Chemical Activity Profiles for Obesity

Background

There are thousands of chemicals currently used, and about 2,000 new chemicals are introduced per year. In order to maintain chemical inventory and regulate hazardous chemicals, the US Environmental Protection Agency (EPA) operates rapid chemical screening programs called Toxicity Forecaster (ToxCast) and Toxicology in the 21st Century (Tox21) using high-throughput screening (HTS) technology. The primary aim of these programs is to differentiate active vs inactive chemicals for different biological endpoints. However, even with the HTS technology, it is impossible to test all possible pairs of chemicals and assay endpoints at a fine grid of doses, resulting in many missing pairs and doses.

化驗

Data Information

The 'ToxCast.RDS' has

i) 'obese_data': a subset of ToxCast/Tox21 data for 30 chemicals and 271 assay endpoints known to be relevant to obesity.
   - 'aeid': assay endpoint ID
   - 'aenm': assay endpoint name
   - 'casn': CAS registry number (unique identifier for chemicals)
   - 'code': casn in different format
   - 'chnm': chemical name
   - 'logc': dose in log base 10
   - 'resp': response value

ii) 'hitc_data': hit-call value based on invitroDBv2 of EPA, which is unique to each (chemical, assay endpoint) pair
   - 'code': casn in different format
   - 'aenm': assay endpoint name
   - 'hitc': hit-call value where 1 if active, 0 if inactive, -1 if cannot be determined, and NA if not tested.

iii) 'aenm_data': an organism and a tissue that each of the 271 assay endpoints is extracted from
   - 'organism': e.g. human, pig, etc.
   - 'tissue': e.g. breast, liver, etc.
   - 'aenm': assay endpoint name

Questions and Goals

Which chemicals are active in relation to obesity?

- Identify chemical activity profiles across the assay endpoints.
- Predict activity for non-tested pairs, if possible.
- Compare your results with EPA's hit-call and justify any differences.
- Explain patterns of activity by chemicals, assay endpoints or their organisms and tissues, if there is any.
- Your methods should account for sparsity of the data.