Debilitats ORA

* Les estadístiques utilitzades pel mètode ORA, com ara la distribució hipergeomètrica, són independent dels canvis mesurats. Això vol dir que aquests tests ignoren tots els valors associats amb ells com ara les intensitats (logRatios)
* Típicament ORA utilitza només els gens més significatius i descarta els altres.
* ORA tracta cada gen igual i per tant assumeix que tots els gens són independent un de l’altre.
* L’ORA assumeix que totes les rutes són independents una de l’altra

Debilitats GSEA

* Tan com l’anàlisi ORA, també GSEA assumeix que les rutes són independents;
* GSEA utilitza els canvis d’expressió gènica dins de la ruta específica i descarta els canvis d’altres anàlisis. Per exemple, agafem dos gens A i B, que canvien dos vegades un (2-fold) i 20 vegades l’altre (20-fold) respectivament. Fins que aquests gens tenen el mateix ranc comparant amb altres gens de la ruta, GSEA els tractarà igual, encara que seria més preferible donar més pes al gen B.

Mapes d’enriquement.

L’anàlisi ORA resulta en una llista de les rutes significativament enriquides. Els gens dins dels conjunts o les rutes poden encavalcar i descriuen gairebé els conceptes biològics idèntics \cite{merico2010enrichment}. Aquest problema de redundància és més evident als conceptes que són organitzats jeràrquicament, com ho fan conceptes de la base da dades GO. El mapa d’enriquiment redueix la redundància als conjunts de gens. Els conjunt de gens estan representats com a nodes els radis dels quals són proporcionalment relacionats amb el nombre de gens que formen part d’aquests conjunts. Els cantells indiquen les nodes que tenen gens compartits, on el seu gruix depèn del nombre de gens compartits. A més a més es pot utilitzar el color de les nodes per representar una altra dimensió com ara el nivell de significació expressat per el valor de P. Si no hi ha cap gen compartit entre els conceptes (o rutes) les nodes no són connectades via cantells. Aquest mètode de representació és molt útil per poder reduir/simplificar la informació obtinguda mitjançant els mètodes ORA o GSEA i per tant m’he posat l’objectiu també implementar-lo a l’aplicació.

Instal·lacio

He decidit no forçar la instal·lació de totes les bases d’anotacions per GO i Reactome. Al servidor sí que ho faria però per a instal·lació local podria resultar en una experiència desagradable, perquè encara que l’usuari necessités per la seva anàlisi només un genoma anotat específic, hauria d’instal·lar tots altes innecessàriament, i per tant aguantar el temps d’instal·lació que iniciaria amb la primera crida de la funció \helvetica{runPathwayApp} i també l’espai al seu disc dur. Recordem que cada base d’anotació té un pes important. La base de dades org.Mm.eg.db per a ratolí, per exemple, ocuparà aprox. 275 megabytes al disc dur. Si anteriorment les anotacions no són descarades per a espècia que l’usuari vol analitzar, l’usuari rebrà un error: \textbf{ERROR: object 'org.Mm.eg.db’ per not found”.

Exemple d’estudi. Enrichment Map.

Observem que totes les rutes identificades comparteixen els gens.

Exemple d’estudi. Category-Gene network.

Aquí identifiquem els gens que pertanyen a cada de les rutes, com per exemple Elovl3, el component important per a reclutament dels lípids al teixit adipós marró \cite{westerberg2006elovl3}.