equations

Brian M. Schilder

2024-07-22

Tutorials

- https://lpembleton.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
```

or using quarto:

```
quarto install tinytex
```

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:

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

```
{=tex #eq-evidence-scores} \begin{equation*}   \eqnmarkbox[NavyBlue]{n1}{M_{g,p}}   =
\frac{     \eqnmarkbox[Cerulean]{n3a}{\sum_{d \in D}}     \eqnmarkbox[blue]{n4a}{R(g,p,d)}
\times      \eqnmarkbox[BlueViolet]{n5}{E(g,d)}     }{     \eqnmarkbox[Cerulean]{n3b}{\sum_{d
\in D}}     \eqnmarkbox[blue]{n4b}{R(g,p,d)}   } \end{equation*} \annotate[yshift=1em]{left}{n1}{Weight
gene-by-phenotype \\evidence score matrix}  \annotate[yshift=-2em]{below,left}{n3a,n3b}{Iterate
over all diseases} \annotate[yshift=-2.5em,xshift=2.5em]{below,right}{n4a,n4b}{Binary
gene-by-phenotype \\relationship matrix,\\ (1=relationship, 0=no relationship)} \annotate[yshift=2em]{l
gene-by-disease \\evidence score matrix}
```

Single-cell transcriptomic atlases

Let us denote:

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

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

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

```
{=tex #eq-ctd-filter} \begin{equation*}   \eqnmarkbox[purple]{f1}{F(g,i,c)}   =   \begin{cases}
\eqnmarkbox[WildStrawberry]{f2}{r_{g,i}},   \text{ }l_i = c\\0,   \text{ }l_i \neq c
\end{cases} \end{equation*} \annotate[yshift=1em]{left}{f1}{Filtered gene-by-cell expression
matrix}  \annotate[yshift=2em]{left}{f2}{Expression of gene $g$ in cell $i$}
```

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

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

{=tex #eq-ctd-specificity} \begin{equation*} \eqnmarkbox[orange]{s1}{S_{g,c}} =
\frac{ \eqnmarkbox[purple]{s3a}{ \frac{ \sum_{i=1}^{|L|} F(g,i,c)
}{ N_c } } }{ \eqnmarkbox[OrangeRed]{s6}{\sum_{r=1}^{k}}(
\eqnmarkbox[purple]{s3b}{ \frac{ \sum_{i=1}^{|L|} F(g,i,c) }{ N_c
} }) } \end{equation*} \annotate[yshift=1em]{left}{s1}{Gene-by-cell type
specificity matrix} \annotate[yshift=2em]{left}{s3a,s3b}{Compute mean expression of each
gene per cell type} \annotate{below,left}{s6}{Compute row sums of \\mean gene-by-cell
type matrix}

Symptom-cell type associations

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

{=tex #eq-symptoms} \begin{equation*} \frac{ \eqnmarkbox[Chartreuse3]{g1}{|G_{d
\cap p \cap c} |} }{ \eqnmarkbox[Emerald]{g2}{|G_{d,p}|}} \geq \eqnmarkbox[SeaGreen]{g3}{
\end{equation*} \annotate[yshift=1em]{left}{g1}{Intersect between \\symptom genes ($G_{d,p}$)
and driver genes ($G_{p,c}$)} \annotate[yshift=-1em]{below,left}{g2}{Symptom genes
\\(i.e. genes annotated to a phenotype\\ via a specific disease)} \annotate[yshift=-1em]{below,right}{
proportion of overlap \\between $G_{d,p,c}$ and $G_{d,p}$}

Annotation of phenotypes using generative large language models

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

```
{=tex #eq-gpt} \begin{equation*}   \eqnmarkbox[Brown4]{nss}{NSS_p}   =   \frac{      \eqnmarkbox[Goldenr
(       \eqnmarkbox[Goldenrod4]{nss3}{F_{pj}}       \times        \eqnmarkbox[IndianRed4]{nss4}{W_j}
)     }{     \eqnmarkbox[Tan]{nss5}{\sum_{j=1}^{m}(\max\{F_j\} \times W_j)}    } \times
100 \end{equation*} \annotate[yshift=1em]{left}{nss}{Normalised Severity Score \\for each
phenotype} \annotate[yshift=3em]{left}{nss2}{Sum of weighted annotation values \\across
all metrics} \annotate[yshift=3em]{right}{nss3}{Numerically encoded annotation value \\of
metric $j$ for phenotype $p$} \annotate[yshift=1em]{right}{nss4}{Weight for metric $j$}
\annotate[yshift=-1em]{below,right}{nss5}{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 14.5

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;  LAPACK versi

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_0.5.8.1 rstudioapi_0.16.0 yaml_2.3.8        rmarkdown_2.26
 [9] knitr_1.45        jsonlite_1.8.8    xfun_0.43         digest_0.6.35
[13] rlang_1.1.3       evaluate_0.23
```