

equations

Brian M. Schilder

2024-03-06

Tutorials

- <https://lplembleton.rbind.io/posts/annotate-equations/>
- <https://github.com/st--/annotate-equations/blob/main/annotate-equations.pdf>

Install latex

Only need to do this once. Instructions using `homebrew`.

```
brew install texlive
```

Methods

Human Phenotype Ontology

Let us denote:

- D as the set of d diseases.
- p as a phenotype.
- g as a gene.

The final evidence-weighted gene x phenotype matrix $M_{g,p}$ can be expressed as:

$$M_{g,p} = \frac{\sum_{d \in D} R(g, p, d) \times E(g, d)}{\sum_{d \in D} R(g, p, d)}$$

$$M_{g,p} = \frac{\sum_{d \in D} R(g, p, d) \times E(g, d)}{\sum_{d \in D} R(g, p, d)}$$

Weighted gene-by-disease evidence score matrix

Weighted gene-by-phenotype evidence score matrix

Average across all diseases

Binary gene-by-phenotype relationship matrix, (1=relationship, 0=no relationship)

Single-cell transcriptomic atlases

Let us denote:

- g as a gene.
- c as a cell type.
- i as a single cell.

The gene expression specificity matrix construction can be mathematically described as the following:

$$S_{g,c} = \frac{\sum_{i=1}^{|L|} F(g, i, c) / N_c}{\sum_{r=1}^k \left(\sum_{i=1}^{|L|} F(g, i, r) / N_r \right)}$$

Average gene expression by cell type

Gene-by-cell type specificity matrix

Number of cell types

Row sums

Gene-by-cell expression matrix

Number of rows

Genes with very no expression across any cell types were considered to be uninformative and were therefore removed before computing the specificity matrix.

$$F(g, i, c) = \begin{cases} r_{g,i}, & l_i = c \\ 0, & l_i \neq c \end{cases}$$

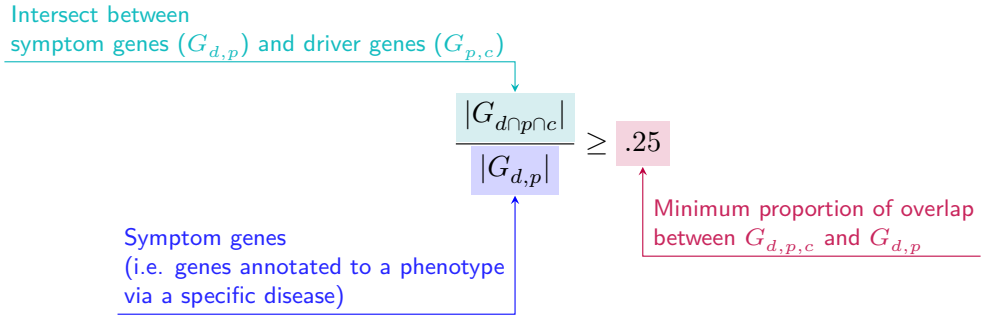
Expression of gene g in cell i

Filtered gene-by-cell type specificity matrix

$$F(g, i, c) = \begin{cases} r_{g,i}, & l_i = c \\ 0, & l_i \neq c \end{cases}$$

Symptom-cell type associations

$$\frac{|G_{d \cap p \cap c}|}{|G_{d,p}|} \geq 0.25$$



Annotation of phenotypes using generative large language models

$$NSS_p = \frac{\sum_{j=1}^m (F_{pj} \times W_j)}{\sum_{j=1}^m (\max\{F_j\} \times W_j)} \times 100$$

Sum of weighted annotation values across all metrics

Numerically encoded annotation value of metric j for phenotype p

Weight for metric j

Normalised Severity Score for each phenotype

$$NSS_p = \frac{\sum_{j=1}^m (F_{pj} \times W_j)}{\sum_{j=1}^m (\max\{F_j\} \times W_j)} \times 100$$

Theoretical maximum severity score

Session Info

```
utils::sessionInfo()
```

R version 4.3.1 (2023-06-16)

Platform: aarch64-apple-darwin20 (64-bit)

Running under: macOS Sonoma 14.3.1

Matrix products: default

BLAS: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib

LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;

locale:

[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: Europe/London

tzcode source: internal

attached base packages:

[1] stats graphics grDevices utils datasets methods base

loaded via a namespace (and not attached):

[1] compiler_4.3.1	fastmap_1.1.1	cli_3.6.2	tools_4.3.1
[5] htmltools_0.5.7	rstudioapi_0.15.0	yaml_2.3.8	rmarkdown_2.25
[9] knitr_1.45	jsonlite_1.8.8	xfun_0.42	digest_0.6.34
[13] rlang_1.1.3	evaluate_0.23		