

# PCA and hierarchical clustering of LFQ intensities of significantly changed proteins in uninfected, IAV-infected, *S. pneumoniae*-infected, and IAV and *S. pneumoniae*-infected hearts

Platt et al (2022) mBio; Topic: flu Strep co-infection

2022-05-11

Salma Elhassa

## Table of Contents

<b>1</b>	<b><i>Introduction</i></b>	<b>1</b>
1.1	Background	1
1.2	Biological hypothesis	2
1.3	Significance	2
<b>2</b>	<b><i>Materials and Methods</i></b>	<b>2</b>
2.1	Samples	2
2.2	Experimental procedure	2
2.3	Statistical methods	2
<b>3</b>	<b><i>Exploratory data analysis</i></b>	<b>2</b>
<b>4</b>	<b><i>Results</i></b>	<b>3</b>
<b>5</b>	<b><i>Conclusions</i></b>	<b>9</b>
5.1	Biological conclusions	9
5.2	Future work	9
<b>6</b>	<b><i>References</i></b>	<b>9</b>
6.1	Cite paper	9
6.2	Cite code repository	9
6.3	Cite data file	<a href="#">9-10</a>

## 1 Introduction

For over a century, it has been reported that influenza infection leads to a deadly form of pneumonia caused by bacteria. We present the first in vivo evidence that influenza infection promotes *Streptococcus pneumoniae* translocation to the myocardium, necrotic cell damage, and proteomic remodeling of the heart.

### 1.1 Background

Primary influenza infection leads to a potentiation of the pulmonary bacterial necrosis in a ply-dependent manner, suggesting that ply could be targeted to reduce the *S. pneumoniae*-driven cellular toxicity in heart tissue.

## **1.2 Biological hypothesis**

Proteomic data of IAV-infected hearts showed differential changes in proteins associated with oxidative stress. We hypothesized that such stress may influence cytotoxicity of cardiomyocytes upon *S. pneumoniae* infection.

## **1.3 Significance**

Adverse cardiac events are a complication of viral and bacterial pneumonia, and necroptosis (i.e., necrosis inhibition) reduces damage and proteome changes associated with health.

# **2 Materials and Methods**

Male and female 6-week-old C57BL/6N (B6NTac) mice from Taconic Biosciences were used in this study (Rensselaer, New York). Warren Alexander, who works at the Walter and Eliza Hall Institute of Medical Research Parkville, Victoria, Australia, agreed to give us MLKL KO mice (24). Mice were given 250 PFU of pandemic influenza virus A/California/7/2009 through their noses. TIGR4 *S. pneumoniae* was given to the mice intracranially on day 10 as before (16). At 12 days after they were infected with influenza, mice were killed to get heart tissue. The data shown is from two separate experiments that used 6 mice, three of each sex.

## **2.1 Samples**

DATA SET S1 Global proteome changes in the heart of mice during secondary bacterial infection to influenza. Download Data Set S1, XLSX file, 0.2 MB. Copyright © 2022 Platt et al. This content is distributed under the terms of the Creative Commons Attribution 4.0 International license. Principal-component analysis (PCA) and hierarchical clustering of LFQ intensities of significantly changed proteins (ANOVA, FDR of 0.05) among uninfected, IAV-infected, *S. pneumoniae*-infected, and IAV and *S. pneumoniae*-infected hearts.

## **2.2 Experimental procedure**

To identify and quantify the mouse proteome, we performed a database search for the UniProt protein database, used the MaxQuant-Andromeda software suite (version 1.6.5.0) and applied the following parameters: 4.5 ppm and 20 ppm mass tolerances for precursors and fragments, respectively.

## **2.3 Statistical methods**

A nonparametric Kruskal-Wallis H test with Dunn's post hoc analysis was used for multiple-group analyses. A two-way ANOVA with Sidak's post hoc analysis was used for grouped analyses. Prism 7 (GraphPad Software, La Jolla, CA) was used to do these statistical analyses. , *P* 0.05; , **P** 0.01; , *P* 0.001.

# **3 Exploratory data analysis**

Gene ontology analysis of proteins that were up a lot during SBI showed more innate immune activity, oxidative processes, and changes to ion homeostasis. Immunoblots showed that the angiotensin-converting enzyme 2 was more active and had more of it, as well as more of it being made. In a model of sequential infections in human cardiomyocytes, we found that influenza makes *S. pneumoniae* more toxic by causing oxidative stress, which makes bacterial toxin-induced necrotic cell death more likely. Influenza caused heart cells to have more receptors that help bacteria attach,

like polymeric immunoglobulin receptor and fibronectin leucine-rich transmembrane protein 1. Finally, mice that didn't have necroptosis had better innate immune responses, less virus-related pathways, and better mitochondrial function after SBI.

## 4 Results

Analysis of variance (ANOVA) and a permutation-based false-discovery rate (FDR) of 0.05 led to 288 proteins that showed a difference between the four conditions. Principal-component analysis (PCA) clustered biological replicates, suggesting that each group had a different proteomic profile. Single-infection groups were found to be clustered together, which suggests that the host's heart responses changed. When mice had SBI, their heart proteins did not cluster together, which suggests that their bodies are going through more changes to their proteomes. Based on their abundance profiles, proteins could be divided into three groups that could be shown in a heat map.

```
setwd("~/Library/CloudStorage/OneDrive-Hunter-CUNY/BIO 47120")
getwd() # show working directory to confirm

## [1] "/Users/salmaelhassa/Library/CloudStorage/OneDrive-Hunter-CUNY/BIO 47120"

# load libraries and read data
library(tidyverse) # load a package into memory

## — Attaching packages ————— tidyverse 1.3.1 —

## ✓ggplot2 3.3.5   ✓purrr  0.3.4
## ✓tibble 3.1.6   ✓dplyr  1.0.8
## ✓tidyr  1.2.0   ✓stringr 1.4.0
## ✓readr  2.1.2   ✓forcats 0.5.1

## — Conflicts ————— tidyverse_conflicts() —
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

library(readxl) # load another package
library(pheatmap)
library(ggrepel)
library(broom)
s1 <- read_xlsx("mbio.03257-21-sd001.xlsx")
glimpse(s1)

## Rows: 288
## Columns: 64
## $ WTheart_1      <dbl> 25.2771, 21.6610, 22.2139, 2...
## $ WTheart_2      <dbl> 26.4097, 21.9625, 20.6939, 1...
## $ WTheart_3      <dbl> 25.6467, 22.7041, 22.0564, 2...
## $ WTheart_Flu_1  <dbl> 33.4129, 27.2926, 27.7662, 2...
## $ WTheart_Flu_2  <dbl> 33.3624, 27.3277, 28.0178, 2...
## $ WTheart_Flu_3  <dbl> 33.3950, 27.4114, 28.0660, 2...
## $ WTheart_Flu_4  <dbl> 33.3460, 27.2185, 27.9679, 2...
## $ WTheart_Flu_5  <dbl> 33.3616, 27.3427, 27.7902, 2...
## $ WTheart_Flu_6  <dbl> 33.3576, 27.1108, 28.0834, 2...
## $ WTheart_Flu_7  <dbl> 33.3221, 27.2311, 28.0112, 2...
## $ WTheart_Flu_8  <dbl> 33.3188, 27.1038, 27.8887, 2...
## $ WTheart_Spn_1  <dbl> 33.0757, 27.6763, 27.6525, 2...
## $ WTheart_Spn_2  <dbl> 32.7469, 27.7278, 27.4031, 2...
## $ WTheart_Spn_3  <dbl> 33.2891, 26.4799, 26.6309, 2...
## $ WTheart_Spn_4  <dbl> 33.2930, 25.7215, 26.1927, 2...
## $ `WTheart_Spn+Flu_1` <dbl> 32.7470, 29.1517, 27.1300, 2...
## $ `WTheart_Spn+Flu_2` <dbl> 32.8580, 29.1811, 27.1426, 2...
## $ `WTheart_Spn+Flu_3` <dbl> 32.9041, 29.1930, 24.7301, 2...
## $ `WTheart_Spn+Flu_4` <dbl> 33.2642, 29.4660, 25.3363, 2...
## $ `WTheart_Spn+Flu_5` <dbl> 32.8956, 28.8063, 26.9364, 2...
```

```

## $ `WTheart_Spn+Flu_6`      <dbl> 33.1618, 28.8994, 26.5128, 2...
## $ `C: ANOVA Significant`    <chr> "+", "+", "+", "+", "+", "+" ...
## $ `N: Peptides`            <dbl> 1400, 29, 129, 97, 18, 17, 2...
## $ `N: Razor + unique peptides` <dbl> 1400, 29, 129, 97, 18, 17, 1...
## $ `N: Unique peptides`      <dbl> 1400, 29, 121, 97, 18, 17, 1...
## $ `N: Sequence coverage [%]` <dbl> 48.3, 45.9, 34.9, 38.8, 32.5...
## $ `N: Unique + razor sequence coverage [%]` <dbl> 48.3, 45.9, 34.9, 38.8, 32.5...
## $ `N: Unique sequence coverage [%]` <dbl> 48.3, 45.9, 32.8, 38.8, 32.5...
## $ `N: Mol. weight [kDa]`    <dbl> 3906.4000, 87.4280, 564.8100...
## $ `N: Q-value`              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ `N: Score`                <dbl> 323.310, 323.310, 323.310, 3...
## $ `N: Intensity`            <dbl> 3.4109e+12, 1.4369e+11, 7.25...
## $ `N: MS/MS count`          <dbl> 61066, 1699, 2501, 1408, 554...
## $ `N: KOheart_Flu_1`        <dbl> 33.5055, 30.1726, 28.6582, 2...
## $ `N: KOheart_Flu_2`        <dbl> 33.3996, 30.0208, 28.5983, 2...
## $ `N: KOheart_Flu_3`        <dbl> 33.1025, 27.5168, 27.3319, 2...
## $ `N: KOheart_Flu_4`        <dbl> 33.1150, 27.7262, 27.4882, 2...
## $ `N: KOheart_Flu_5`        <dbl> 33.1680, 28.5474, 27.0021, 2...
## $ `N: KOheart_Flu_6`        <dbl> 33.1592, 28.3598, 26.5028, 2...
## $ `N: KOheart_Spn_1`        <dbl> 32.7800, 25.8975, 27.4959, 2...
## $ `N: KOheart_Spn_2`        <dbl> 32.6056, 26.0503, 27.2593, 2...
## $ `N: KOheart_Spn_3`        <dbl> 33.1464, 27.2903, 27.3860, 2...
## $ `N: KOheart_Spn_4`        <dbl> 33.1319, 27.3353, 27.2431, 2...
## $ `N: KOheart_Spn_5`        <dbl> 33.0282, 27.8088, 27.6174, 2...
## $ `N: KOheart_Spn_6`        <dbl> 33.0186, 28.0666, 27.5100, 2...
## $ `N: KOheart_Spn+Flu_1`    <dbl> 32.9661, 29.5358, 27.5820, 2...
## $ `N: KOheart_Spn+Flu_2`    <dbl> 32.9142, 29.2381, 27.6580, 2...
## $ `N: KOheart_Spn+Flu_3`    <dbl> 33.0433, 29.5942, 26.2962, 2...
## $ `N: KOheart_Spn+Flu_4`    <dbl> 32.8150, 29.5519, 26.3991, 2...
## $ `N: KOheart_Spn+Flu_5`    <dbl> 33.0610, 28.8014, 26.2630, 2...
## $ `N: KOheart_Spn+Flu_6`    <dbl> 33.0924, 28.9945, 26.0309, 2...
## $ `N: -Log ANOVA p value`   <dbl> 21.19040, 12.90160, 8.20353,...
## $ `N: ANOVA q-value`        <dbl> 0.00000e+00, 0.00000e+00, 0...
## $ `N: WTheart_Flu_9`        <dbl> 33.1990, 27.5386, 26.8193, 2...
## $ `N: WTheart_Flu_10`       <dbl> 33.2188, 27.9519, 26.4961, 2...
## $ `N: WTheart_Flu_11`       <dbl> 33.3648, 27.2435, 26.2031, 2...
## $ `N: WTheart_Flu_12`       <dbl> 32.9569, 26.2007, 26.3605, 2...
## $ `T: Protein IDs`          <chr> "sp|A2ASS6|TITIN_MOUSE;REV__...
## $ `T: Majority protein IDs`  <chr> "sp|A2ASS6|TITIN_MOUSE", "sp...
## $ `T: id`                   <dbl> 61, 72, 75, 78, 90, 92, 99, ...
## $ `T: Protein`              <chr> "UBP24", "ACACB", "AGRF5", "...
## $ `T: Accession`            <chr> "B1AY13", "E9Q4Z2", "G5E8Q8"...
## $ `T: Protein Name`         <chr> "Ubiquitin carboxyl-terminal...
## $ `T: Gene`                 <chr> "Usp24", "Acacb", "Adgrf5", ...

```

```
names(s1) # show colnames
```

```

## [1] "WTheart_1"
## [2] "WTheart_2"
## [3] "WTheart_3"
## [4] "WTheart_Flu_1"
## [5] "WTheart_Flu_2"
## [6] "WTheart_Flu_3"
## [7] "WTheart_Flu_4"
## [8] "WTheart_Flu_5"
## [9] "WTheart_Flu_6"
## [10] "WTheart_Flu_7"
## [11] "WTheart_Flu_8"
## [12] "WTheart_Spn_1"
## [13] "WTheart_Spn_2"
## [14] "WTheart_Spn_3"
## [15] "WTheart_Spn_4"
## [16] "WTheart_Spn+Flu_1"
## [17] "WTheart_Spn+Flu_2"
## [18] "WTheart_Spn+Flu_3"
## [19] "WTheart_Spn+Flu_4"
## [20] "WTheart_Spn+Flu_5"
## [21] "WTheart_Spn+Flu_6"

```

```

## [22] "C: ANOVA Significant"
## [23] "N: Peptides"
## [24] "N: Razor + unique peptides"
## [25] "N: Unique peptides"
## [26] "N: Sequence coverage [%]"
## [27] "N: Unique + razor sequence coverage [%]"
## [28] "N: Unique sequence coverage [%]"
## [29] "N: Mol. weight [kDa]"
## [30] "N: Q-value"
## [31] "N: Score"
## [32] "N: Intensity"
## [33] "N: MS/MS count"
## [34] "N: KOheart_Flu_1"
## [35] "N: KOheart_Flu_2"
## [36] "N: KOheart_Flu_3"
## [37] "N: KOheart_Flu_4"
## [38] "N: KOheart_Flu_5"
## [39] "N: KOheart_Flu_6"
## [40] "N: KOheart_Spn_1"
## [41] "N: KOheart_Spn_2"
## [42] "N: KOheart_Spn_3"
## [43] "N: KOheart_Spn_4"
## [44] "N: KOheart_Spn_5"
## [45] "N: KOheart_Spn_6"
## [46] "N: KOheart_Spn+Flu_1"
## [47] "N: KOheart_Spn+Flu_2"
## [48] "N: KOheart_Spn+Flu_3"
## [49] "N: KOheart_Spn+Flu_4"
## [50] "N: KOheart_Spn+Flu_5"
## [51] "N: KOheart_Spn+Flu_6"
## [52] "N: -Log ANOVA p value"
## [53] "N: ANOVA q-value"
## [54] "N: WTheart_Flu_9"
## [55] "N: WTheart_Flu_10"
## [56] "N: WTheart_Flu_11"
## [57] "N: WTheart_Flu_12"
## [58] "T: Protein IDs"
## [59] "T: Majority protein IDs"
## [60] "T: id"
## [61] "T: Protein"
## [62] "T: Accession"
## [63] "T: Protein Name"
## [64] "T: Gene"

```

```

# Data wrangling: turn into a tidy table (one row one value)

```

```

## select intensity and protein id and make a long table
s1.long <- s1 %>%
  select(1:21, 34:51, 54:57, 61) %>%
  pivot_longer(1:43, values_to = "intensity", names_to = "sample_id")

```

```

## make colnames computer-friendly
colnames(s1.long)[1] <- c("protein_id")

```

```

## Extract and add categorical variable columns (col headings are not supposed to contain variables)

```

```

#### add a genotype column
s1.long <- s1.long %>%
  mutate(genotype = if_else(str_detect(sample_id, "KO"), "knock_out", "wild_type"))

```

```

#### checksum to validate
s1.long %>% group_by(genotype) %>% count()

```

```

## # A tibble: 2 × 2
## # Groups:   genotype [2]
##   genotype     n
##   <chr>   <int>
## 1 knock_out 5184
## 2 wild_type 7200

```

```

#### add a pathogen column
s1.long <- s1.long %>%
  mutate(pathogen = if_else(str_detect(sample_id, "heart_Flu_"), "flu",
    if_else(str_detect(sample_id, "heart_Spn_"), "strep",
      if_else(str_detect(sample_id, "Spn.Flu_"), "both", "control"))))

```

```

#### checksum to validate
s1.long %>% group_by(pathogen) %>% count()

```

```

## # A tibble: 4 × 2
## # Groups:   pathogen [4]
##   pathogen     n
##   <chr>   <int>
## 1 both    3456
## 2 control  864
## 3 flu     5184
## 4 strep   2880

```

```

tmp <- s1.long %>% group_by(sample_id, pathogen) %>% count()

```

```

#### add a replicate column
s1.long <- s1.long %>%
  mutate(rep = as.numeric(str_replace(sample_id, "^._(\\d+)$", "\\1")))

```

```

#### checksum to validate
s1.long %>% group_by(pathogen) %>% count()

```

```

## # A tibble: 4 × 2
## # Groups:   pathogen [4]
##   pathogen     n
##   <chr>   <int>
## 1 both    3456
## 2 control  864
## 3 flu     5184
## 4 strep   2880

```

```

#tmp <- s1.long %>% group_by(sample_id, pathogen) %>% count()
#### checksum to validate
s1.long %>% group_by(rep) %>% count()

```

```

## # A tibble: 12 × 2
## # Groups:   rep [12]
##   rep     n
##   <dbl> <int>
## 1     1  2016
## 2     2  2016
## 3     3  2016
## 4     4  1728
## 5     5  1440
## 6     6  1440
## 7     7   288
## 8     8   288
## 9     9   288
## 10    10   288
## 11    11   288
## 12    12   288

```

```

# cluster analysis of genes and samples by heatmap

```

```

## heatmap by geom_tile (not very informative)
s1.long <- s1.long %>% filter(!is.na(protein_id))
s1.long %>% ggplot(aes(x = sample_id, y = protein_id, fill = intensity)) +
  geom_tile() +
  scale_fill_gradient(low="white", high="blue") +
  theme_bw()

```

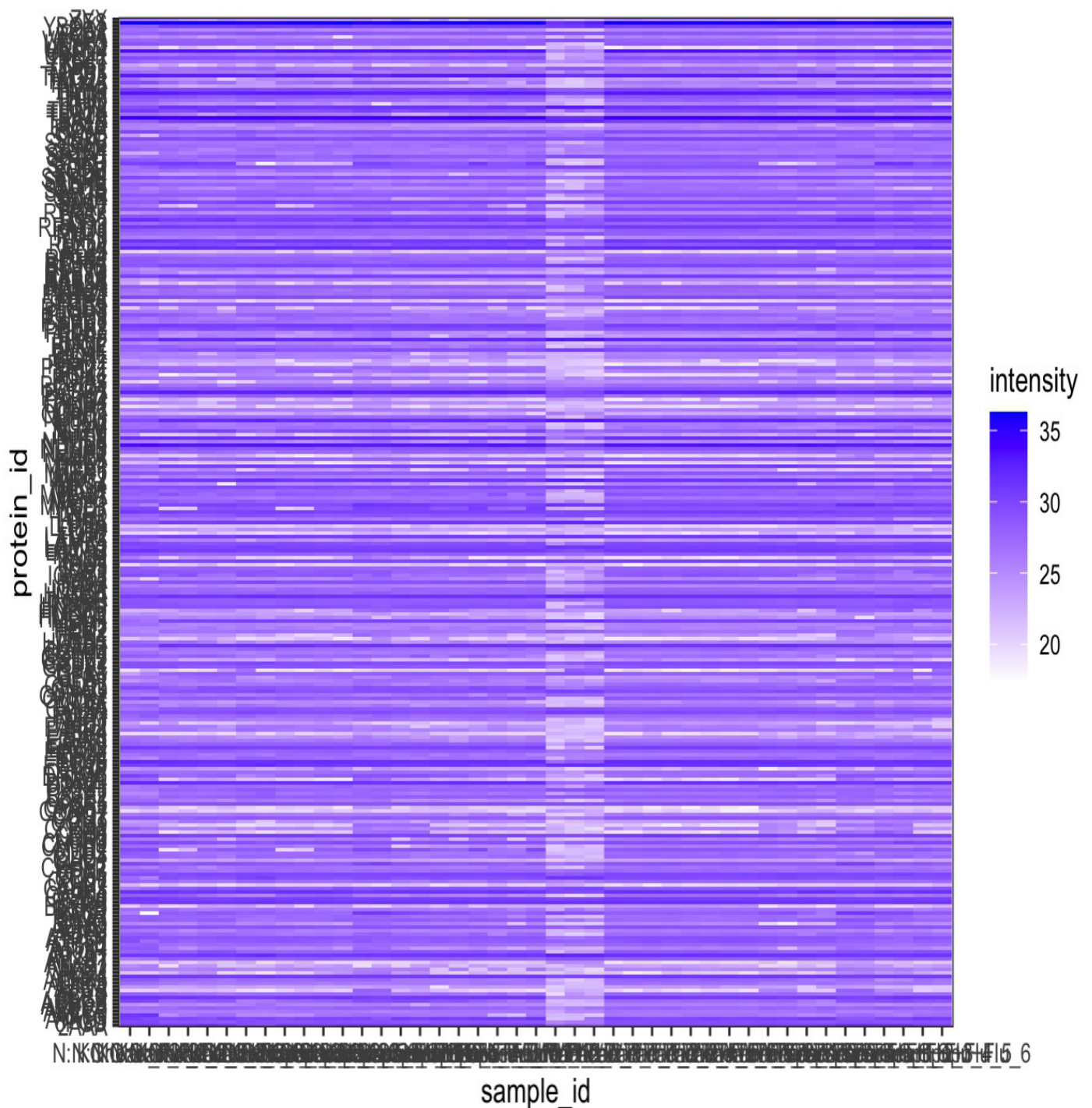
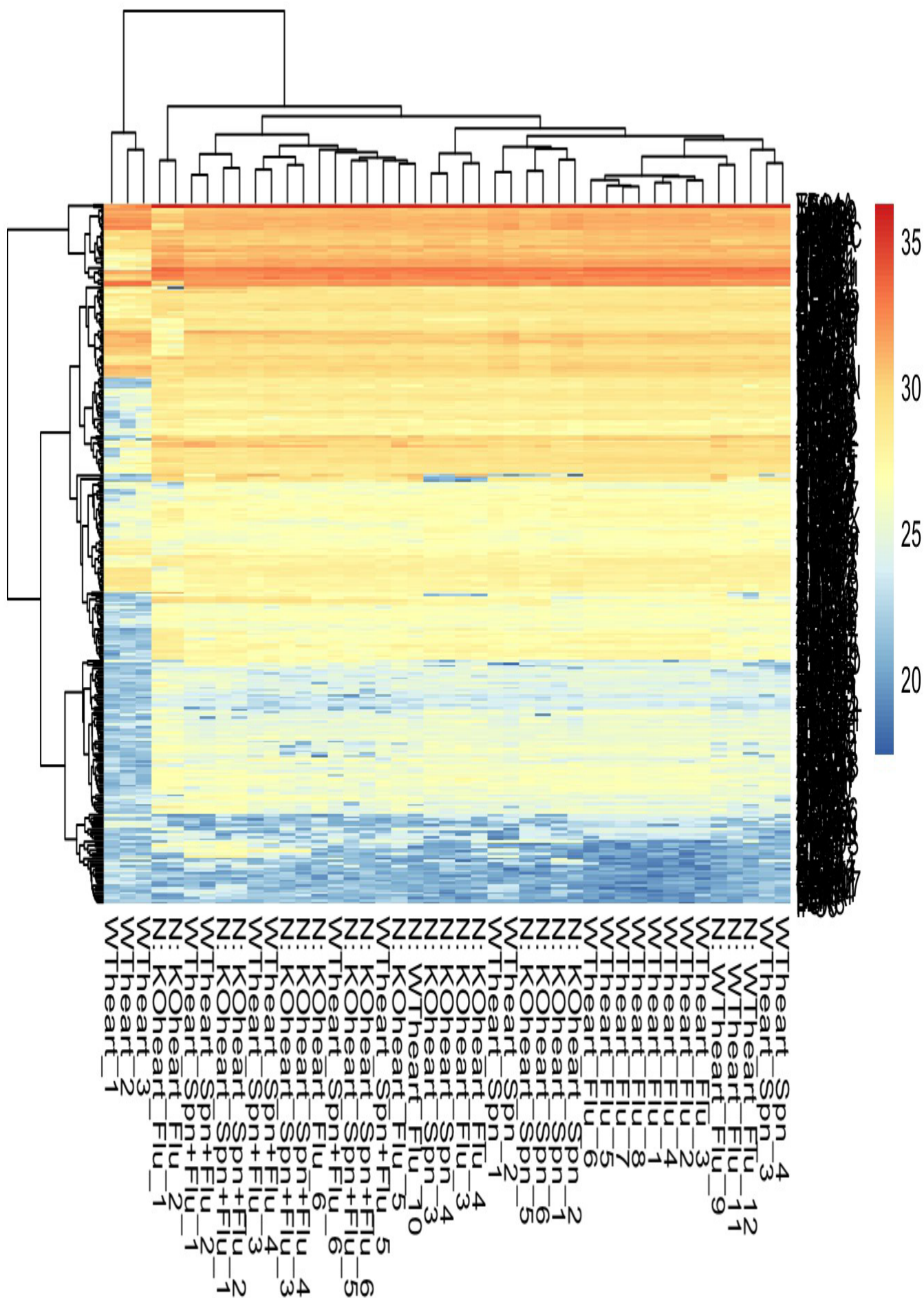


Figure A: cluster analysis of genes and samples by heatmap. PCA and hierarchical clustering of LFQ intensities of significantly changed proteins (ANOVA, FDR of 0.05) in uninfected, IAV-infected, *S. pneumoniae*-infected, and IAV and *S. pneumoniae*-infected hearts. The interpretation of a heat map is dependent on the data represented on that particular map. Remember that warmer colors represent higher values, while cooler colors represent lower values. In this map, blue is the warmest color and white is the coolest. You must analyze colors and comprehend the map's intensity. In the image above, the intensity is at its peak (blue) during the initial searches comparing the sample\_id to the protein\_id.

```
## heatmap by pheatmap, need a matrix (not table)
s1.wide <- s1.long %>% select(1:3) %>% pivot_wider(names_from = "sample_id", values_from = "intensity")
s1.mat <- as.matrix(s1.wide[,2:44])
rownames(s1.mat) <- s1.wide %>% pull(protein_id)
pheatmap(s1.mat, scale = "none")
```







Figures B: Heatmap by pheatmap. Proteins could be classified into three distinct clusters based on abundance profiles. These clusters were enriched in terms associated with T-cell differentiation, immune responses, and metabolic processes, among others. Warm colors fall within the red-yellow spectrum (intense), whereas cool colors fall within the blue-green spectrum (less intense). The dendrograms along the sides illustrate how the variables and rows are clustered independently. The heat map depicts each row and column's data value (possibly standardized so they all fit in the same range). Any patterns in the heat map could suggest a relationship between the rows and columns. A heatmap is a graphical representation used to depict the intensity of data at specific geographical locations. A colored overlay will appear on top of the map when the Heatmap Layer is enabled.

## 5 Conclusions

### 5.1 Biological conclusions

Our findings show that influenza virus and *S. pneumoniae* both change the proteins in the heart, and this changes even more during SBI. This causes more toxicity, necrotic cell death, and major changes in the heart's proteins. Taken together, our findings show how pulmonary pathogens work together to make extrapulmonary organs sick.

### 5.2 Future work

Secondary bacterial infections to influenza virus aren't clear about how they affect the body outside of the lungs. In this study, we used quantitative proteomics and molecular methods to figure out how influenza causes bacteria to damage the heart and change the proteins in the body's proteome.

## 6 References

Platt, Maryann P., et al. "Pandemic Influenza Infection Promotes Streptococcus Pneumoniae Infiltration, Necrotic Damage, and Proteomic Remodeling in the Heart." *MBio*, vol. 13, no. 1, 2022, <https://doi.org/10.1128/mbio.03257-21>.

Platt, Maryann P., et al. "Pandemic Influenza Infection Promotes Streptococcus Pneumoniae Infiltration, Necrotic Damage, and Proteomic Remodeling in the Heart." *MBio*, vol. 13, no. 1, 2022, Fig. 1G, <https://doi.org/10.1128/mbio.03257-21>.

### 6.1 Cite paper

Platt, M. P., Lin, Y.-H., Wiscovitch-Russo, R., Yu, Y., & Gonzalez-Juarbe, N. (2022). Pandemic influenza infection promotes streptococcus pneumoniae infiltration, necrotic damage, and proteomic remodeling in the heart. *MBio*, 13(1). <https://doi.org/10.1128/mbio.03257-21>

### 6.2 Cite code repository

We say that all of the data that backs up the findings of this study can be found in the manuscript and from the corresponding author if you make a reasonable request. The original proteomic data has been added to ProteomeXchange with the accession number PXD016137.

### 6.3 Cite data file

DATA SET S1 Global proteome changes in the heart of mice during secondary bacterial infection to influenza.  
Download Data Set S1, XLSX file, 0.2 MB. Copyright © 2022 Platt et al. This content is distributed under the terms of the Creative Commons Attribution 4.0 International license.