Method Sheet 23

How to define and identify hits from high throughput screening data

Overview

This method sheet explains how to define and identify hits from high throughput screening (HTS) assay data. Hits are those compounds or natural extracts that are discovered in a bioassay screen to have favourable bioactivity, and which could therefore form the basis for development of a new drug from that compound or extract. There are three common approaches for the definition of hits from HTS data in academia and industry. Each has its own strengths and weaknesses, depending on the difficulty of the target and the variability of the assay, as discussed in the sections below.

1. The Arbitrary Cutoff Method

A fixed percentage threshold is chosen as the definition of a hit before looking at the data. For example, you could choose to select as hits all extracts or compounds showing >80% inhibition of growth or enzyme activity in your assays.

- *Advantages*: A very simple approach that can be completed quickly in the analysis phase.
- *Disadvantages*: This approach does not account for how variable the assay is, or how difficult the target is. Some screens may have high variability, giving rise to many false positives. Other screens may reveal that it is difficult to reach the specified level of inhibition for that particular target, so no hits are discovered at all.

2. The Ranked List Approach

In this method, all of the extracts or compounds are ranked from top to bottom by the extent of inhibition they achieve in your assay. You simply pick how many of the top performing extracts or compounds you would like to take forward as hits.

- *Advantages*: Ensures you always have samples to move forward with further testing (replication) of hits or compounds.
- *Disadvantages*: If the assay failed, or none of the compounds or extracts are capable of inhibiting the target of interest at all, you may not be finding true inhibitors.

3. The Statistical Approach

This approach uses the Standard Deviation (SD) of your negative controls to calculate a threshold percentage value, below which the samples are considered to be hits. For smaller screens (such as in a student project), this threshold may be set at 2 SDs lower than the mean of the controls (giving 95% certainty that the hits are real).

- *Advantages*: This is the most scientifically rigorous method, as it adapts to the quality of your assay. If the assay is highly variably, the threshold to qualify for a hit will be lower.
- *Disadvantages*: Can result in a very high hit rate if the target is easily inhibited, and too many extracts or compounds to take forward for replication.

Which method should you use?

In professional drug discovery, the statistical approach using a 3 SD difference from the control is preferred. This approach gives high confidence (99.7% certainty) that the hits are real and not just artefacts arising from random variability in the assay.

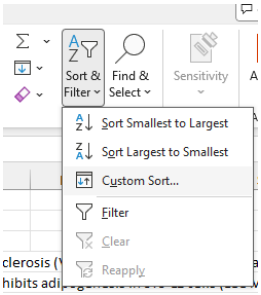
As your screen is relatively small compared to the many thousands or millions of compounds screened in industrial HTS projects, we recommend you use the Ranked List method for defining which hits to take forward for further investigation.

Method for ranked list hit selection

1) This method should be performed after you have completed the data normalisation and ID alignment tasks for your assay data.
2) Your data should now be in a list of the mean (of three or four experiments) percentage activity for every compound or extract ID, with the list sorted by extract ID in the first column.
3) Create a new worksheet (tab) in the Excel analysis file and call it “Hit finding”.
4) Highlight all of the columns in the data table from the previous worksheet.
5) Make sure you include the “Mean” column, and also the compound or extract descriptor columns (such as plant name, traditional uses etc.) when you select this.
6) Paste this table as VALUES into the new “Hits” worksheet.
7) Select this entire table in the “Hits” worksheet, including the headings at the top of the columns.
8) It should look something like this at this stage (note the ascending extract ID):

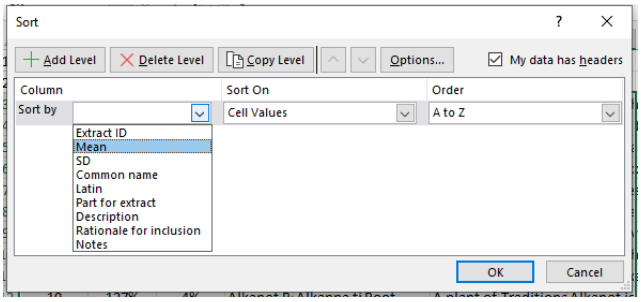
	A	B	C	D	E	F	G	H	I	J
1										
2	Extract ID	Mean	SD	Common	Latin	Part for e:	Descriptic	Rationale	Notes	
3	1	99%	3%	Acacia Gu	Acacia ser	Resin	A small, th	Traditiona	Contains hentriacontar	
4	2	103%	4%	Yarrow	Achillea m	Leaves	A small flo	Traditiona	Reported to reduce dis	
5	3	120%	4%	Calamus R	Acorus cal	Root	A tall pere	Used in Ch	Contains alpha-asaron	
6	4	123%	4%	Kiwi fruit	Actinidia c	Fruit	The kiwifr	Dietary in	Ethanol extract of peel	
7	5	129%	5%	Horse Che	Aesculus h	Seed	A large de	A Cochran	Horse chestnut seed is	
8	6	106%	4%	Button mu	Agaricus b	Fungal fru	An edible	Dietary in	This is the most comm	
9	7	88%	3%	Agrimony	Agrimonia	Leaf	A deciduo	Traditiona	Contains volatile oils, f	
10	8	90%	3%	Couch Gra	Agropyror	Rhizome	A commor	Traditiona	A major chemical const	
11	9	84%	3%	Lady's Ma	Alchemill	Leaf	A herbace	Traditiona	Vasorelaxant propertie	
12	10	127%	4%	Alkanet R	Alkanna ti	Root	A plant of	Traditiona	Alkanet is also used in	
13	11	123%	4%	Onion	Allium ce	Bulb	A cultivate	Dietary in	Rich in phenolic compo	
14	12	115%	4%	Nigella se	Nigella sa	Seed	An annual	Traditiona	Despite the common n	
15	13	125%	4%	Leek	Allium po	Stem	A cultivate	Dietary in	A. porrum was reporte	
16	14	118%	4%	Garlic	Allium sat	Bulb	A cultivate	Dietary in	Antibacterial and chole	
17	15	125%	4%	Chives	Allium sch	Leaf	A cultivate	Dietary in	Traditionally used to re	
18	16	103%	4%	Lemon Ve	Aloysia cit	Leaf	A flowerir	Traditiona	Contains verbascoside,	

9) In the Editing section of the Home ribbon, click on the A-Z Sort and Filter button, and select the ‘Custom Sort’ option.



10) In the dialogue box that appears, use the pulldown menu in the ‘Sort by’ section and select ‘Mean’, leave the other pulldown menus as ‘Cell Values’ and ‘Smallest to Largest’

11) Click ‘OK’ to begin the sorting function.



12) Your data should now be sorted with the lowest values for percentage response at the top of the list and the highest values at the bottom of the list.

13) Your hits are the extracts or compounds at the top of this list.

14) Simply choose as many hits as you want from the uppermost rows in this list.

15) For laboratory-based *Phytotitre* projects we recommend taking forward at least 1 hit, preferably the top 3 hits, for further analysis in the next stages of your project.

Notes

- The approach given above works for the antibiotic discovery, cancer cell drug discovery and enzyme inhibitor discovery *Phytotitre* projects.
- If you are seeking hits for the TNFα production inhibitor discovery project, the hits will be those that do NOT kill the cells, and give wells that are darker in colour than the DMSO control, since less production of TNFα results in more cell survival.
- Therefore, for the TNFα production inhibitor project only, you should sort the data in descending order (not ascending order as shown above), and in this case your hits will be those with the highest percentage absorbance values (not the lowest).