
Examining the Immunological and Transcriptional Response upon SARS-CoV-2 Infection using scRNA-Seq

— Madeline Iseminger, Trevor Kwan, Jackson Moore, Yilin Qiu —



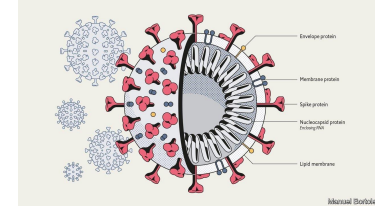
Project Outline

1. **Background:** Dynamics and biases of SARS-CoV-2 infection
2. **Dataset + Aims:** scRNA-Sequencing of SARS-CoV-2 infection in 54 patients
3. **Methods:** Our statistical approach
4. **Results:** Immunological and transcriptional responses
5. **Considerations:** Limitations and future directions
6. **Conclusion**



Background:

- The likelihood and severity of SARS-CoV-2 infection is associated with several different factors:
 - Patient age
 - Comorbidities (obesity, diabetes, hypertension)
 - Socioeconomic inequities
 - **Patient Sex**
- A study published in the *New England Journal of Medicine* last year screened nearly 10000 Icelandic individuals for SARS-CoV-2 infection:
 - 16.7% of males tested positive
 - 11% of females tested positive
- Jin et al. in 2020 found that:
 - Males tended to have more serious complications from infection
 - Males accounted for 2.4x more deaths than females in a cohort of patients from China
- Elgendy et al 2020 reports that:
 - The outbreak in Italy saw 1.7x more deaths in males than females



Background:

- Potential explanations for increased morbidity and mortality in males:
 - Higher expression of angiotensin-converting enzyme-2 (ACE2) receptor; entry receptor for SARS-CoV-2
 - Lifestyle & behaviour: studies report that men are less likely to follow preventative and social distancing measures; increased smoking and alcohol consumption in males, etc.
 - Sex-based immunological differences: immunological features of X and Y chromosomes



Background:

- Sex-based immunological differences that may contribute to SARS-CoV-2 infection rates:
 - **Decreased expression of CD200R**, a receptor on monocytes and lymphocytes, in turn promotes Type 1 IFN response. Female mice with limited CD200R had reduced mortality upon hepatitis coronavirus infection
 - Sex-based genes implicated in differential immune and humoral responses: **Human leucocytes antigen (HLA), IL-4, IL-10, IL-12 receptor**
 - **Cell-type specific responses to SARS-CoV-2 infection:** male mice were found to have more abundant inflammatory cell-types (macrophages and neutrophils) and increased mortality
 - **Estrogen receptor signaling** associated decreased mortality
 - **Increased humoral response** in severe female covid patients



Dataset

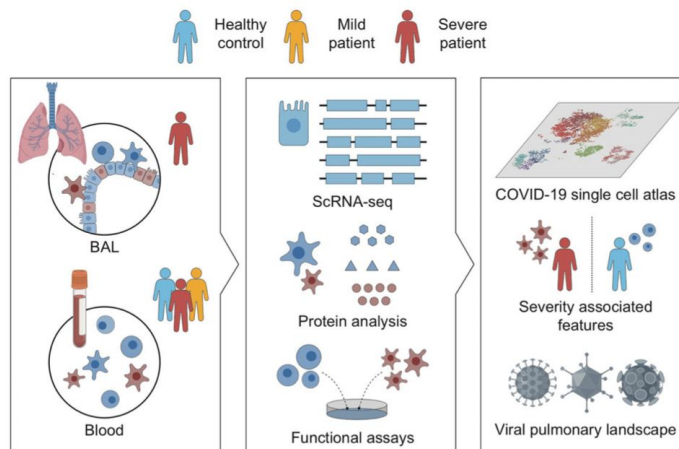
scRNA-Seq data:

- Bost et al. 2021, Deciphering the State of Immune Silence in Fatal COVID-19 Patients, *Nature Communications*

54 scRNA-Seq samples from 36 patients

- Separated by: sex, tissue source (blood or lung), COVID-severity
- Limitations of dataset: 9 female patients, 27 male patients. Only 1 female control

Table 1 Anagraphic and coexisting disorders of enrolled patients and healthy donors.			
Characteristics	Healthy controls N = 5	Mild patients N = 10	Severe patients (ICU) N = 21
<i>Anagraphic</i>			
Age, yr: median (IQR)*	66 (64-73)	69 (56-80)	67 (58-70)
Male, no. (%)	4 (80)	6 (60)	17 (81)
<i>Coexisting disorder, no. (%)</i>			
Any	2 (40)	10 (100)	17 (81)
Obesity	0 (0)	2 (22)	3 (14)
Hypertension	2 (40)	10 (100)	11 (52)
Diabetes	0 (0)	3 (30)	7 (33)
Chronic obstructive pulmonary disease	0 (0)	2 (20)	1 (5)
Cardiovascular disease	0 (0)	5 (50)	3 (14)
Chronic kidney disease	0 (0)	2 (20)	1 (5)
Active malignancies	0 (0)	0 (0)	2 (10)

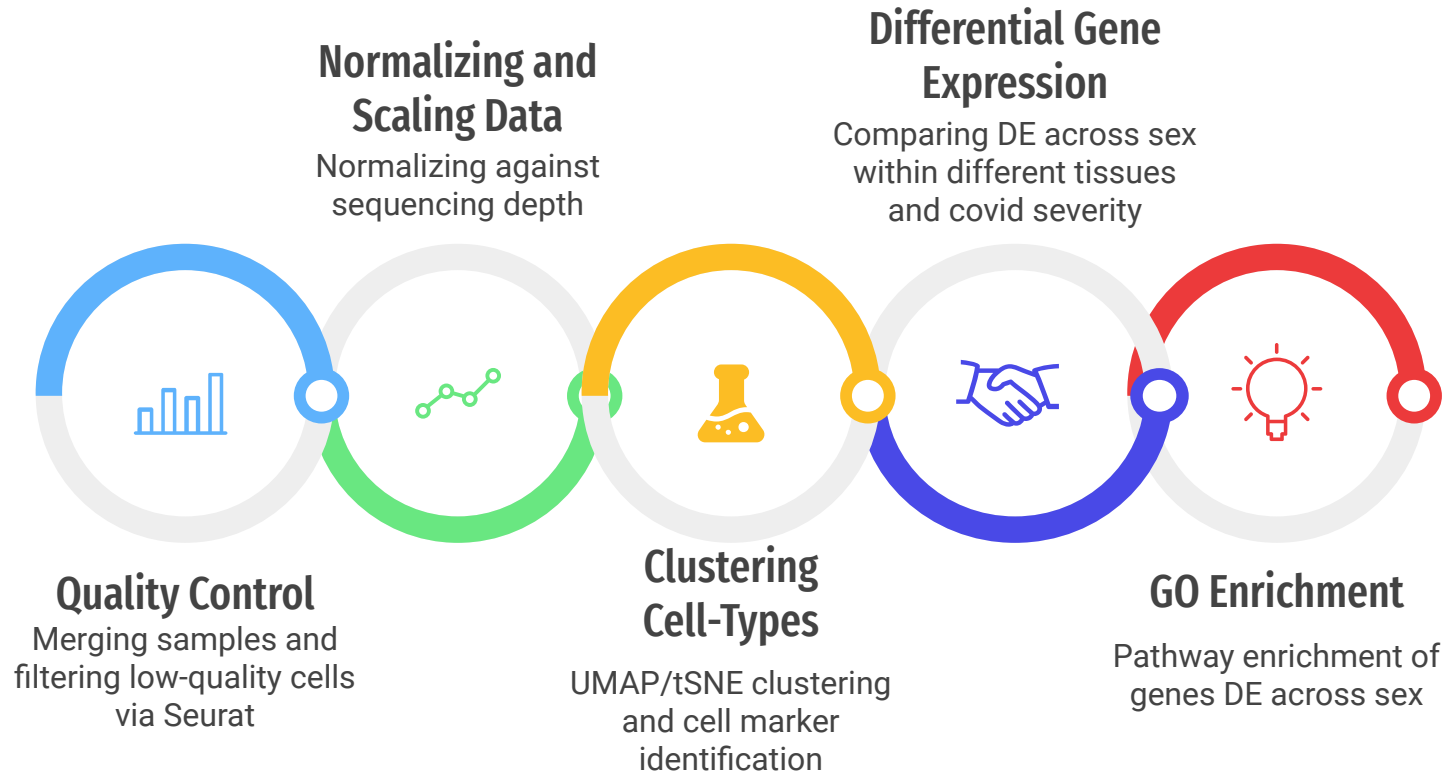


Aims

1. Determine the top differentially expressed genes in each immune cell type cluster between male and female patients across SARS-CoV-2 virus status.
 - Subset by disease severity: mild, severe, and healthy
 - Subset by T-cells, B-cells, NK cells, and Neutrophils
2. Determine the top differentially expressed genes in each immune cell type cluster between male and female patients across tissue type.
 - Subset by tissue type: Blood and Lavage/BAL (lung)
 - Subset by T-cells, B-cells, NK cells, and Neutrophils
3. Examine what immunological pathways may be differentially implicated between sex
 - Perform GO enrichment on DE genes from aim 2

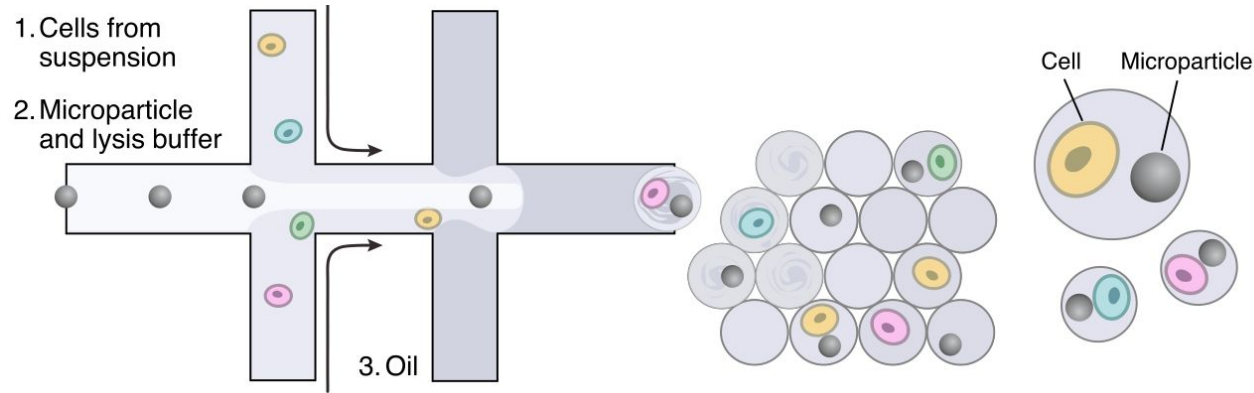


Methods & Workflow



scRNA Filtering Parameters

Single cell RNA seq technology combines individual cells with barcoded microparticles within small droplets



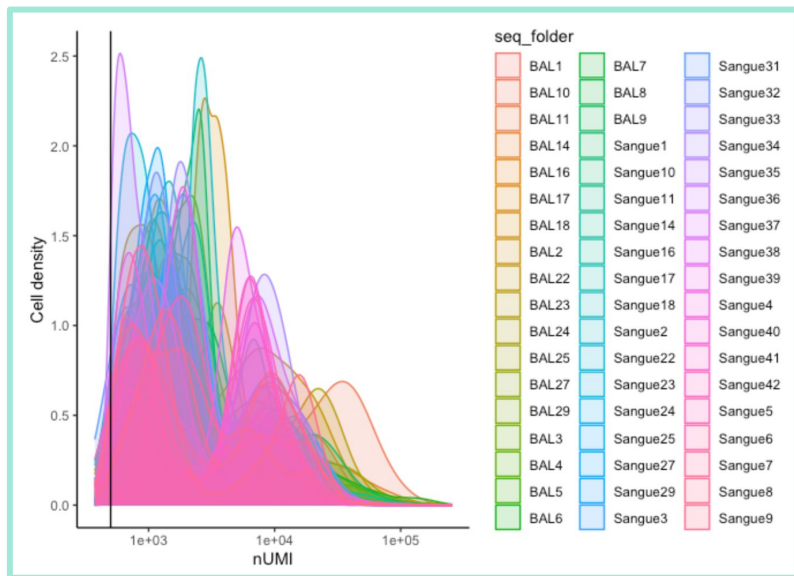
Need to consider:

- Empty droplets
- Doublets (droplets with more than one cell)
- Dying cells
- Red blood cells

Cell-Level Filtering: Empty droplets

Droplets with no cells are expected to have very low counts/transcripts. We can filter these out by setting:

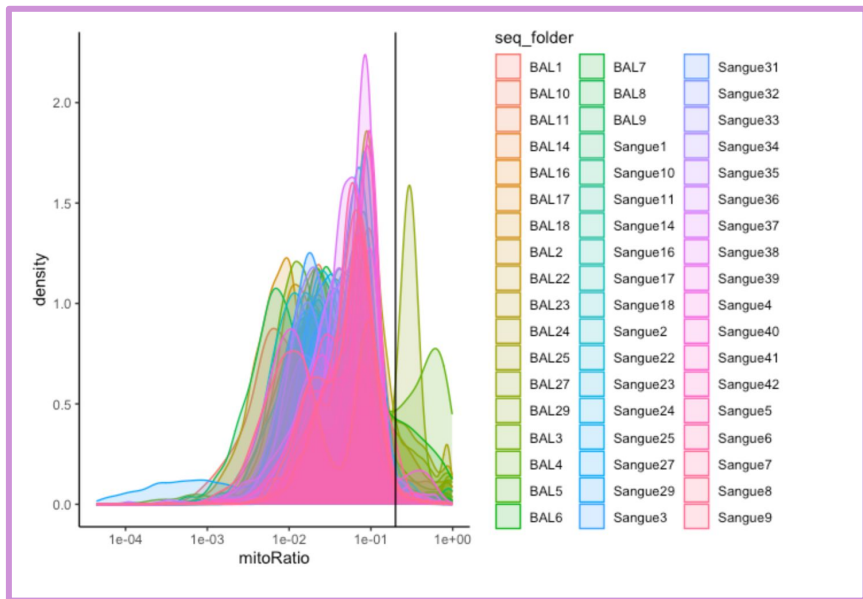
- The number of transcripts per cell (nUMI) > 500
- The number of expressed genes per cell > 250



Cell-Level Filtering: Dying Cells and Red Blood Cells

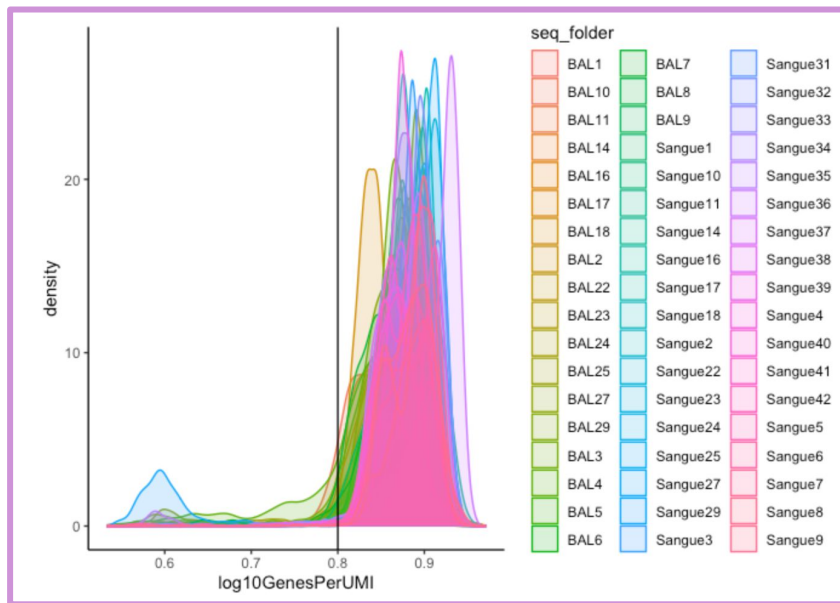
Mitochondrial expressed genes are enriched in dying or apoptotic cells.

- Can filter these out by removing cells with >20% mitochondrial gene expression



Red blood cells lack a nucleus and don't actively express genes. They should thus have a low gene:count ratio.

- Cells < 0.80 log10 Genes per UMI removed



Gene-Level Filtering: Lowly-expressed genes

Genes with low expression can also be removed as they are:

- Unlikely to provide statistically/biologically significant differential gene expression
- Unlikely to provide clustering information for the cell types

We filtered these out by removing genes that were expressed in less than 10 cells per sample



Normalization and Scaling

After filtering, counts were **normalized** using a regular negative binomial model:

- Based off sequencing depth (# of transcripts per cell)
- Log transformed

Then were **scaled** using a linear transformation:

- Shift the expression of each gene to have a mean of 0 across cells and a variance of 1 across cells
- This removes the impact of highly expressed genes washing out the signal of lowly expressed genes and limits variation from technical noise

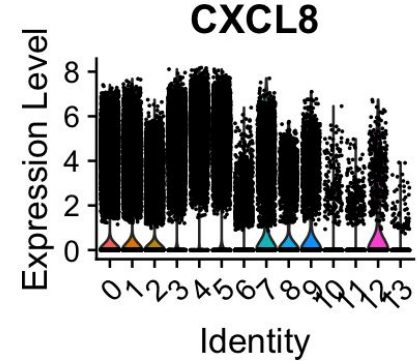
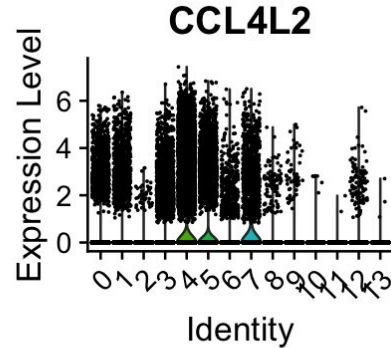
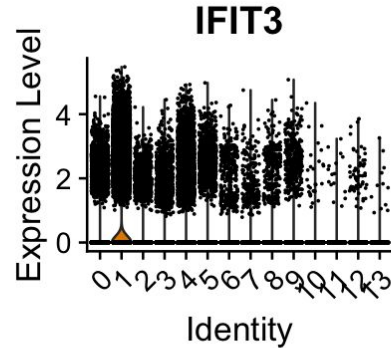
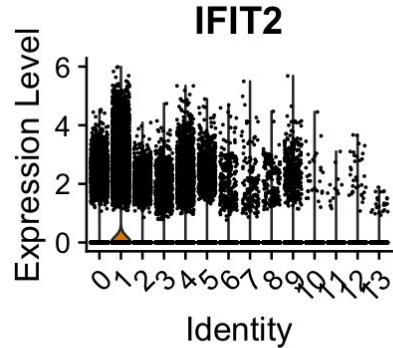


Finding differentially expressed features (cluster biomarkers)

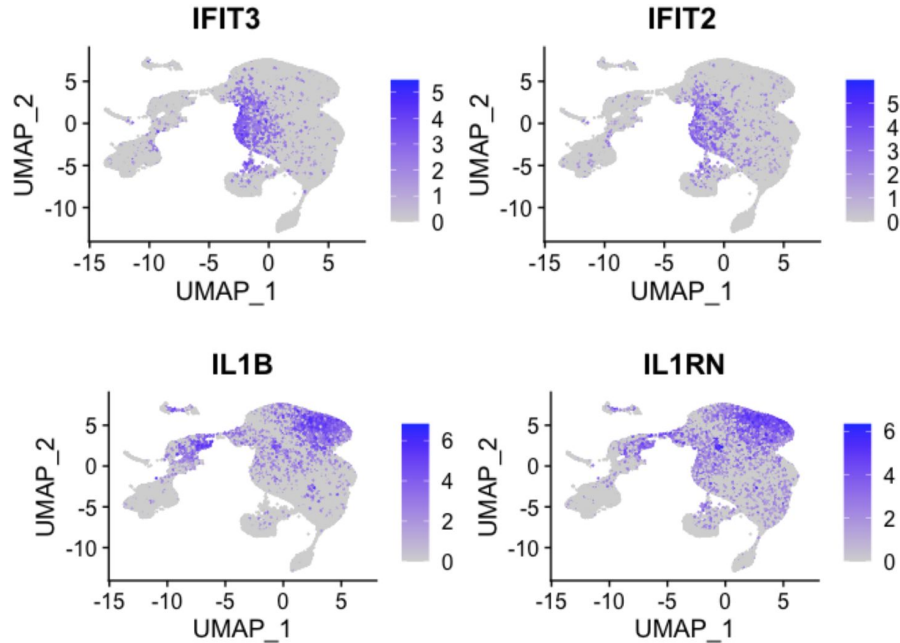
- Find markers for every cluster compared to all remaining cells
- 5149 biomarkers were identified across 13 clusters



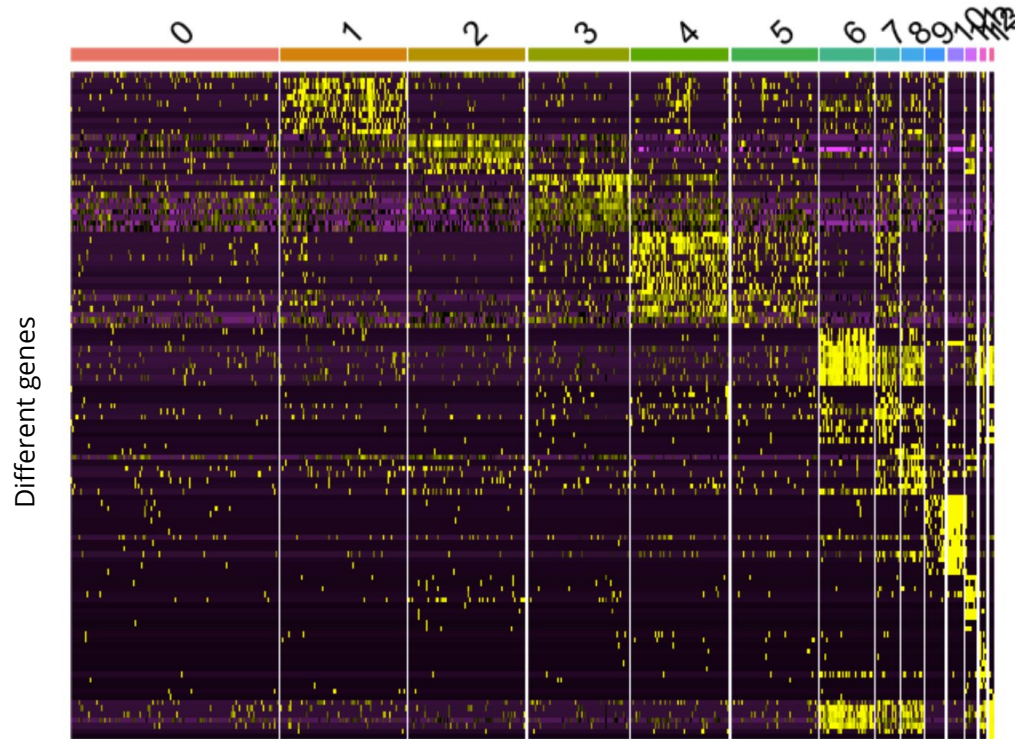
Expression probability distributions across clusters



Visualize gene expression

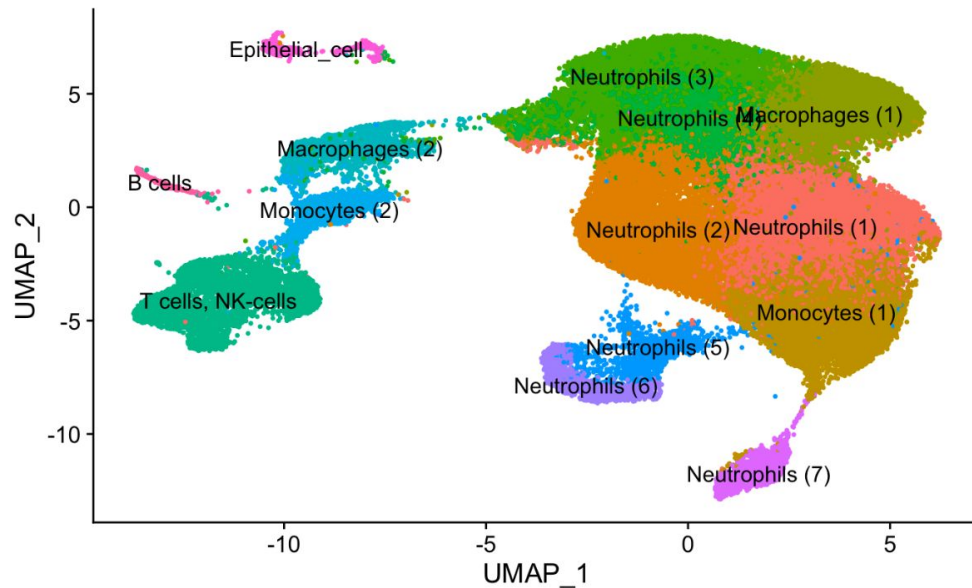


Expression heatmap for given cells and features



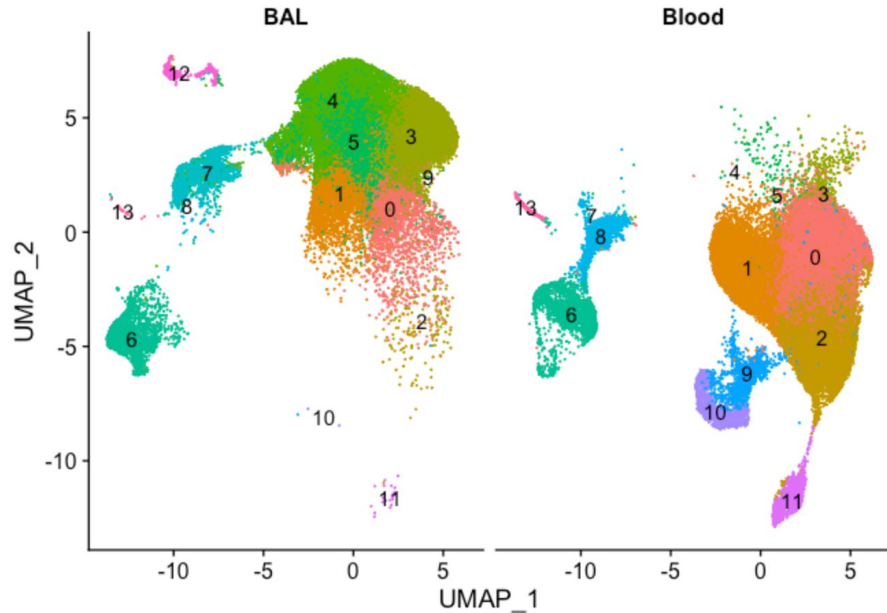
Assigning cell type identity to clusters

Cluster ID	Cell Type
0	Neutrophils (1)
1	Neutrophils (2)
2	Monocytes (1)
3	Macrophages (1)
4	Neutrophils (3)
5	Neutrophils (4)
6	T cells, NK-cells
7	Macrophages (2)
8	Monocytes (2)
9	Neutrophils (5)
10	Neutrophils (6)
11	Neutrophils (7)
12	Epithelial cells
13	B cells

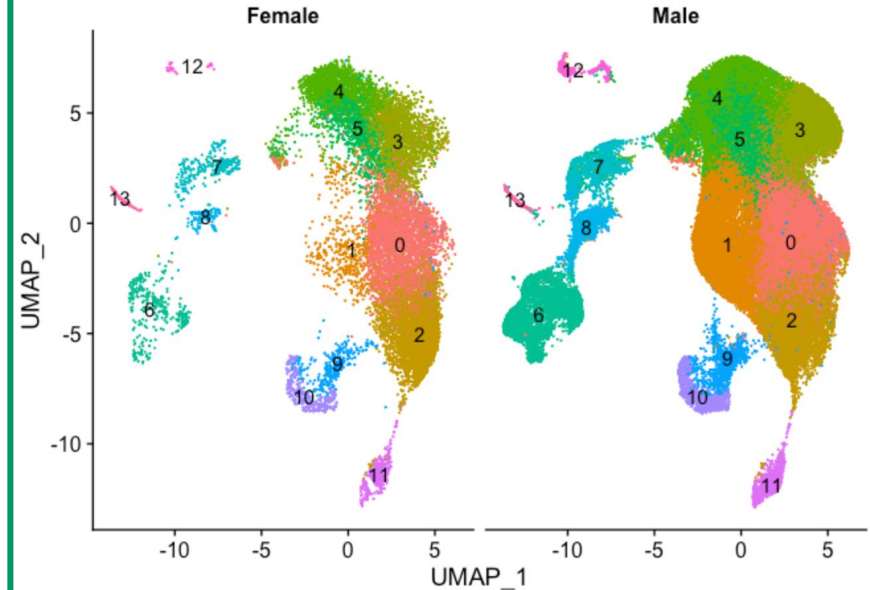


Cell-Type Clustering

- Different cell types present in different tissues



- Similar cell types seen in different sexes (bias towards male with cell depth)



Identifying differentially expressed genes across COVID severity

Identify differentially expressed genes between male and female patients across **COVID severity type (severe vs. mild. healthy)** and within cell types.

Clinic status: “Severe COVID”, “Mild COVID”, “Healthy control”

Cell type: “T_cells_NK_cells”, “B_cells”



Identifying differentially expressed genes across tissue type

Identify differentially expressed genes between male and female patients across **tissue type** (**lavage fluid vs. blood**) and within cell types.

Tissue type: “BAL”, “Blood”

Cell type: “T_cells_NK_cells”, “B_cells”



Clinic status: Severe Covid

```
[r]
get_DE_genes(COVID_severity = "Severe COVID", tissue_type = NA, cell_type = "T_cells_NK_cells")
```

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
XCL2	0.07326542	-0.2678863	0.099	0.120	1
HSPA1A	0.10776260	0.6376632	0.156	0.178	1
PLCG2	0.27900111	0.3068325	0.225	0.208	1
S100A8	0.38752105	-0.3026388	0.106	0.097	1
CCL4L2	0.60627892	-0.4404375	0.125	0.131	1
CXCL8	0.87181007	0.2514535	0.271	0.276	1

6 rows

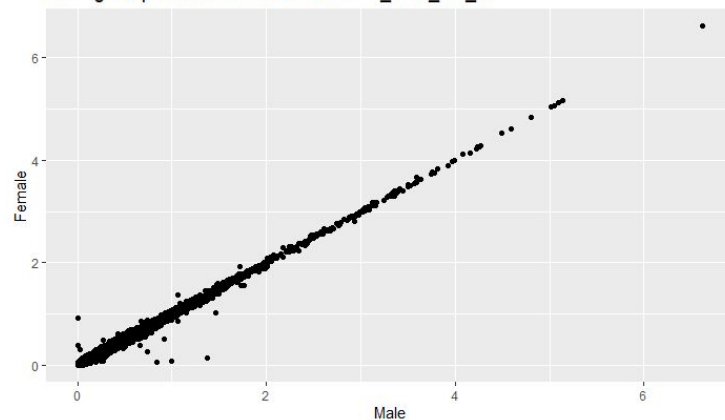
```
[r]
get_DE_genes(COVID_severity = "Severe COVID", tissue_type = NA, cell_type = "B_cells")
```

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
RLIM	0.0003237115	0.6233978	0.200	0.000	1
SLC25A5	0.0013358515	-0.6913124	0.333	0.644	1
GTF2H5	0.0017815939	-0.5469001	0.000	0.153	1
GBP4	0.0021245587	0.6011261	0.150	0.000	1
CHMP4B	0.0021764795	0.5230262	0.250	0.051	1
MPRIIP	0.0023689293	0.5419930	0.283	0.068	1
HARS	0.0033635147	-0.4115142	0.000	0.136	1
TTC7A	0.0037524275	-0.5357617	0.017	0.169	1
SNRPC	0.0038195849	-0.7454215	0.083	0.288	1
RBMX2	0.0038809497	-0.4987613	0.017	0.169	1

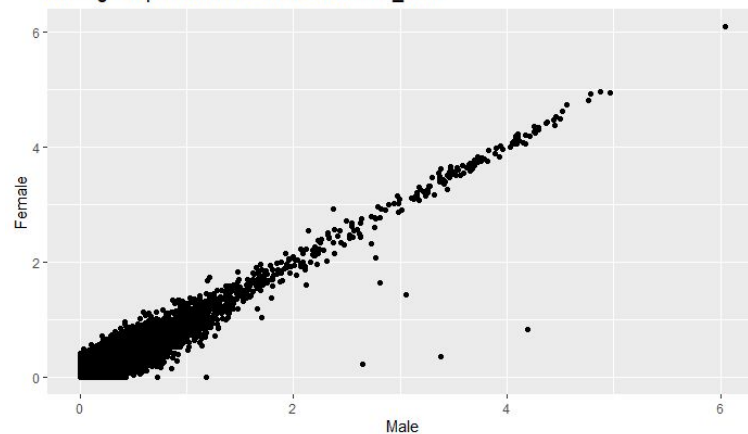
1-10 of 15 rows

Previous 1 2 Next

Average Expression of Severe COVID T_cells_NK_cells



Average Expression of Severe COVID B_cells



Clinic status: Mild Covid

```
{r}
get_DE_genes(COVID_severity = "Mild COVID", tissue_type = NA, cell_type = "T_cells_NK_cells")
```

	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
PRDX2	0.0007831340	0.3099341	0.158	0.092	1
SECISBP2	0.0009473272	0.3708796	0.128	0.069	1
BTG1	0.0009494570	-0.2599985	0.678	0.741	1
C9orf78	0.0021917360	0.4762247	0.303	0.229	1
MYL6	0.0025021442	0.3365664	0.593	0.514	1
CTorf56	0.0028176204	-0.4316946	0.166	0.234	1
ANAPC16	0.0033879804	0.2738744	0.232	0.161	1
TBCB	0.0036591108	-0.3733270	0.080	0.132	1
FAM32A	0.0042556857	-0.3811977	0.057	0.102	1
FGFR1OP2	0.0043827811	-0.3789195	0.076	0.127	1

1-10 of 15 rows

Previous 1 2 Next

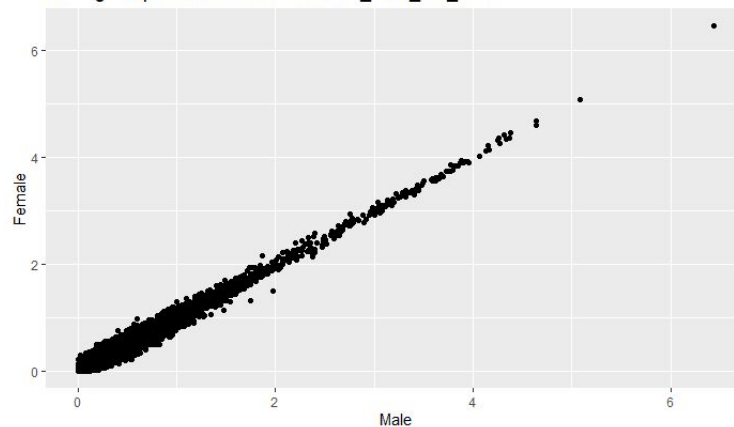
```
{r}
get_DE_genes(COVID_severity = "Mild COVID", tissue_type = NA, cell_type = "B_cells")
```

	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
DNAJC21	3.405102e-05	0.9330482	0.477	0.070	0.670805
TMED5	9.244106e-04	-0.9471448	0.023	0.279	1.000000
ALG5	1.512698e-03	-0.6946345	0.000	0.209	1.000000
EMP3	1.572057e-03	-0.9141623	0.409	0.721	1.000000
SIK3	1.815888e-03	0.9308566	0.386	0.116	1.000000
DHX29	1.880373e-03	1.0671006	0.250	0.023	1.000000
RHBDF2	2.597366e-03	-0.5720724	0.045	0.302	1.000000
RICTOR	2.667464e-03	1.2391335	0.455	0.186	1.000000
TCF3	2.907652e-03	0.8977589	0.364	0.093	1.000000
QKI	3.080770e-03	0.9121023	0.341	0.093	1.000000

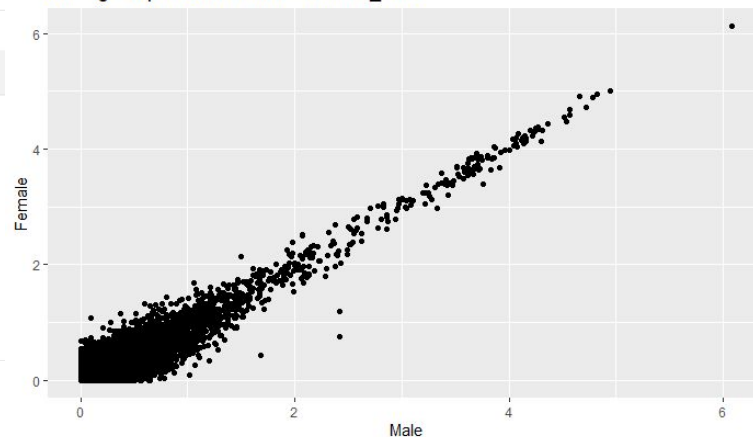
1-10 of 15 rows

Previous 1 2 Next

Average Expression of Mild COVID T_cells_NK_cells



Average Expression of Mild COVID B_cells



Clinic status: Healthy Patient

```
{r}
get_DE_genes(COVID_severity = "Healthy control", tissue_type = NA, cell_type = "T_cells_NK_cells")
```

	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
FAM78A	0.002758873	-0.2558314	0.077	0.130	1
TLK2	0.031744841	0.2601118	0.160	0.117	1
CCL4	0.090295920	-0.2951792	0.477	0.536	1
OTULIN	0.099212386	-0.2979508	0.189	0.224	1
FOSL2	0.151086177	-0.2541826	0.326	0.356	1
PTGDS	0.159161024	-0.4583331	0.107	0.133	1
CCDC88C	0.286281267	-0.2781645	0.397	0.413	1
CXCL8	0.304992585	-0.3981069	0.121	0.144	1
FBXW5	0.346441568	-0.2989194	0.254	0.270	1
GZMK	0.474925317	-0.2662136	0.181	0.192	1

1-10 of 14 rows

Previous 1 2 Next

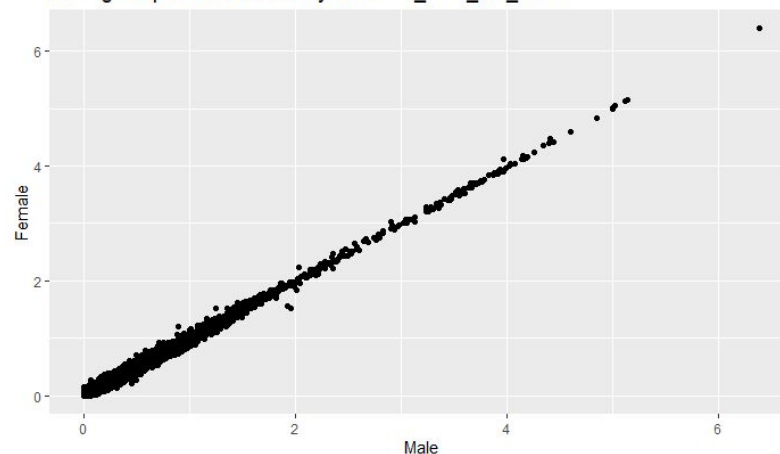
```
{r}
get_DE_genes(COVID_severity = "Healthy control", tissue_type = NA, cell_type = "B_cells")
```

	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
SHC1	0.0003113618	-0.3143295	0.025	0.130	1
DFFA	0.0004455782	0.3086889	0.105	0.012	1
HSPB1	0.0005888497	0.7690106	0.204	0.074	1
UBE2W	0.0006487893	-0.3169690	0.025	0.123	1
THRA	0.0011821624	-0.2784393	0.019	0.105	1
TMEM123	0.0017676534	-0.3955096	0.605	0.741	1
SPTAN1	0.0018814081	-0.4049537	0.117	0.247	1
BLOC1S4	0.0020363836	-0.2748959	0.160	0.309	1
ZFP36L1	0.0023086794	-0.3404585	0.494	0.636	1
DRAP1	0.0023461200	0.3898420	0.451	0.290	1

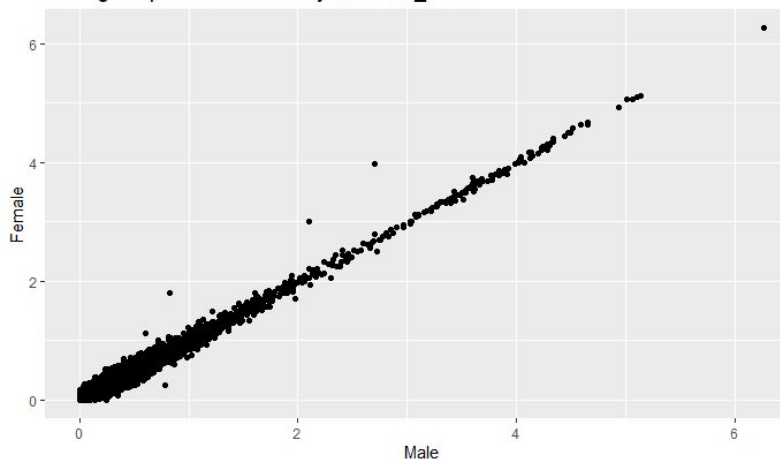
1-10 of 15 rows

Previous 1 2 Next

Average Expression of Healthy control T_cells_NK_cells



Average Expression of Healthy control B_cells



Tissue type: BAL

```
get_DE_genes(COVID_severity = NA, tissue_type = "BAL", cell_type = "T_cells_NK_cells")
```

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
XCL2	0.09492618	-0.3073118	0.098	0.118	1
HSPA1A	0.10467073	0.6628469	0.165	0.189	1
CCL4L2	0.72106753	-0.4454339	0.130	0.135	1
CXCL8	0.88115072	0.2963193	0.284	0.290	1

4 rows

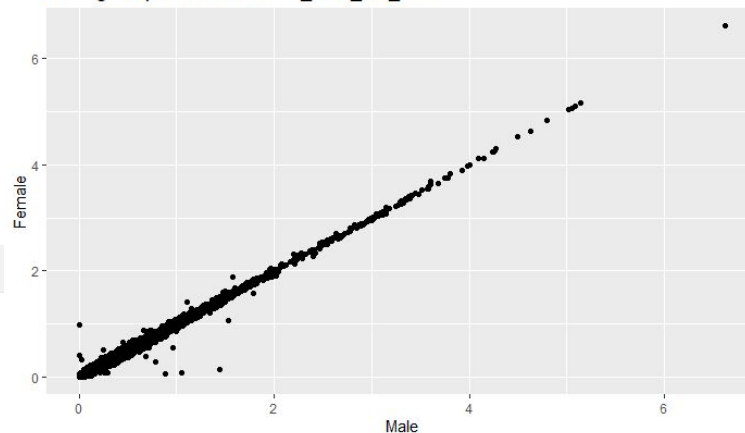
```
get_DE_genes(COVID_severity = NA, tissue_type = "BAL", cell_type = "B_cells")
```

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
IDS	0.01204181	1.8122268	0.667	0.111	1
POLD2	0.01395012	-2.0818360	0.000	0.556	1
NUP98	0.01395012	-1.6379338	0.000	0.556	1
ETS1	0.01395012	-1.5964805	0.000	0.556	1
GPR183	0.01395012	-2.3181879	0.000	0.556	1
ARPP19	0.01395012	1.4170432	0.556	0.000	1
FCRL2	0.01395012	1.9451903	0.556	0.000	1
MT-ND4	0.01511703	-0.7393468	0.889	0.889	1
NDUFA4	0.01715132	1.3138166	0.889	0.444	1
STK17A	0.01835339	-1.5427822	0.222	0.778	1

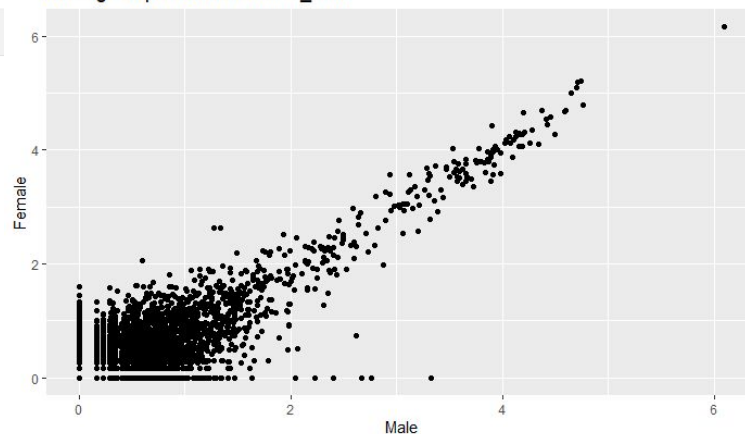
1-10 of 15 rows

Previous 1 2 Next

Average Expression of BAL T_cells_NK_cells



Average Expression of BAL B_cells



Tissue type: Blood

```
{r}  
get_DE_genes(COVID_severity = NA, tissue_type = "Blood", cell_type = "T_cells_NK_cells")
```

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
PTGDS	0.003399829	-0.5474796	0.110	0.148	1
PLCG2	0.008270798	-0.3034187	0.211	0.169	1
NUP210	0.014403389	-0.2795832	0.136	0.168	1
CCL4	0.046129635	-0.3190844	0.337	0.374	1
LYZ	0.050000962	0.3397268	0.160	0.134	1
GZMK	0.152683304	-0.2546166	0.114	0.130	1

6 rows

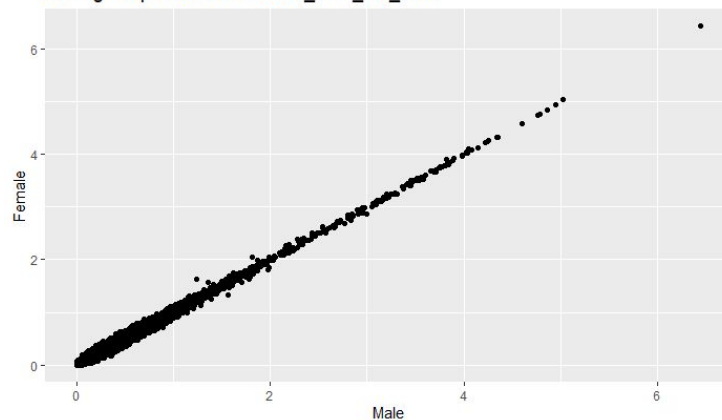
```
{r}  
get_DE_genes(COVID_severity = NA, tissue_type = "Blood", cell_type = "B_cells")
```

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
CEP57	0.0002512441	-0.4095586	0.105	0.227	1
GSPT1	0.0011950166	0.3124394	0.215	0.109	1
HEXDC	0.0012531821	-0.2774245	0.094	0.195	1
KLF6	0.0023649292	0.4515659	0.730	0.664	1
TGIF2	0.0027639563	-0.3381344	0.078	0.164	1
KANSL1	0.0033806592	0.3165460	0.305	0.195	1
TAP1	0.0037661920	0.2674972	0.184	0.098	1
ATP5C1	0.0048342499	-0.2753629	0.152	0.254	1
RNF38	0.0049046539	-0.2689811	0.066	0.141	1
TMEM219	0.0049749455	0.2644433	0.359	0.254	1

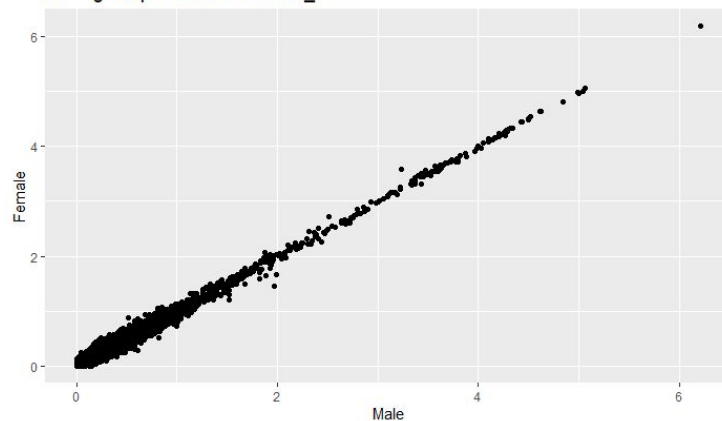
1-10 of 15 rows

Previous 1 2 Next

Average Expression of Blood T_cells_NK_cells



Average Expression of Blood B_cells



Identifying differentially expressed genes for GO Enrichment

- Identify differentially expressed genes between male and female patients across **tissue type (lavage fluid vs. blood)** and **COVID severity type (severe vs. mild. healthy)** for **Neutrophils cell type**.
- **Clinic status:** “Severe COVID”, “Mild COVID”, “Healthy control”
- **Tissue type:** “BAL”, “Blood”
- **Cell type:** “T_cells_NK_cells”, “B_cells”



Differential Expression: Neutrophils Across Clinic Status

```
{r}
get_DE_genes(COVID_severity = "Severe COVID", tissue_type = NA, cell_type = "Neutrophils")
```

R Console

data.frame
4 x 5

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
MT2A	0.02359572	0.5302435	0.107	0.090	1
SERPINB9	0.02961437	-0.2558125	0.089	0.104	1
HBB	0.11423782	0.8645935	0.201	0.219	1
HBA2	0.22441599	0.3979185	0.106	0.116	1

4 rows

```
{r}
get_DE_genes(COVID_severity = "Mild COVID", tissue_type = NA, cell_type = "Neutrophils")
```

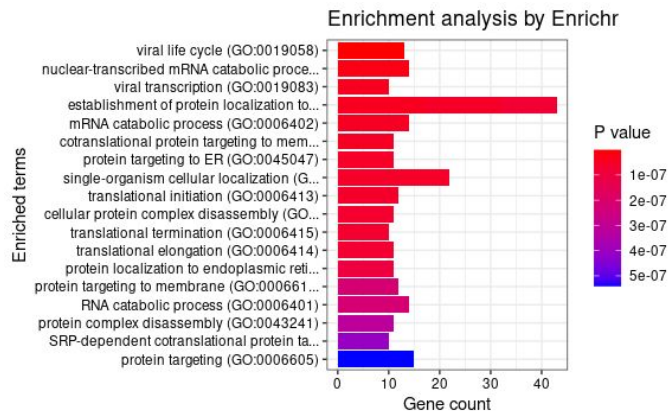
R Console

data.frame
15 x 5

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
TLR4	0.0005003369	-0.6658023	0.102	0.198	1
PPM1F	0.0014030599	-0.5627361	0.096	0.180	1
PPP1CA	0.0015894704	0.5949803	0.171	0.090	1
ERV3-1	0.0028915142	-0.5836980	0.066	0.135	1
PRDX5	0.0039864151	0.3559666	0.267	0.171	1
TLE4	0.0046451474	-0.4433869	0.087	0.159	1
STK4	0.0057567243	-0.4436662	0.375	0.465	1
RPS15	0.0062097078	0.3451678	0.189	0.111	1
PPP3CA	0.0066416166	-0.3754000	0.093	0.165	1
EIF1AY	0.0090601416	0.6253398	0.147	0.084	1

1-10 of 15 rows

Previous 1 2 Next



Differential Expression: Neutrophils Across Clinic Status

```
{r}
get_DE_genes(COVID_severity = "Healthy control", tissue_type = NA, cell_type = "Neutrophils")
```

R Console

data.frame
15 x 5

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
CYTIP	0.0002530143	-0.4967614	0.243	0.349	1
NQO2	0.0011748752	0.5682489	0.110	0.053	1
DDIT3	0.0015019217	-0.3370850	0.139	0.222	1
PELI2	0.0032691982	0.5225801	0.129	0.072	1
C10orf54	0.0036485097	-0.2670204	0.487	0.588	1
PHIP	0.0036636023	-0.3775757	0.135	0.207	1
GGA1	0.0043053089	-0.4460086	0.061	0.114	1
IQSEC1	0.0047316254	-0.4554058	0.152	0.220	1
PYCARD	0.0047971163	-0.3391046	0.148	0.220	1
SH3BGRL3	0.0057518139	0.2680015	0.814	0.763	1

1-10 of 15 rows

Previous 1 2 Next

Differential Expression: Neutrophils Across Tissue Type

```
{r}  
get_DE_genes(COVID_severity = NA, tissue_type = "BAL", cell_type = "Neutrophils")
```

R Console

data.frame
4 x 5

	p_val <dbl>	avg_log2FC <dbl>	pct.1 <dbl>	pct.2 <dbl>	p_val_adj <dbl>
IFNGR2	0.0004144043	0.2702332	0.289	0.239	1
INSIG1	0.0072533704	-0.2928060	0.141	0.173	1
SERPINB9	0.0086938837	-0.3849594	0.135	0.165	1
HBB	0.3872940541	0.6098652	0.117	0.125	1

4 rows

```
{r}  
get_DE_genes(COVID_severity = NA, tissue_type = "Blood", cell_type = "Neutrophils")
```

R Console

data.frame
0 x 3

0 rows

Limitations

- **Batch effect:**

- It is recommended to “integrate” samples together using a KNN-graphical cluster approach to account for batch effect
- This is both time and memory intensive (more than our computers can handle), and we were forced to only merge datasets together without accounting for batch

- **Filtering:**

- All 54 samples were hard-filtered using same parameters. Ideally each sample would be filtered individually with own parameters

- **Dataset**

- Limited number of female patients
- All lavage/lung samples were from severe covid patients; none for control or mild patients



Conclusions and Future Directions

- Differentially expressed genes were identified
 - Between male and females with the same clinic status and cell type, or tissue type and cell type, some differentially expressed genes were identified.
- More data and higher computational power
 - Utilize datasets with larger patient samples sizes, and gain access to higher computational power for sample-sample integration (remove batch effects)

