definition for HTS

Johannes Knop, Renee Emkey, Violeta Yu, Noemi Sealock, Seamus Ragan, Josh Xiao, and Cheng Su, Amgen Inc.*
**Correspondence should be addressed to J.K. (jknop@amgen.com)*

The role of High-Throughput Screening (HTS) is to identify as many drug discovery start points (compounds) as possible as fast as practicable. Thresholds used to define hits are selected with the aim of maximizing the number of “true actives” - compounds with intrinsic activity against the target of interest. In an ideal assay, “true actives” are expected to form a distribution separated from the bulk of inactive compounds. Real-world HTS assays however, typically display a seamless mixed distribution of the two species. This lack of separation hampers defining a meaningful threshold for selecting compounds for secondary screening. Selection of a hit threshold becomes a trade-off between “false-positives” – signals resulting from assay noise - and “false-negatives” – compounds with intrinsic activity that are missed by the chosen threshold.

Different methods for selecting hits are commonly employed in HTS labs (reviewed in [1]). Hits can be selected as a percentage (e.g. top 1%) of the compounds with the highest signal. Since the number of true actives is not expected to be constant, the sensitivity of this approach may vary widely between assays.
Alternatively, hits are often selected based on a fixed ‘activity’ threshold (e.g. >50%), assuming that the measured signal correlates with the intrinsic activity of a compound. From a statistical perspective, single-point measurements cannot be expected to give a reasonable estimate of true activity. Additionally, both approaches ignore the assay variability which may lead to high rates of false-positives or false-negatives.
A more probabilistic method is to estimate the assay error from the distribution of all valid, normalized measurements and to select compounds with measured activities deviating from the bulk of (presumably inactive) compounds. Conventionally, the hit threshold is based on Mean ± 3 Standard Deviations, which approximates the 99.87 percentile for a normal distribution. Since Mean and Standard Deviation are sensitive to ‘outliers’, the assay variability will be overestimated in the presence of active compounds.

To overcome the limitations of these ‘classical’ approaches and to make hit selection objective and consistent, we aimed to identify a method that i) applies sufficient statistical rigor, ii) is easy to understand and compute, and iii) produces hit-rates in a reasonable range.
To test the performance of hit selection criteria based on activity distributions, a meta-analysis of 41 screening campaigns from our different Lead Discovery labs was performed. The screens represented a balanced cross-section of assay types (biochemical vs. cell-based), target classes, and assay modalities (activator vs. inhibitor). All campaigns were conducted using the same compound library. Result values were normalized to ‘percent activity’ for both, activation and inhibition assays using in-plate controls and statistical parameters were calculated using all valid compound measurements from each of the screens.

We used the median and inter-quartile range (IQR) as robust estimates of the assay’s center and variability, respectively. As median and IQR are widely unaffected by ‘outliers’, they are not significantly impacted by active compounds. The IQR was normalized (NIQR = 0.741*IQR) so that Median ± 3*NIQR approximates the 99.87 percentile for a normal distribution. Since this threshold aims at identifying non-zero activities, a large fraction of the resulting hits display very low activity values which may be biologically meaningless. Especially assays with low variability produced unreasonably high hit-rates with a very low average activity.

To overcome these limitations, we introduced the concept of a ‘minimum desired response (MDR)’, a constant which shifts the threshold away from the distribution of inactive compounds while yielding biologically meaningful hits and manageable hit-rates. The optimal MDR value was determined empirically by analyzing resulting hit-rates from all 41 HTS

campaigns. Based on our experience with the compound library, hit-rates of 0.1-1.0% were deemed reasonable (Figure 1).

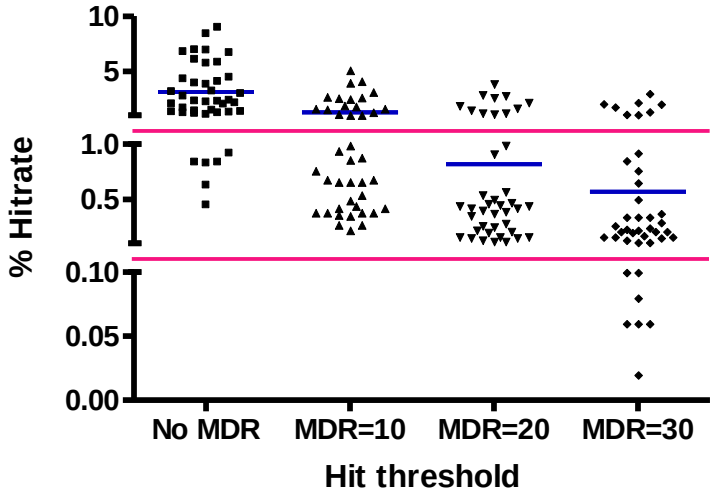

Figure 1: Resulting hit-rates from 41 screening campaigns using the hit threshold Median+3*NIQR+MDR for variable MDR values. Each dot represents one screening campaign.

Minimum Desired Response (MDR)	0	10	20	30
Screens in hit-rate range 0.1-1.0%	15%	59%	73%	66%
Average hit-rate	3.1%	1.3%	0.8%	0.6%

Table 1: Percentage of screening campaigns producing hit-rates in the range 0.1-1.0% as a function of the MDR

With 73% of the assays falling into the desired hit-rate range and an average hit-rate of 0.8%, MDR=20 turned out to be most efficient (Table 1).
A generic hit threshold definition should reflect variations in the number of 'true actives' between assays and target classes. To test this hypothesis, the correlation between the robust threshold values and resulting hit-rates were analyzed (Figure 2). Hit-rates varied widely between assays at similar threshold values. Additionally, we found that differences in hit-rates matched our experience for various assay types and target classes: Average hit-rates for biochemical screens were higher compared to cell-based screens, activation assays produced less hits than inhibition assays, and ion channel screens had, on average, higher hit rates than GPCR screens.

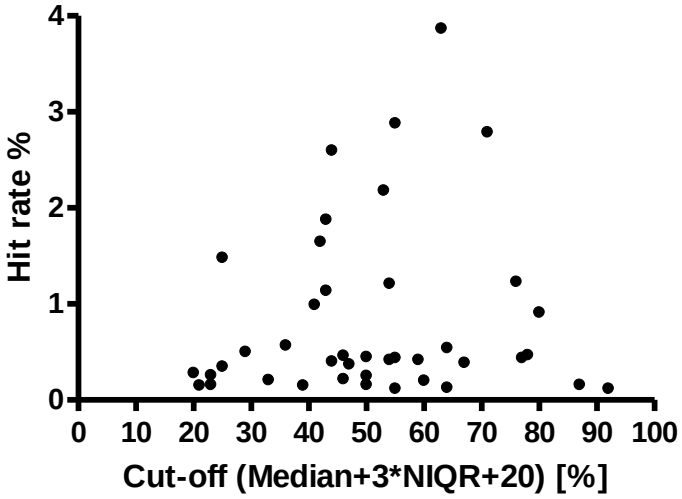

Figure 2: Correlation between 'robust hit threshold' and resulting hit-rate for 41 screening campaigns. Each dot represents one screening campaign.

The ‘robust hit threshold’ Median+3*NIQR+20 calculated from all valid, normalized measurements of a screening campaign is an objective, statistics-driven hit-threshold definition that yields manageable hit rates for the majority of our diverse screening operations. It is easy to understand and uses standard statistics that are easy to compute. Introducing the ‘robust hit threshold’ as a standard method across our Lead Discovery organization increased consistency and comparability of hit- and confirmation-rates between assays and across research groups.

---

[1] Malo, N. *et al.*, Statistical practice in high-throughput screening data analysis, Nature Biotechnology, 24, 2, 2006, p.167