What are Perturbagen Classes (PCLs)?

A **P**erturbagen **CL**ass (PCL) refers to a CMap-designated grouping of compounds or genetic perturbagens. Compound PCLs are identified by first grouping compounds that share the same mechanism of action (MoA) or biological functions as determined from the literature. Then L1000 data is made (or existing data analyzed) for these groupings to assess whether the members give similar gene expression signatures, thus confirming their shared activity. Below are examples from CMap data:

In cases where there is agreement from multiple (usually 3 or more) members of the class, but not all (i.e some literature nominated perturbagens produce signatures inconsistent with the other members), the classes are pruned to a core set of strongly interconnected perturbagens.

PCLs can also be created using genetic perturbagens. A genetic PCL contains strongly interconnected genetic perturbagens corresponding to genes that belong to the same gene family or which are commonly targeted by the same compounds.

Using PCLs to interpret CMap results

A strong connectivity to a PCL, as opposed to a single perturbagen, provides users a higher level of confidence in predicticting biological functions from CMap analyses.

A typical question the CMap compendium can be used to answer is, "What is the MoA/target for unknown compound X?". One can query CMap with a signature of compound X and examine its list of Touchstone connections, but this list can be long and may not always yield immediate insights. PCLs help by 'reducing the dimensionality' of connections and showing connectivity at the MoA level as opposed to individual perturbagens.

Caveats in using PCLs

- Gene expression might not be the optimal readout for all MoAs. That is, an activity may not elicit a transcriptional response and hence the a coherent class in L1000 data isn't in itself definitive.
- PCLs are inherently limited to the data we've made. We may not have a PCL for a certain class simply because we haven't yet profiled enough of its members to observe a coherent response.
- Annotations and literature can be ambiguous about the MoA for some compounds. For example, some compounds have multiple activities and different sources report the different activities. Using data can help refine these discrepancies by examining which activities are actually observed.
- Importantly, while already notably larger than earlier datasets, the current L1000 data almost certainly doesn't have the right cellular context for several biological contexts. For example, many neuroactive libraries show limited response in the current set of 9 cancer cell lines.

In summary: PCLs (perturbagen classes) are high-level sensors of certain activities. Members of a compound PCL share the same mechanism of action and have been shown empirically to connect with their class members in L1000 data. Similarly, members of a genetic PCL belong to

the same gene family or are commonly targeted by the same compound, and show strong interconnectivity in CMap. Strong connectivity to a PCL, as opposed to an individual perturbagen, offers a reductive but more interpretable view of connections because it represents connectivity to a group of related perturbagens rather than just a single perturbagen.

Last modified: Mon Aug 06 2018 12:21:19 GMT-0400 (EDT)