SUPLEMENTARY INFORMATION (I)

M. musculus analysis

Expression Data

The dataset used for this example comes from a *Mus musculus* experiment performed in an Affymetrix microarray platform (Mouse Genome 430 2.0). It is available both at GEO (accession number GSE18115) and ArrayExpress (accession number E-GEOD-18115).
The experiment explores dendritic cell (DC) stimulation to lypopolisaccharide (LPS, a complex commonly present in pathogens) and to polyriboinosinic:polyribocytidylic acid (pI:C, a dsRNA mimic with strong immunostimulatory capacity).

The experiment has the following resulting conditions:

- 0h – no LPS
- 4 and 16h, with LPS

4 and 16h, pI:C (soluble)

The expression data (processed by the authors) has a mean expression value of 165, with a minimum of 0 and a maximum of 10639. This asymmetry is common among expression data and could be interesting to normalize them, but we will use the data as provided, instead of re-processing the data.

The corresponding publication[1], according to the abstract, states that *"in response to polyriboinosinic:polyribocytidylic acid (pI:C), DCs mount a specific integrated stress response during which the transcription factor ATF4 and the growth arrest and DNA damage-inducible protein 34 (GADD34/Ppp1r15a), a phosphatase 1 (PP1) cofactor, are expressed. In agreement with increased GADD34 levels, an extensive dephosphorylation of the translation initiation factor eIF2α was observed during DC activation*".

Therefore, some genes of interest, and annotated terms, are as follows:

Name	Entrez ID	Related terms
GADD34/Ppp1r15a	17872	Cell death, response to stress, regulation of biological process.
eIF2a	229317	Translation initiation factor, regulation of translation, ribosome assembly.
ATF4	11911	Regulation of transcription from RNA polymerase, DNA binding
CHOP	13198	Response to different stimuli, cell cycle, regulation of biological process

[1] Clavarino G, Cláudio N, Dalet A, Terawaki S, Couderc T, Chasson L, Ceppi M, Schmidt EK, Wenger T, Lecuit M, Gatti E, Pierre P. *Protein phosphatase 1 subunit Ppp1r15a/GADD34 regulates cytokine production in polyinosinic:polycytidylic acid-stimulated dendritic cells.* PNAS 2012;109(8):3006-11.

IFN-beta	15977	Immune response, response to different stimuli, regulation of biological process
Interleukin-6	16193	Cytokine activity, defense reponse, negative regulation of apoptosis

Possibly relevant biological terms, according to the paper:

- Specific DC stress response
- RNA transcription
- Translation initiation
- DNA-damage
- Growth arrest
- Immune response
- Type-1 Interferon

Considerations

We expect to find relevant patterns on some of the related biological terms. Hopefully, we will find also find the genes of interest annotated on some terms, although it might happen that they won't be annotated for a determinate ontology.

We will focus on GO, a very complete and well-established ontology, and Reactome, a trusted ontology by the immunologists involved in the analysis. It must be noted that Reactome has no standard ontology, but it is hierarchically organized on pathways that contain genes, and therefore can be considered as a ontology for the purpose of Voronto.

Gene Ontology analysis

Following, we will show how the tool can be used, rather than for a general exploratory top-down analysis (see Supp. Information II), for a guided exploratory analysis that focuses on relevant terms.

The protocol is as follows:

1. Locate relevant terms, either by static labels ('l' key), by hovering non-labelled (because of size) terms, or by searching them ('f' key).
2. Once a relevant term is found, in the case that the ontology is very large (e.g. GO), we can filter out every other term, by hovering to the term and pressing the 'enter' key. This will generate a new tessellation that will occupy the whole visualization with the term's children.
3. If the visualization is not satisfactory, we can repeat 1 and 2 to go deeper in the hierarchy, or go back ('del' key) to explore other terms. If it is satisfactory we can proceed to analyse the term (step 4).
4. We can analyse the term on several ways:
    - By hovering a term, we can see its whole profile for every experimental condition, at the bottom.
    - We can use 'left' and 'right' arrow keys to see the whole aspect of the hierarchy level under a single experimental condition
    - We can alt-click on a hovered term to show a heatmap with its annotated genes, hierarchically clustered. The genes in the heatmap can be double-clicked to open its Entrez Gene entry.

- We can double-click on a hovered term to show its entry on the original resource (KEGG, GO or Reactome webpages)

With this simple interaction routine, we found some interesting aspects that are discussed below.

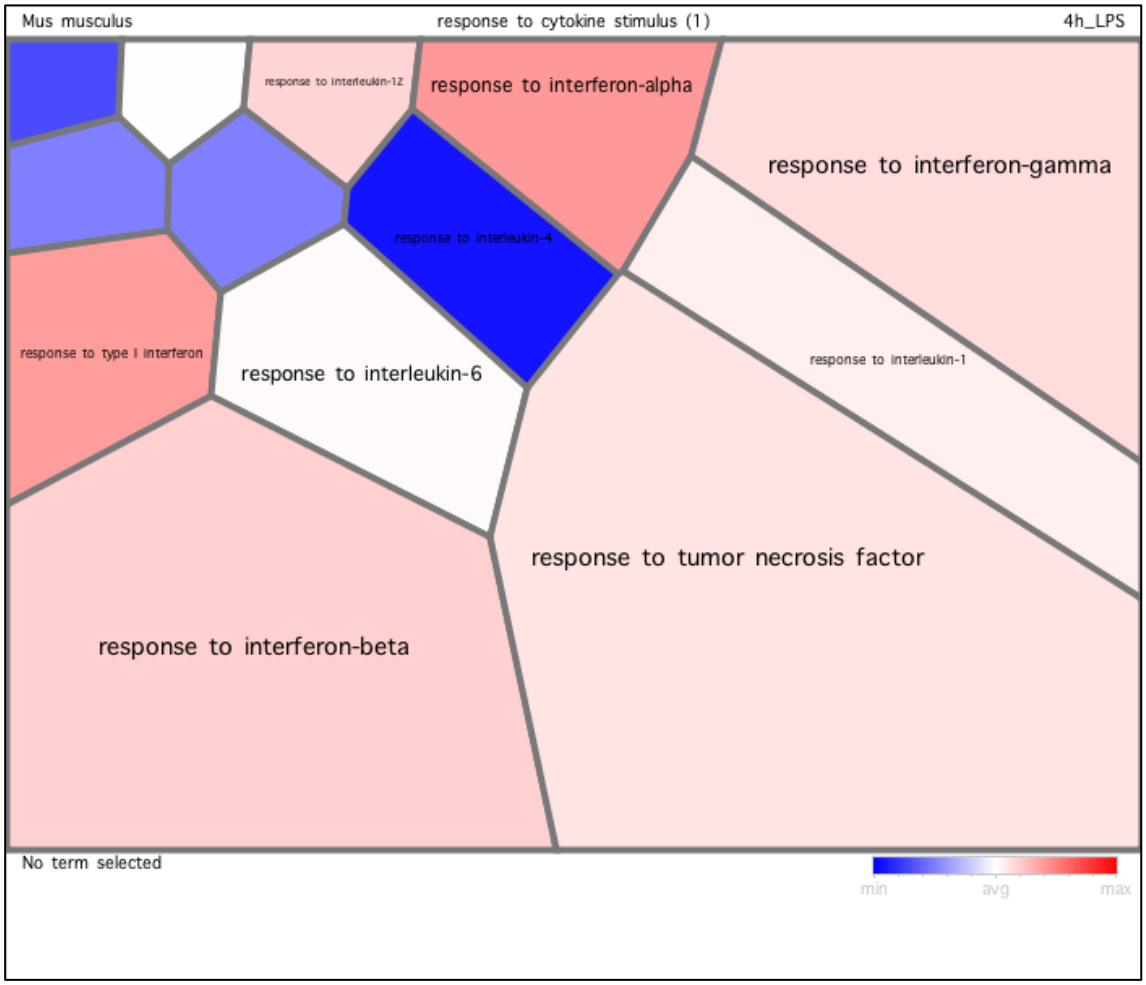
General view

On the general view, top terms do not show very relevant changes. This is something also partially observed in the Supp. Information II, and it is attributable to the large amount of genes annotated on the top terms, whose expressions counteract to give mostly average expression.

Response to cytokine stimulus

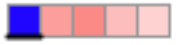
We use the 'enter' key to filter out terms until the whole visualization represents 'response to cytokine stimulus' (whole path is 'response to stimulus' -> 'response to chemical stimulus' -> 'response to organic substance' -> 'response to cytokine stimulus').

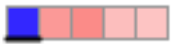
First of all, we can check that the terms are much more deviated from the average expression than in higher levels (brighter expression colours).

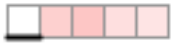


Voronoi diagram for 'response to cytokine stimulus'

Hovering over the terms, most of them show changes because of LPS, possibly the most clear ones are 'response to interferon-alpha' and 'response to type I interferon':


response to type I interferon

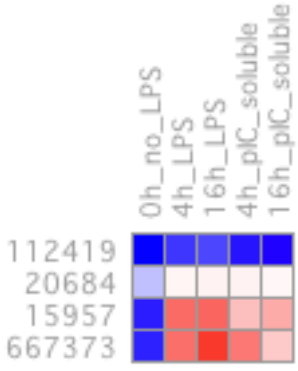

response to interferon-alpha


response to inferferon-beta

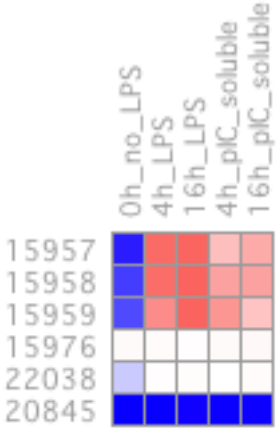
0 4 16 4 16 (time)
no|LPS | pI:C (stimulus)

Type I interferon and interpheron-alpha responses are under-regulated on a normal situation (first square, on blue) but overexpressed with LPS. With the immune-activator pI:C (squares 4 and 5) the expression is also high but milder. Thus although the type-I interferon production and autocrine responses are expected from the biology of stimulated DCs *in vitro*, Voronto analysis reveals differences in the intensity and the kinetics of the response to the different microbial stimuli.

If we get to the actual genes by alt-clicking on their corresponding terms:



response to type I interferon

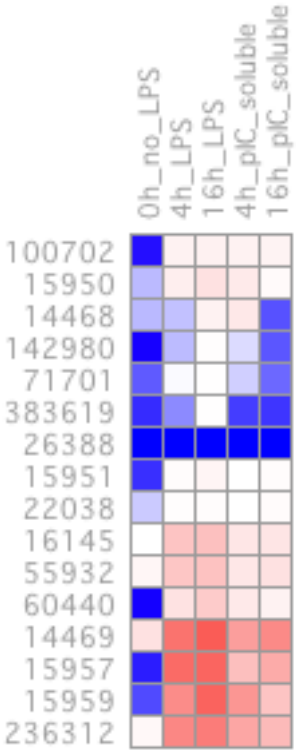


response to interferon-alpha

Type I response is due to two out of four genes, one of which (15957 – Ifit1) is also one of the three main genes responsible for the interferon-alpha response identification.

Note that some genes are never expressed, this a common pattern in several observed terms: only some genes in the term are responsible of its over-expression. Note also the less specific response of 20684 (Plscr1) and 22038 (Sp100), which are under-expressed when there is no stimulation and have average expression when it is present, independently of its type.

'Response to interferon-beta' has a similar, but milder pattern. It is not induced, but neither under-expressed when no stimuli is present, and up-regulates with stimuli (more with LPS and pI:C). There are 15 genes annotated with this term, and most of them are behaving differently with stimulation, and some of them show changes with time too:

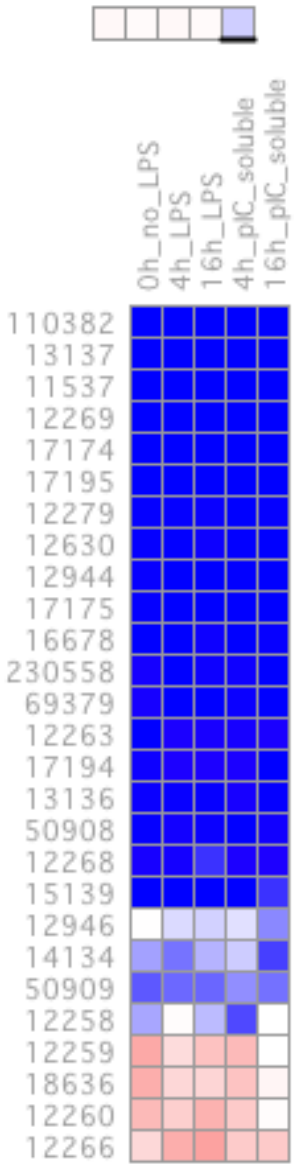


Heatmap of 'response to interferon-beta'

Note that, again Ifit1 is annotated with this third cytokine-related term.

Immune system process

One of the subterms of the 'Immune system process' is 'Complement activation', with 27 annotated genes. This term has an interesting pattern showing under-expression only for 16h pI:C:



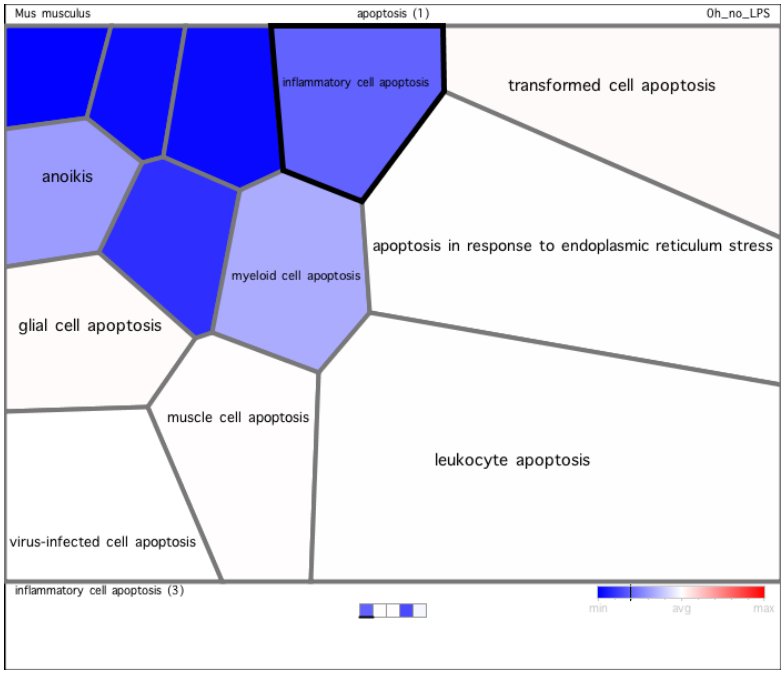
'Complement activation' (under 'Immune system process'→'Immune effector process') profile and heatmap for its annotated genes.

Most of the genes in 'Complement activation' are not expressed at all in DCs, but the combination of the lower expression at 16h+LPS or +pI:C (sol.) of the generally over-expressed genes (12259, 12260, 18636) and the specific under-expression of mildly inactive genes (especially 14134 and 12946) produce this expression pattern[2].

[2] Note that, because the data are not centered, the average expression (white color) is much closer to the minimum expression (blue) than to the maximum (red), so mild reds counteracts bright blues easily.

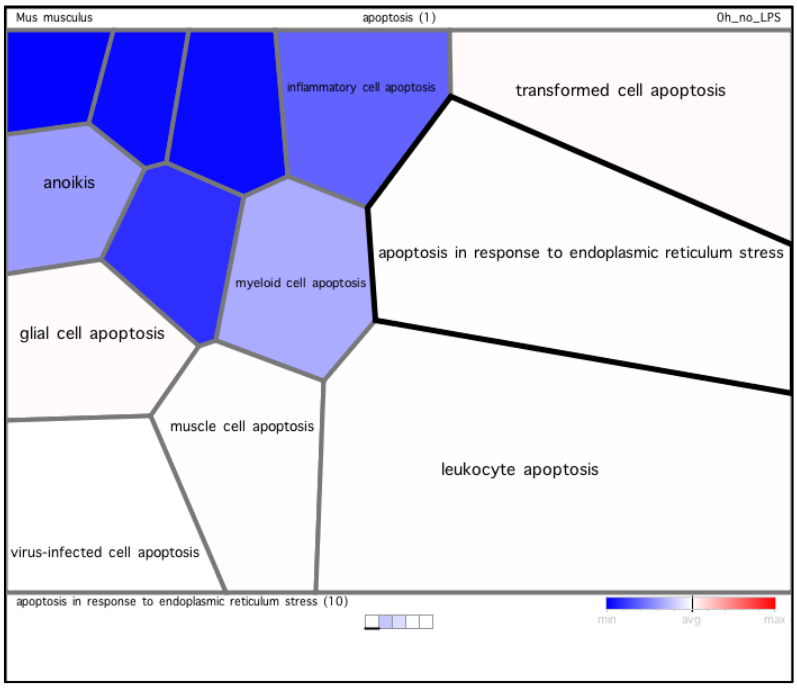
Cell death

GADD34 (17872) is annotated with 'cell death', among other terms. Exploring 'apoptosis' subterms, we find that the term 'inflammatory cell apoptosis' goes from under-expression to average expression for LPS. In the case of pI:C, it keeps under-expressed at pI:C, (soluble) initial point (4h) and goes to average later (16h).



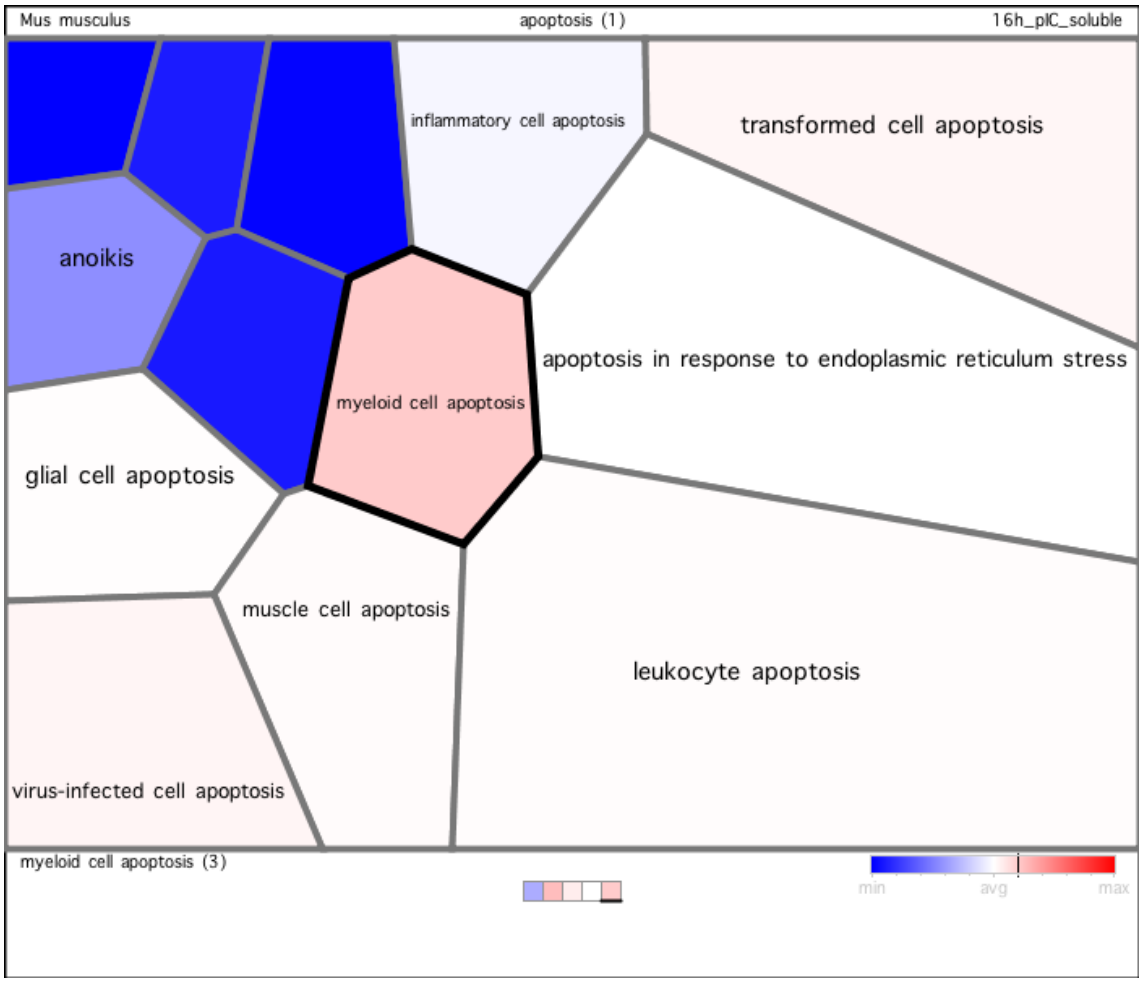
'apoptosis' subterms at 0h. No apoptosis is active, and several ones are under-expressed. 'inflammatory cell apoptosis' highlighted

The 'inflammatory cell apoptosis' pattern is almost opposite to the pattern of 'apoptosis in response to endoplasmic reticulum stress': it is average at 0h, under-expressed with LPS and (mostly) average with pI:C



'apoptosis in response to endoplasmic reticulum stress'

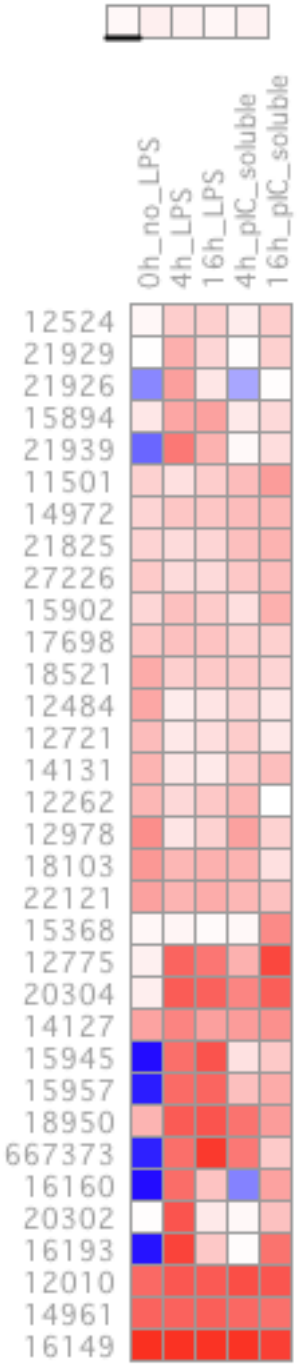
Apoptotic overexpression when LPS is present only appears in 'transformed cell apoptosis' (very mild) and 'myeloid cell apoptosis':



'Apoptosis' subterms at 16h (iP:C). Most of the terms have expression similar to 0h (compare with previous figures). The only one that clearly up-regulates is 'myeloid cell apoptosis'.

Regulation of immune system process

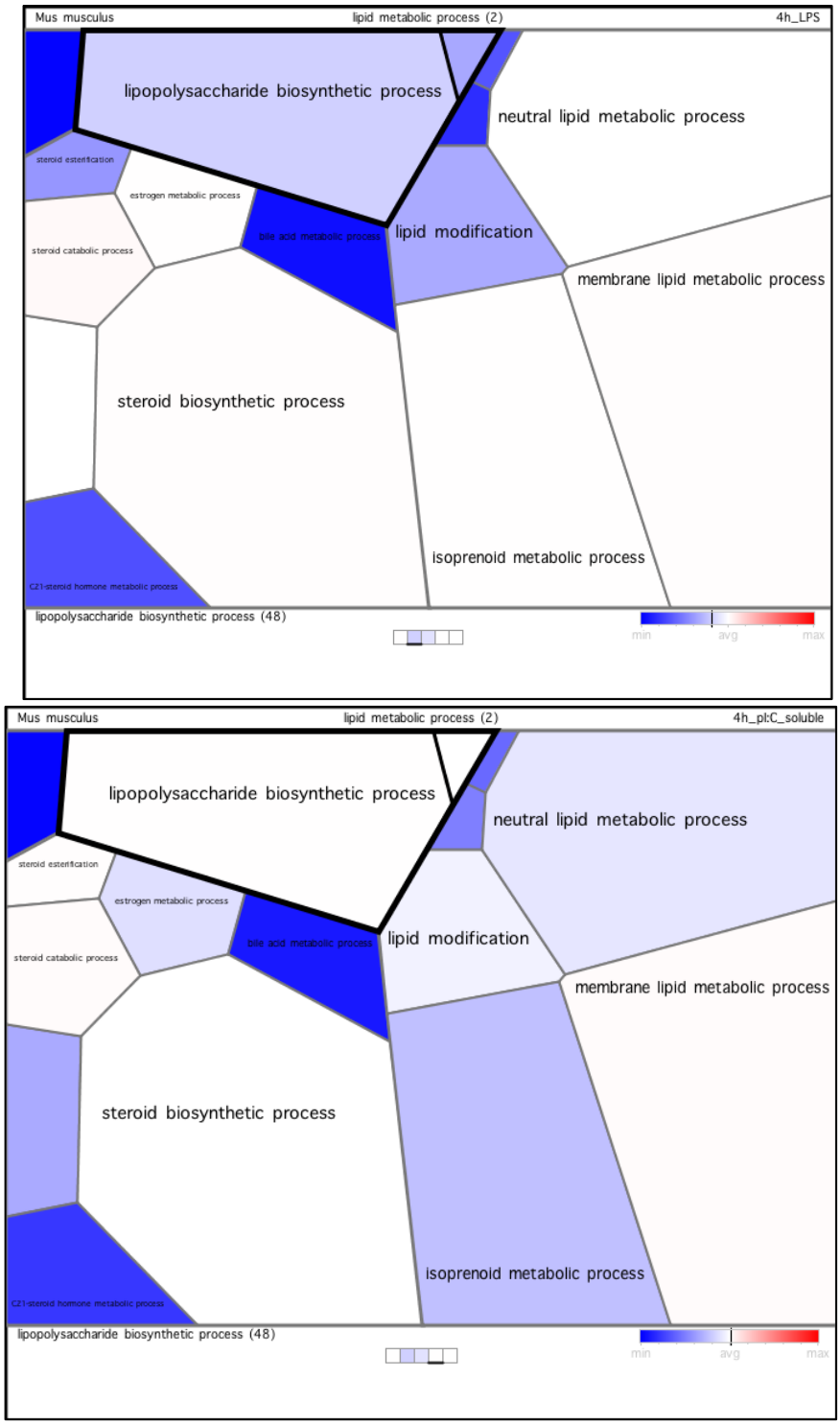
This 'biological regulation' → 'regulation of biological process' is very slightly overexpressed in the presence of LPS or pI:C. It includes over 500 genes, but only a few genes change from very low to very high expression, including IFN-beta (15977) and IL6 (16193), two of the genes of interest according to the original paper:



Detail of the 'regulation of immune system' heatmap, showing the 6 most differentially expressed genes between no stimulation and stimulation conditions.

Lipoprotein metabolic process

Since LPS is a lipid, we decided to check 'lipid metabolic process' subterms. Although mostly under average expression, there are some clear patterns differentiating LPS from pI:C conditions. Following figures show expression levels at 4h, for LPS and pI:C.



'lipid metabolic process' subterms at 4h when LPS is present (top) and when pI:C is present (bottom)

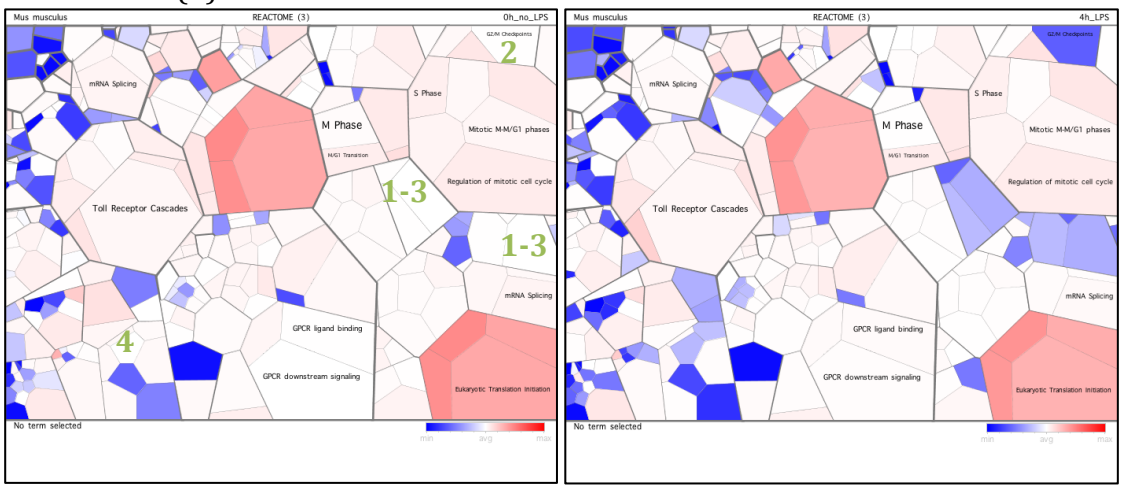
'Neutral lipid metabolic process' and 'isoprenoid metabolic process' are underexpressed for pI:C but not for LPS. Inversely, 'lipopolysaccharide biosynthetic process' and 'neutral lipid metabolic process' are underexpressed for LPS but not for pI:C. This sound especially strange, cause the basic assumption is that LPS byosinthesis shoud be more active if LPS is present.

Reactome analysis

Reactome has fewer terms and annotations that GO, but is possibly a more curated ontology. In this case we proceed with a more open analysis, by browsing experimental conditions in the first place (overview) and then focus on some terms that, because of the overview (RNA polymerase transcription) or because of the biological problem (immune systems) could be interesting.

Overview

A first glance of the ontology shows that several large terms are underexpressed on LPS presence. Possibly relevant ones are 'RNA polymerase I transcription' (1), 'G2/M Checkpoints' (2), 'RNA polymerase III transcription' (3) and 'Sphingolipid metabolism' (4):



Voronoi visualization of Reactome, up to the 3rd level, at 0h (left) and 4h+LPS (right). Note that the elements labelled (1,2,3 and 4) on the left are under-expressed on the right.

RNA polymerase I and III transcription

Both transcriptions show a similar pattern on LPS conditions:

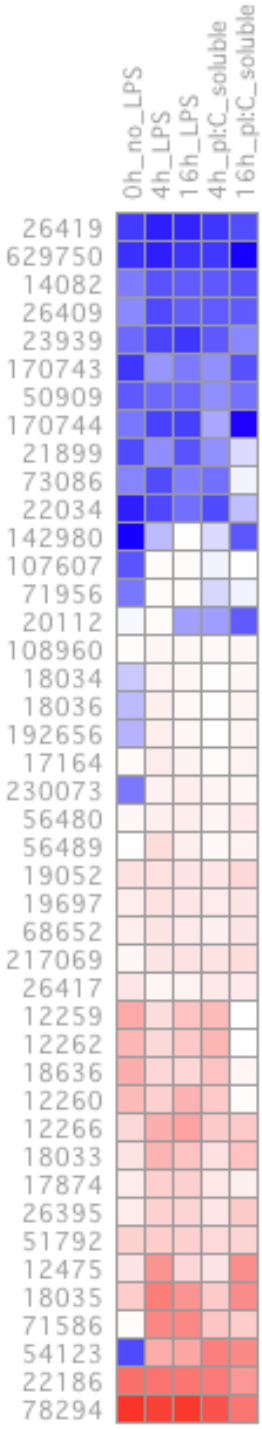

RNA polymerase I


RNA polymerase III

Both patterns are more under-expressed on LPS than on pI:C. However, RNA polymerase I expression goes to average expression rather than to under-expression for pI:C at 4h.

Immune system

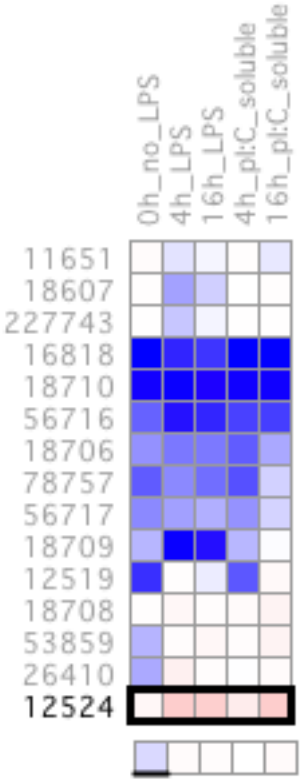
'Immune system', as a whole has an (slightly up) average constant expression. However, if we inspect the individual genes, we can see that there are specific responses of some of them to LPS and pI:C, such as 54123 (Irf7) for 'Innate immune system':



'Innate immune system' heatmap (detail). Note the expression of 54123 near the bottom.

Searching for subterms, Irf7 is present at 'TLR4 signalling' and 'TRIF mediated TLR3 signalling'.

Browsing 'adaptative immune system' we find a 15 gene-term called 'CD28 dependent PI3K/Akt signaling, which goes from under-expression to average expression with LPS, mainly due to the effect of CD86 (12524).



'CD28 dependent PI3K/Akt signaling' heatmap with the gene 12524 highlighted, and average profile (bottom)

Conclusions

This small analysis of the GSE-18115 experiment in the context of GO and Reactome ontologies gives an overall idea of how Voronto can be used to inspect and find relevant patterns in a real case.

In this case, with a few clicks and mouse moves we can check the expression of a large number of ontology terms, finding relevant patterns. It is not our scope to give a deep biological interpretation, but to show the tool usage. However, from this point of view, we detect a strong response on some genes to the presence of LPS, which is usually diluted in the context of large terms, although sometimes a strong response or a combination of responses makes the term pattern to change. Comparing the responses to LPS with pI:C reveals that although many of the induced terms are common and, as expected, modulates the immune system response, several terms such as lipid metabolism are expressed differently.

Voronto analysis has, of course, its limitations. First one is the confidence of the analyst in the ontology. If the analyst suspects that the ontology is not well formed, curated, or does not contain every annotated gene, the whole analysis might become irrelevant or at least inconclusive. Voronto offers the opportunity to load user-defined ontologies if this is the case. Secondly, Voronto is not a statistical tool, so references are always visual and based on expression levels, and no statistical

tests on differential expression or enrichment are performed. However, Voronto complements and enhances statistical tests, whose outcomes usually show just p-value lists, disregarding (or putting in second place) the number of genes present in the term, which is very important for the statistical interpretation of p-values, and the actual expression levels and hierarchy relations, which are crucial for biological interpretations.