Long Title: Long non-coding RNAs (lncRNAs) NEAT1 and MALAT1 are differentially expressed in severe COVID-19 patients: An integrated single cell analysis

Short Title: LncRNAs NEAT1 and MALAT1 differentiate inflammation in severe COVID-19 patients

Authors: Kai Huang BS[1,2], Catherine Wang MS[3], Christen Vagts MD[1], Vanitha Raguveer BS[3], Patricia W. Finn MD[1,2,4*], David L. Perkins MD PhD[2,5,6]

[1]Division of Pulmonary, Critical Care, Sleep, and Allergy, Department of Medicine, University of Illinois at Chicago, Chicago IL, 60612 United States

[2]Department of Bioengineering, University of Illinois at Chicago, Chicago IL, 60612 United States

[3]College of Medicine, University of Illinois at Chicago, Chicago IL, 60612 United States

[4] Department of Microbiology and Immunology, University of Illinois at Chicago, Chicago IL, 60612 United States

[5]Division of Nephrology, Department of Medicine, University of Illinois at Chicago, Chicago IL, 60612 United States

[6]Department of Surgery, University of Illinois at Chicago, Chicago IL, 60612 United States

*Corresponding author: pwfinn@uic.edu

Author Contributions:

Conceptualization, K.H.; Methodology, K.H.; Software, K.H.; Investigation, K.H, C.W., and V.R.; Writing – Original Draft, K.H., C.W.; Writing Review & Editing, K.H., C.W., V.R., C.V., D.L.P., and P.W.F; Visualization, K.H.; Supervision, C.V., D.L.P., and P.W.F.; Funding Acquisition, K.H., D.L.P., and P.W.F.

25

Abstract:

Hyperactive and damaging inflammation is a hallmark of severe rather than mild COVID-19 syndrome. To uncover key inflammatory differentiators between severe and mild COVID-19 disease, we applied an unbiased single-cell transcriptomic analysis. We integrated a bronchoalveolar lavage (BAL) dataset with a peripheral blood mononuclear cell dataset (PBMC) and analyzed the combined cell population, focusing on genes associated with disease severity. Distinct cell populations were detected in both BAL and PBMC where the immunomodulatory long non-coding RNAs (lncRNAs) NEAT1 and MALAT1 were highly differentially expressed between mild and severe patients. The detection of other severity associated genes involved in cellular stress response and apoptosis regulation suggests that the pro-inflammatory functions of these lncRNAs may foster cell stress and damage. The lncRNAs NEAT1 and MALAT1 are potential components of immune dysregulation in COVID-19 that may provide targets for severity related diagnostic measures or therapy.

42 **Introduction:**

43 The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) pandemic continues around the
44 world,(1,2) but the underlying pathophysiology of coronavirus disease 2019 (COVID-19) is ill-defined.
45 Symptoms and progression of COVID-19 vary widely(3) as some patients may be asymptomatic while
46 others exhibit disease with varying severity.(4,5) Common symptoms include fever, cough, and fatigue,
47 which generally appear 2 to 14 days after exposure,(6,7) while rarer symptoms include dyspnea,
48 headache/dizziness, nausea, diarrhea, and hemoptysis.(8) Severe cases of COVID-19 are distinguished by
49 strong inflammatory responses that can lead to multiorgan damage and death.(9) The mechanisms that
50 separate mild and severe disease remain poorly understood.

51 After viral exposure, the inflammatory response to COVID-19 commences with signaling
52 cascades that lead to secretion of type I interferons, cytokines and chemokines.(10) This initial exposure
53 also activates inflammasomes, multimeric protein complexes that play an important role in triggering
54 inflammation and the subsequent initiation of an adaptive immune response.(11–13) The Nod-like
55 receptor pyrin domain-containing 3 (NLRP3) inflammasome is a major cause of cytokine storm
56 associated the with clinical manifestations of severe COVID-19 disease.(14) Furthermore, coronavirus
57 viroporin proteins activate the NLRP3 inflammasome which regulates the secretion of IL-1β and IL-
58 18.(15) Pyroptosis, a programmed cell death pathway that leads to immune cell depletion, is also
59 regulated by activation of the NLRP3 inflammasome and is an important mechanism of viral
60 pathogenesis in both SARS-CoV-2 and SARS-CoV.(16–18) These studies suggest that investigation of
61 inflammasome regulation may elucidate understanding of COVID-19 disease pathophysiology.

62 Single-cell studies of COVID-19 patients have found dysregulated immune compartments in the
63 respiratory tract as well as peripheral blood.(19–24) However, it is challenging to directly compare
64 results across studies in different tissues due to differences in cell cluster identification between
65 physiological compartments. We postulated that simultaneous analysis of severe versus mild COVID-19

66 patients across respiratory and peripheral immune compartments using integrated clustering would
67 uncover overall effectors of immune dysregulation in the COVID-19 immune response. To achieve this
68 goal, we integrated single-cell datasets, one from bronchoalveolar lavage (BAL) and one from peripheral
69 blood mononuclear cells (PBMC),(19,20) in order to examine disease transcriptomics across severities as
70 well as between local and peripheral cellular environments. We utilized an unbiased analytical strategy
71 that was agnostic to specific gene functions and focused on genes with severity dependent expression
72 across different cell types. Taken together, we uncovered genes contributing to the dysregulated COVID-
73 19 immune response prominent in severe relative to mild disease. Moreover, we identified cell types
74 where these inflammatory regulators manifest in local and peripheral compartments.
75

76 **Methods:**

77 *Dataset preprocessing and integration*

78 We selected publicly available single-cell datasets with patient severity metrics and ample
79 sequencing depth to pass our quality filters for integration into our combined dataset. The raw count
80 matrices for BAL cells and PBMC cells were downloaded from the NCBI Gene Expression Omnibus
81 (accession number GSE145926) and the COVID-19 Cell Atlas (https:/www.covid19cellatlas.org/#wilk20),
82 respectively. Patients who were mechanically ventilated or had PaO_2/FiO_2 ≤ 300 mmHg indicating
83 hypoxemia consistent with acute respiratory distress syndrome (ARDS)(25) were designated as severe
84 patients while all others were considered to have mild disease (**Table S1**). The BAL dataset contained
85 three healthy controls, while the PBMC dataset contained six (**Table S2**). Both datasets were
86 preprocessed using the R program Seurat.(26) Briefly, cells were filtered to only include cells with
87 unique molecular identifier (UMI) counts greater than 1000, gene count between 200 and 6000, and less
88 than 10% of genes mapping to mitochondrial genes. The function SCTransform from the Seurat package
89 was applied to each dataset separately to regress out technical variability as well as the percentage of
90 mitochondrial gene expression.(27) Transformed BAL and PBMC datasets were integrated with 3000
91 integration features and 50 integration anchors as recommended in Seurat.(28) We found that the "M3"
92 mild patient sample from the BAL dataset contained only 369 total cells, while every other patient
93 sample for BAL or PBMC had at least 1200 cells. The M3 sample was removed before differential
94 expression analysis to avoid skewing results due to extremely low cell counts.

95

96 *Clustering and Identification*

97 The integrated dataset was dimension reduced using principal component analysis (PCA) and
98 clustered with a resolution set to 0.5 and including the top 30 principle components. The clustering was

visualized using uniform manifold approximation and projection (UMAP).(29) The raw integrated dataset was normalized by applying SCTransform to the full integrated dataset. This normalized count matrix is utilized for all subsequent analysis. Marker genes for each cluster were computed using the FindAllMarkers function with the Model-based Analysis of Single-cell Transcriptomics algorithm (MAST) for differential expression with UMI count as a latent variable.(30,31) Cluster markers were then inspected and labeled according to known cell markers(Supplemental file 1).(32,33) A second round of clustering with a resolution of 1 was then conducted to further classify subtypes of identified cells. Clusters with fewer than 300 cells were reassigned to larger clusters using Seurat integration label transfer. Cluster identities were scored and verified using a signature matrix generated from flow cytometry sorted RNA-seq data of immune cells.(34) Plots of cell clusters and key cell type markers were generated using Seurat's plotting functions.

Cell Proportions

Cell types were tallied for each sample, and the percentage abundance of each cell type was calculated. Cell proportions for healthy controls, mild patients, and severe patients were compared using a two-sided pairwise test of equal proportion with false discovery rate (FDR) p-value adjustment. The resulting proportions were plotted using the ggplot2 R package.(35)

Differential Expression

For each cell type, differentially expressed genes (DEGs) were calculated separately for BAL and PBMC cells using MAST with UMI count as a latent variable. To support MAST differential gene expression analysis between three sample groups, the Seurat built-in differential expression function FindMarkers was modified to generating iterations of the hurdle model corresponding to each set of

122 two compared conditions. Genes from each cell type that were differentially expressed across all three
123 comparisons between healthy controls, mild, and severe patients were tallied. For BAL, only genes with
124 FDR adjusted p-values less than 1e-7 across all comparisons were considered (17.4% of all DEGs in BAL),
125 while PBMC DEGs were considered if all FDR adjusted p-values across comparisons for that gene were
126 less than 0.05 (16.4% of all DEGs in PBMCs). The difference in p-value threshold was used to filter a
127 similar proportion of genes from BAL and PBMC data due to the smaller number of DEGs detected in
128 PBMCs (200-400 in PBMCs, 1000+ in BAL). These highly differentially expressed genes were further
129 filtered by removing genes that were not differentially expressed across all conditions in at least 5
130 different cell types (1/3 of our total cell types). The lists of PBMC and BAL highly differentially expressed
131 genes were then combined, removing duplicates.

132 Some of the genes identified were highly cell type specific. These genes also had the highest
133 residual variance. Since these genes represent intrinsic cell type differences rather than biologically
134 interesting differentially expressed genes, they were removed. To determine this filter threshold, the
135 residual variance of the top 100 variable genes were plotted in decreasing order to determine the
136 "elbow" point where the variance stops decreasing at a rapid rate.(**Figure S1**) This resulted in the 21
137 most variable genes being removed. We termed the 50 remaining differentially expressed genes
138 recurrent differentially expressed genes (rDEGs) since they were found in multiple cell types and showed
139 differential expression between patients and healthy controls as well as between severities. The rDEG
140 expression data was exported to Monocle 3 to generate modules for gene ontology (GO) enrichment
141 analysis. We generated 4 modules from the 50 rDEGs using the find_gene_modules function in Monocle
142 3 with 30 principle components and a resolution of 0.8.(36–38) Differential module expression was
143 calculated using ANOVA using aggregated module expression levels and processed with a tukey posthoc
144 test. Module gene ontology enrichment was computed topGO with default settings.(39) Module and
145 gene level plots were generated using the R packages ggplot2, ComplexHeatmap, and Circlize.(35,40,41)

146

147 *Validation*

148 We compared rDEG trends from our analysis with two additional COVID-19 datasets, one with
149 nasopharyngeal data to compare against the local inflammatory environment of our BAL data, and one
150 in PBMCs to compare against the peripheral environment in our PBMC data.(22,23) Cell types from our
151 analysis were transferred onto the validation datasets using Seurat to identify the validation cells for
152 comparison. Each validation dataset was filtered and preprocessed separately using the same
153 parameters as the main dataset. After preprocessing and label transfer, DEGs were generated
154 independently for each validation set for our transferred cell types. We compared the rDEGs we focused
155 on in the manuscript with DEG results from each validation set, noting whether the same DEG was
156 detected and whether the direction of change was the same.

157

158 **Results:**

159 *Integrated PBMC and BAL analysis identified 26 clusters consolidated into 15 cell types.*

160 After quality filtering, we recovered 100,739 single cell transcriptomes. From these, we
161 recovered 26 cell clusters from Seurat. (**Figure S2**). Since we identified cell types from the integrated
162 dataset containing both PBMC and BAL cells, we were able to examine how each of our cell clusters
163 behaves across both physiological compartments. The clusters did not aggregate based on sample type
164 or patient condition (**Figure S2**), indicating successful integration clustering of the two datasets.

165 From the 26 clusters, 11 were identified as monocyte/macrophage (Mo/Ma) clusters. Since our
166 clusters contain both monocytes from the PBMC sample as well as macrophages from the BAL, we
167 designated them as MoMa clusters. Six of the MoMa clusters showed classically M1 associated
168 transcriptomes with increased expression of VCAN, FCN1 and CD14 expression.(42,43) These clusters

169 also expressed other pro-inflammatory factors such as S100A8, CCL2, CCL3, CCL7, and CCL8.(44,45)
170 Three other MoMa clusters showed M2 polarization with increased FN1 expression along with
171 decreased VCAN and FCN1 expression. These M2 MoMa clusters also expressed Th2 associated
172 inflammatory factors such as MRC1 and CCL18.(45) All MoMa clusters expressed FCGR3A (CD16a).(42)
173 Two additional clusters were labeled as intermediate MoMa because they did not show distinct
174 transcriptomes corresponding to either M1 or M2 groups. One intermediate MoMa cluster
175 overexpressed MALAT1, while the other overexpressed metallothionein proteins including MT1F and
176 MT1G.

177 We also identified two clusters of CD4+ T cells (CD4 and IL7R), one cluster of T regulatory cells
178 (IL2RA and LAG3),(46,47) and three clusters of CD8+ T cells (CD8A). Two of the CD8+ T cell clusters were
179 labeled as CD8+ memory cells due to their high CCL5 and GZMH expression.(48) Other identified
180 immune cell clusters include natural killer (NK) cells (SPON2 and NCAM1), neutrophils (NAMPT),(34)
181 naïve B cells (MS4A1), plasmablasts (IGJ and MZB1), plasmacytoid dendritic cells (IRF8 and PLD4),(49,50)
182 and myeloid dendritic cells (CD1C)and LGALS2).(34,51) In addition to immune cell types, we also found
183 two epithelial clusters. One contained a mixture of epithelial and granulocyte markers including KRT19
184 and SLPI(34,52) while the other also contained the additional markers PPIL6 and CFAP300 for
185 pneumocytes and ciliary cells, respectively.(53,54)

186 The 26 clusters were consolidated into 15 cell types (**Figure S2)** to streamline further analysis by
187 combining clusters that are not distinguishable when examining their canonical marker expression
188 levels. This consolidation also prevents cell groups with many clusters such as the M1 MoMa group from
189 overshadowing those with fewer clusters in our subsequent differential gene expression ranking
190 analysis. Cells within each cluster were compared against their original identifications from both their
191 respective dataset, and most clusters identifications were consistent. The one exception was the

192 intermediate MoMa type, which was predominantly composed of macrophages from BAL, but also
193 contained a mixture of monocytes, CD4+ and CD8+ T cell identifications from PBMCs (**Table S3)**.
194
195 *Proinflammatory cell types are enriched in severe COVID-19 patients.*
196 Cell type proportions within BAL and PBMC sample groups showed distinct differences between
197 patients and healthy controls as well as between mild and severe patients (**Figure 1**). In BAL samples of
198 patients with severe disease, proinflammatory cells such as M1 MoMa and neutrophils showed
199 increased abundance (p<<<.01, **Table S4**), while immunoregulatory cell types including M2, intermediate
200 MoMa, and Tregs were less abundant (p<<<.01). BAL samples of patients with mild disease showed
201 decreased abundance of M1 MoMa and neutrophils and increased Tregs and CD8+ memory T cells
202 compared to healthy controls and severe patients.
203 **Fig 1: Severe patients show increased proportions of proinflammatory cell types. A.** Overall average abundance of each major
204 cell type for all cells. **B.** Per patient abundance of all major cell types for all cells. **C.** Per cohort (bronchoaveolar lavage (BAL) and
205 peripheral blood mononuclear cells (PBMC)) and per condition (healthy, mild, severe) abundance for each cell type. Conditions
206 that are significant versus their respective controls are labeled with a triangle (p<.05). Conditions that are significant between
207 severe and both mild as well as healthy controls are labeled with a star (p<.05). Conditions that are significant between severe
208 and mild, but not between severe and healthy controls are labeled with a diamond (p<.05).
209
210 In PBMCs, the trends in M1 MoMa and neutrophils are reversed. Tregs and CD8+ memory T cells
211 are less abundant in PBMCs of mild patients. These opposing patterns may illustrate heavy recruitment
212 of the cell types abundant in BAL, resulting in depletion in the PBMCs that results in an increase in
213 relative abundance of non-recruited cells in PBMCs. Mild patients also showed an increase of
214 intermediate MoMa in PBMCs, reinforcing the pattern of relative increases in abundance of
215 immunoregulatory cell types in mild patients in both BAL and PBMC compartments.

216

217 *Recurrent DEG (rDEG) modules highlight key pathways in COVID-19 immune response.*

218 We identified an average of 1158 DEGs per cell type for BAL samples, and 260 DEGs per cell type
219 for PBMC samples.(**Table S5,S6**) After filtering, we identified 50 rDEGs across our 15 cell types that
220 formed 4 distinct modules (**Figure 2**). Module 1 showed significant GO enrichment for developmental
221 processes (p<.05) but did not show differential expression between conditions. Module 2 showed
222 significance for viral defense and Type I interferon GO terms. Most genes in this module were interferon
223 induced genes including the first three IFIT family genes, ISG15, CXCL10, and MX1. This module was
224 significantly overexpressed in BAL of all patients versus healthy controls (p<.01).

225

226

227 **Fig 2: rDEGs grouped into four distinct modules with immune regulation enriched GO terms. A.** Heatmap of modules
228 generated from the recurrent differentially expressed geens (rDEGs). Triangles indicate significance (p<.05) versus healthy
229 control within the sample cohort; diamonds indicate significance between severe and mild patients (p<.05). **B.** Module
230 membership for each module. Modules 1 and 4 contain a mixture of metabolic and immune response related genes. Module 2
231 contains genes related to interferon activated viral defense. Module 3 contains other inflammatory regulation genes and stress
232 response genes (generated using the Circlize R package). **C.** Per module GO term enrichment showing the top enriched terms
233 for each module and their respective p-values with the red line indicating -log10(0.05). The first three modules contain
234 inflammation related terms in their most enriched terms, while module 4 only contains metabolism related terms.

235

236 Module 3 was enriched for macromolecule synthesis and cellular processes. This module
237 includes the immunomodulatory lncRNAs NEAT1 and MALAT1.(55,56) It also includes MTRNR2L12, an
238 anti-apoptotic lncRNA, and NFKBIA which is an NF-κB inhibitor. The module was significantly
239 underexpressed in BAL of mild patients versus healthy controls, and it was overexpressed in BAL of
240 severe patients versus mild patients. Module 4 had significant terms related to negative regulation of

241 metabolic processes. This module included the NUPR1 stress response gene, and CSTB which is an
242 inhibitor of cathepsins like CTSL and CTSB that are involved in COVID-19 viral entry.(57) Module 4 was
243 significantly underexpressed in BAL of severe patients versus healthy controls.

244

245 *Stress response, apoptosis, and viral entry related genes show severity dependent expression.*

246 We analyzed individual rDEGs in each of our five most abundant cell types: M1 MoMa, M2
247 MoMa, CD4+ T cells, NK cells, and CD8+ memory T cells (**Figure 3**). This analysis confirmed previous
248 reports of downregulation of HLA genes(20,58) such as HLA-DRA and HLA-DRB5 in COVID-19 patients,
249 with severe patients showing the most downregulation. We also saw upregulation of interferon related
250 genes including MX1, and IFIT1-3. This increase was greatest in mild patients, correlating with previous
251 findings of immune exhaustion (**Table S7**).(59,60) Further examination showed additional severity
252 dependent patterns of differential expression of transcripts related to the stress response, cell death,
253 and viral entry in cell types involved in the viral immune response.

254 **Fig 3: rDEG expression in most abundant cell types highlights differential immune regulation between mild and severe**
255 **patients in both BAL and PBMC cohorts. A.** Heatmaps visualizing rDEGs within each of the top five most abundant cell types in
256 our dataset (generated using the ComplexHeatmap R package). For each cell type, the full rDEG list was filtered via the same p-
257 values (p<10e-7 for BAL, p<.05 for PBMC) and only rDEGs that are differentially expressed below these thresholds for either BAL
258 or PBMC are included in the plot. Expression levels are normalized separately for each cohort. The first sidebar indicates which
259 cohort the particular gene passed the rDEG threshold for, while the second sidebar indicates the ratio of expression of the
260 particular gene between BAL and PBMC with green (positive values) indicating higher expression levels detected in BAL. **B.**
261 Visualization of select rDEGs representing pathways outside of the main interferon activated gene group that are relevant to
262 disease. These genes are visualized separately for each cohort and condition using the sample UMAP projection of cell types
263 from Figure 1. Each gene shows cell type, cohort, and condition specific differences in localization across the dataset.

264

265 The NF-κB inhibitor NFKBIA was upregulated in all five most abundant cell types within the BAL
266 of severe patients compared to healthy controls and mild groups. In PBMCs of severe patients, NFKBIA
267 was downregulated compared to healthy controls and mild patients except in CD8+ memory T cells. This
268 pattern of localized overexpression in BAL may indicate increased NFKBIA activity in response to local
269 hyperactivity of NF-κB. Furthermore, the stress response gene NUPR1, whose downregulation leads to
270 cell death, was downregulated in M1 and M2 MoMa in the BAL of severe patients and upregulated in
271 mild patients, indicating a pro-apoptotic shift in severe patient MoMa clusters. NUPR1 was
272 downregulated in BAL of both mild and severe patients for NK cells, CD4+ T cells, and CD8+ Memory T
273 cells.

274 Mild and severe patients also had variable expression of two anti-apoptotic genes, the BCL2
275 inhibitor BCL2A1 and the lncRNA MTRNR2L12. BCL2A1 was significantly upregulated in BAL of severe
276 patients over healthy controls and mild groups for M1 and M2 MoMa, NK cells, and CD4+ T cells. Mild
277 patients showed downregulation of BCL2A1 versus healthy controls in NK and CD4+ T cells. Additionally,
278 MTRNR2L12 was upregulated in BAL of both mild and severe patients in M1 and M2 MoMa, NK cells,
279 CD4+ T cells and CD8+ Memory T cells. The upregulation of these anti-apoptotic genes shows a
280 defensive response to apoptotic cell stresses, particularly in BAL.

281 CTSL, which is a critical protein in the viral entry pathway for COVID-19, was upregulated in BAL
282 of severe patients in M1 and M2 MoMa in mild patients and healthy controls. This suggests a faster viral
283 entry pathway in severe patients, which may contribute to the formation of a hyperinflammatory
284 response. In BAL of NK, CD4+ T cells, and CD8+ Memory T cells, CTSL was downregulated in mild patients
285 and upregulated in severe patients. CTSB, also implicated in viral entry, showed similar patterns.

286

287 *NEAT1 and MALAT1 are differential regulators of inflammation in severe COVID-19.*

288 The pro-inflammatory lncRNA NEAT1 passed our rDEG threshold in BAL samples for nine
289 different cell types, more than any other gene in our analysis. These cell types include M1, M2 and
290 intermediate MoMa, NK cells, CD4+ T cells, CD8+ memory T cells, naïve B cells, myeloid dendritic cells,
291 and epithelium/basal cells (**Figure 4**). NEAT1 is localized to the site of infection and inflammation since it
292 is not differentially expressed in PBMCs. Additionally, among rDEGs, it has one of the highest averages in
293 log2-fold change between severe and mild patients (**Figure 4**). NEAT1 is overexpressed in BAL of severe
294 patients and underexpressed in mild patients. The epithelial/basal cell group is the exception where mild
295 groups also show NEAT1 overexpression over healthy controls, but expression is still significantly higher
296 in severe patients versus mild patients.
297

298

299 **Fig 4: lncRNAs NEAT1 and MALAT1 are strongly differentially expressed between severe and mild patients and represent key**
300 **inflammatory regulators in BAL and PBMC respectively. A.** Violin plots showing overall expression level density across patient
301 conditions in the entire dataset. Even at the full dataset scale, these distributions show that NEAT1 is overexpressed in BAL of
302 severe patients while MALAT1 is underexpressed in PBMCs of severe patients. **B.** Frequency of detection across cell types for
303 rDEGs shows NEAT1 as the most detected rDEG in BAL, with MALAT1 tied for second among rDEGs in PBMC. The top log2-fold
304 change of rDEGs in severe versus mild patients also shows NEAT1 and MALAT1 among the rDEGs with the highest absolute
305 change between severe and mild conditions. **C.** Visualization of NEAT1 and MALAT1 via UMAP projection shows more cell type
306 localized expression in NEAT1. It is also clearly underexpressed in mild BAL cases. MALAT1 also shows a more subtle but
307 significant underexpression in severe patient PBMCs.

308

309 Another immunomodulatory lncRNA, MALAT1, was the second most frequent rDEG in PBMCs. It
310 passed our rDEG threshold in 6 cell types (tied with ISG15) and 3 cell types in BAL. In BAL derived M1
311 and M2 MoMa, MALAT1 was underexpressed in mild patients compared to both healthy controls and
312 severe patients. In CD4+ T cells, MALAT1 shows consistent overexpression in mild patients and
313 underexpression in severe patients. In PBMCs, MALAT1 was underexpressed in severe patients versus
314 both healthy controls and mild patients in M1, M2 and intermediate MoMa, NK cells, plasmablasts, and
315 epithelial/basal cells.

316

317 *Validation of rDEG expression patterns.*

318 We projected the cell type classifications in our analysis via Seurat's label transfer feature to two
319 other COVID-19 datasets, one with nasopharyngeal samples(22) and one with PBMC samples(23).
320 Comparison of cell labels from the original datasets versus our transferred labels shows general
321 agreement among the cell IDs, allowing us to use our cell type labels for direct comparison of rDEG
322 patterns in the validation data (**Figure S3&S4**). We compared rDEG expression patterns of the genes

323 analyzed in the previous two sections in the same five cell types. Among each gene's statistically
324 significant changes in our analysis between healthy, mild, and severe cases, 36.3% were significant in
325 our validation cohorts. However, more than two thirds of the non-significant findings in validation were
326 from comparisons with healthy control in the nasopharyngeal dataset. This is likely due to the small
327 number of cells recovered from these controls after filtering. Only 1148 cells were recovered from
328 control samples after filtering compared to 35715 and 25546 for moderate and severe cases
329 respectively. Among the findings that were significant in both our data and in validation, we found that
330 79% were significant in the same direction. Notably, NEAT1 showed 100% agreement in validation of
331 BAL, while MALAT1 showed 100% agreement in validation of PBMCs (**Table S8,S9 & S10**).

332 **Discussion:**

333 Our analysis of BAL and PBMC single cell data in COVID-19 patients has elucidated key
334 differences between mild and severe disease. We were able to combine cells from both PBMC and BAL
335 in an integrated analysis. Although our intermediate MoMa group had a mixed group of PBMC cells, our
336 overall identifications were consistent across both datasets. Furthermore, the cells in the intermediate
337 MoMa group consisted of cells with weak expression of a wide range of canonical markers. These cells
338 may be intermediate immune cells from different lineages that share a similar transcriptomic profile. By
339 conducting analysis simultaneously on cells from the local infection site in the lung as well as the
340 peripheral immune system, we contrast how the disease manifests and interacts across both
341 compartments. We have identified differentially expressed genes that vary with severity, are highly
342 differentially expressed across multiple cell types, and represent key functions related to the
343 hyperinflammatory disease state. NEAT1 was the most widely differentially expressed gene across cell
344 types within BAL; it also exhibited a high log-fold change that correlated with disease severity. The
345 ubiquity of NEAT1, its specific localization to BAL cells, and pro-inflammatory functions suggests that it
346 may be a key mediator of the inflammation seen in severe COVID-19. NEAT1 is a well characterized

347 activator of the NLRP3 inflammasome, as well as NLRC4 and AIM2 inflammasomes, which in turn
348 amplify the inflammatory response.(55) However, an overactive immune response contributes to lasting
349 tissue damage in severe COVID-19 disease. Intense inflammation through activation of the NLRP3
350 inflammasome can also lead to pyroptosis, driven by the upregulation of NEAT1.(55,61) These highly
351 inflammatory and damaging effects of NEAT1 illustrate how overexpression in severe patients might
352 lead to the inflammatory tissue damage seen in severe COVID-19.

353 MALAT1 also exerts various immunological effects including the mediation of NLRP3
354 inflammasome activation.(62,63) MALAT1 has been linked to M1-like activity in macrophages,
355 promoting inflammation.(64) Our finding that MALAT1 is overexpressed in BAL MoMa of severe versus
356 mild patients suggests that it might be involved in precipitating a shift towards M1 macrophages that
357 exacerbates inflammation. This is further supported by our findings that severe patients show expansion
358 of M1 macrophages and decrease of M2 and intermediate macrophages in BAL, while mild patients
359 show decrease of M1 macrophages. Furthermore, MALAT1 was overexpressed in CD4+ T cells of mild
360 patients. This is also reflective of MALAT1's protective role in T cells. Loss of MALAT1 expression has
361 been shown to push T cells towards the inflammatory Th1 and Th17 phenotype while also decreasing
362 Treg differentiation.(65) This function matches our observed increase in abundance of Tregs in mild
363 patients. Thus, the upregulation of MALAT1 we see in mild patients may be contributing to the more
364 subdued immune response observed in these patients.

365 The severity dependent differential expression of other genes in our analysis provides further
366 evidence of increased cellular stress reflective of a NEAT1 and MALAT1 enhanced hyperinflammatory
367 state. NF-κB is induced in COVID-19 infection.(66) Although we did not detect differential activity of NF-
368 κB directly, we found upregulation of its inhibitor NFKBIA in BAL of severe patients which suggests a
369 feedback response to strong NF-κB activity. NFKBIA's downregulation in PBMCs of severe patients may
370 be due to localization of cells expressing NFKBIA to the site of infection in attempts to regulate the

371 hyperactive inflammatory state.(67) The upregulation of BCL2A1 and MTRNR2L12 is also indicative of
372 extensive cellular stress.(68,69) While MTRNR2L12 is upregulated in both mild and severe disease,
373 BCL2A1 is upregulated exclusively in severe disease. The increased activity of these anti-apoptotic genes,
374 particularly in BAL of severe patients, shows additional evidence of the cellular stress induced by
375 infection and inflammation. These genes may be responding to pyroptosis pathways triggered by
376 inflammasome activation via NEAT1 and MALAT1. Further evidence of inflammatory cell damage is seen
377 in the downregulation of NUPR1 in BAL of M1 and M2 macrophages of severe patients with
378 upregulation in mild patients. Downregulation of this stress response gene has been shown to cause
379 mitochondrial dysfunction and ROS production that can lead to cell death.(70) Lastly, our observation
380 that CTSL, a protein crucial for COVID-19 viral entry is upregulated across multiple cell types in severe
381 patients provides a potential initial mechanism for the induction of the NEAT1 and MALAT1 mediated
382 inflammatory state through increased efficiency of viral entry.(57)

383 Limitations in our study include the small sample size, variable clinical presentation and
384 treatment. Additionally, time from presentation to sample collection varied across patients. The
385 stratification of patients as severe or mild may also introduce unknown factors due to patient variability
386 in presentation and classification. Although our validation shows promising reproduction of expression
387 patterns, additional studies with more subjects and stringent recruiting and sample collection would
388 further elucidate these findings.

389 We have demonstrated a clear ensemble of differential gene activity associated with severe
390 disease in COVID-19 infection that revolves around the lncRNAs NEAT1 and MALAT1. Their specific
391 activity changes in severe patients coupled with inflammasome promoting functions, suggest important
392 roles in the COVID-19 hyperinflammatory process. These findings indicate that NEAT1 and MALAT1 may
393 be candidates for treatment targeting or biological marker exploration.

394

395 **References**

396 1. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. RESEARCH High Contagiousness
397 and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis
398 [Internet]. 2020 Jul 1 [cited 2020 Oct 14];26(7):1470–7. Available from:
399 /pmc/articles/PMC7323562/?report=abstract

400 2. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates
401 the rapid dissemination of novel coronavirus (SARS-CoV-2). Science (80-) [Internet]. 2020 May 1
402 [cited 2020 Oct 14];368(6490):489–93. Available from: http://science.sciencemag.org/

403 3. García LF. Immune Response, Inflammation, and the Clinical Spectrum of COVID-19 [Internet].
404 Vol. 11, Frontiers in Immunology. Frontiers Media S.A.; 2020 [cited 2020 Oct 14]. Available from:
405 https://pubmed.ncbi.nlm.nih.gov/32612615/

406 4. Vetter P, Eckerle I, Kaiser L. Covid-19: A puzzle with many missing pieces [Internet]. Vol. 368, The
407 BMJ. BMJ Publishing Group; 2020 [cited 2020 Oct 14]. Available from:
408 https://twitter.com/WHO/status/

409 5. Wang F, Qu M, Zhou X, Zhao K, Lai C, Tang Q, et al. The timeline and risk factors of clinical
410 progression of COVID-19 in Shenzhen, China. J Transl Med [Internet]. 2020 Jul 3 [cited 2020 Oct
411 14];18(1):270. Available from: https://translational-
412 medicine.biomedcentral.com/articles/10.1186/s12967-020-02423-8

413 6. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019
414 in China. N Engl J Med [Internet]. 2020 Apr 30 [cited 2020 Oct 14];382(18):1708–20. Available
415 from: http://www.nejm.org/doi/10.1056/NEJMoa2002032

416 7. Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A, et al. Clinical Characteristics of Covid-19
417 in New York City. N Engl J Med [Internet]. 2020 Jun 11 [cited 2020 Oct 14];382(24):2372–4.

418 Available from: http://www.nejm.org/doi/10.1056/NEJMc2010419

419 8. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation
420 and intervention [Internet]. Vol. 20, Nature Reviews Immunology. Nature Research; 2020 [cited
421 2020 Oct 14]. p. 363–74. Available from: /pmc/articles/PMC7187672/?report=abstract

422 9. Zaim S, Chong JH, Sankaranarayanan V, Harky A. COVID-19 and Multiorgan Response [Internet].
423 Vol. 45, Current Problems in Cardiology. Mosby Inc.; 2020 [cited 2020 Oct 14]. Available from:
424 /pmc/articles/PMC7187881/?report=abstract

425 10. van den Berg DF, te Velde AA. Severe COVID-19: NLRP3 Inflammasome Dysregulated. Front
426 Immunol [Internet]. 2020 Jun 26 [cited 2020 Oct 14];11:1580. Available from:
427 https://www.frontiersin.org/article/10.3389/fimmu.2020.01580/full

428 11. Evavold CL, Kagan JC. How Inflammasomes Inform Adaptive Immunity [Internet]. Vol. 430,
429 Journal of Molecular Biology. Academic Press; 2018 [cited 2020 Oct 14]. p. 217–37. Available
430 from: /pmc/articles/PMC5766381/?report=abstract

431 12. Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, et al. Coronavirus infections and immune responses
432 [Internet]. Vol. 92, Journal of Medical Virology. John Wiley and Sons Inc.; 2020 [cited 2020 Oct
433 14]. p. 424–32. Available from: https://pubmed.ncbi.nlm.nih.gov/31981224/

434 13. Guo H, Callaway JB, Ting JPY. Inflammasomes: Mechanism of action, role in disease, and
435 therapeutics [Internet]. Vol. 21, Nature Medicine. Nature Publishing Group; 2015 [cited 2020 Oct
436 14]. p. 677–87. Available from: /pmc/articles/PMC4519035/?report=abstract

437 14. Freeman TL, Swartz TH. Targeting the NLRP3 Inflammasome in Severe COVID-19 [Internet]. Vol.
438 11, Frontiers in Immunology. Frontiers Media S.A.; 2020 [cited 2020 Sep 25]. p. 1518. Available
439 from: www.frontiersin.org

440 15. Farag NS, Breitinger U, Breitinger HG, El Azizi MA. Viroporins and inflammasomes: A key to

441 understand virus-induced inflammation [Internet]. Vol. 122, International Journal of Biochemistry
442 and Cell Biology. Elsevier Ltd; 2020 [cited 2021 Mar 2]. p. 105738. Available from:
443 /pmc/articles/PMC7102644/

444 16. Yap JKY, Moriyama M, Iwasaki A. Inflammasomes and Pyroptosis as Therapeutic Targets for
445 COVID-19. J Immunol [Internet]. 2020 Jul 15 [cited 2020 Oct 14];205(2):307–12. Available from:
446 http://www.jimmunol.org/content/205/2/307

447 17. Yue Y, Nabar NR, Shi CS, Kamenyeva O, Xiao X, Hwang IY, et al. SARS-Coronavirus Open Reading
448 Frame-3a drives multimodal necrotic cell death. Cell Death Dis [Internet]. 2018 Sep 1 [cited 2020
449 Oct 14];9(9):1–15. Available from: https://www.nature.com/articles/s41419-018-0917-y

450 18. Bergsbaken T, Fink SL, Cookson BT. Pyroptosis: Host cell death and inflammation [Internet]. Vol.
451 7, Nature Reviews Microbiology. NIH Public Access; 2009 [cited 2020 Sep 25]. p. 99–109.
452 Available from: /pmc/articles/PMC2910423/?report=abstract

453 19. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune
454 cells in patients with COVID-19. Nat Med [Internet]. 2020 May 12 [cited 2020 May 15];1–3.
455 Available from: http://www.nature.com/articles/s41591-020-0901-9

456 20. Wilk AJ, Rustagi A, Zhao NQ. A single-cell atlas of the peripheral immune response in patients
457 with severe COVID-19. Nat Med [Internet]. 2020; Available from:
458 https://doi.org/10.1038/s41591-020-0944-y

459 21. Zhang JY, Wang XM, Xing X, Xu Z, Zhang C, Song JW, et al. Single-cell landscape of immunological
460 responses in patients with COVID-19. Nat Immunol [Internet]. 2020 Sep 1 [cited 2020 Sep
461 25];21(9):1107–18. Available from: https://doi.org/10.1038/s41590-020-0762-x

462 22. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, et al. COVID-19 severity correlates
463 with airway epithelium–immune cell interactions identified by single-cell analysis. Nat Biotechnol

464 [Internet]. 2020 Aug 1 [cited 2020 Sep 25];38(8):970–9. Available from:
465 https://doi.org/10.1038/s41587-020-0602-4

466 23. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Severe COVID-19
467 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell [Internet]. 2020 Sep 17 [cited 2020
468 Sep 25];182(6):1419. Available from: /pmc/articles/PMC7405822/?report=abstract

469 24. Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, et al. Immunophenotyping of covid-19 and
470 influenza highlights the role of type i interferons in development of severe covid-19. Sci Immunol
471 [Internet]. 2020 Jul 1 [cited 2020 Oct 14];5(49):1554. Available from:
472 http://immunology.sciencemag.org/

473 25. Diaz J V, Baller A, Banerjee A, Bertagnolio S, Bonet M, Bosman A, et al. Clinical management of
474 COVID-19: interim guidance. 2020;WHO/2019-nCoV/clinical/2020.5. Available from:
475 https://www.who.int/publications/i/item/clinical-management-of-covid-19

476 26. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive
477 Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21.

478 27. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using
479 regularized negative binomial regression. Genome Biol [Internet]. 2019 Dec 23 [cited 2020 Jun
480 8];20(1):296. Available from:
481 https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1874-1

482 28. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data
483 across different conditions, technologies, and species. Nat Biotechnol [Internet]. 2018 Apr 2
484 [cited 2018 Jul 25];36(5):411–20. Available from:
485 http://www.nature.com/doifinder/10.1038/nbt.4096

486 29. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and

487 Projection. J Open Source Softw [Internet]. 2018 Sep 2 [cited 2020 Oct 14];3(29):861. Available
488 from: http://joss.theoj.org/papers/10.21105/joss.00861

489 30. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: A flexible statistical
490 framework for assessing transcriptional changes and characterizing heterogeneity in single-cell
491 RNA sequencing data. Genome Biol [Internet]. 2015 Dec 10 [cited 2020 Jun 8];16(1):278.
492 Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0844-5

493 31. Soneson C, Robinson MD. Bias, robustness and scalability in single-cell differential expression
494 analysis. Nat Publ Gr. 2018;15.

495 32. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: A manually curated resource of
496 cell markers in human and mouse. Nucleic Acids Res. 2019;47(D1):D721–8.

497 33. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: A web server for exploration of mouse and
498 human single-cell RNA sequencing data. Database [Internet]. 2019 Jan 1 [cited 2020 Oct
499 14];2019(1):46. Available from:
500 https://academic.oup.com/database/article/doi/10.1093/database/baz046/5427041

501 34. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carré C, et al. RNA-Seq Signatures Normalized by
502 mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. Cell Rep
503 [Internet]. 2019 Feb 5 [cited 2019 Apr 4];26(6):1627-1640.e7. Available from:
504 https://www.sciencedirect.com/science/article/pii/S2211124719300592

505 35. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Cham: Springer International
506 Publishing; 2016 [cited 2020 Oct 14]. (Use R!). Available from:
507 http://link.springer.com/10.1007/978-3-319-24277-4

508 36. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional
509 landscape of mammalian organogenesis. Nature [Internet]. 2019 Feb 20 [cited 2019 Mar

510 13];566(7745):496–502. Available from: http://www.nature.com/articles/s41586-019-0969-x

511 37. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves
512 complex single-cell trajectories. Nat Methods [Internet]. 2017;14(10):979–82. Available from:
513 http://dx.doi.org/10.1038/nmeth.4402

514 38. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators
515 of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol
516 [Internet]. 2014 Mar 23 [cited 2020 Oct 14];32(4):381–6. Available from:
517 https://www.nature.com/articles/nbt.2859

518 39. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. Bioconductor. 2020;R
519 package.

520 40. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
521 multidimensional genomic data. Bioinformatics [Internet]. 2016 Sep 15 [cited 2020 Oct
522 16];32(18):2847–9. Available from: https://pubmed.ncbi.nlm.nih.gov/27207943/

523 41. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in
524 R. Bioinformatics [Internet]. 2014 Apr 2 [cited 2020 Oct 16];30(19):2811–2. Available from:
525 https://pubmed.ncbi.nlm.nih.gov/24930139/

526 42. Kapellos TS, Bonaguro L, Gemünd I, Reusch N, Saglam A, Hinkley ER, et al. Human monocyte
527 subsets and phenotypes in major chronic inflammatory diseases [Internet]. Vol. 10, Frontiers in
528 Immunology. Frontiers Media S.A.; 2019 [cited 2020 Aug 19]. p. 2035. Available from:
529 www.frontiersin.org

530 43. Chang MY, Kang I, Gale M, Manicone AM, Kinsella MG, Braun KR, et al. Versican is produced by
531 trif- and type I interferon-dependent signaling in macrophages and contributes to fine control of
532 innate immunity in lungs. Am J Physiol - Lung Cell Mol Physiol [Internet]. 2017 Dec 1 [cited 2020

533 Aug 19];313(6):L1069–86. Available from: /pmc/articles/PMC5814701/?report=abstract

534 44. Wang S, Song R, Wang Z, Jing Z, Wang S, Ma J. S100A8/A9 in inflammation [Internet]. Vol. 9,
535 Frontiers in Immunology. Frontiers Media S.A.; 2018 [cited 2020 Aug 19]. p. 1298. Available from:
536 www.frontiersin.org

537 45. Palomino DC arolin. T, Marti LC avalheir. Chemokines and immunity [Internet]. Vol. 13, Einstein
538 (São Paulo, Brazil). Instituto de Ensino e Pesquisa Albert Einstein; 2015 [cited 2020 Aug 19]. p.
539 469–73. Available from: /pmc/articles/PMC4943798/?report=abstract

540 46. Mohr A, Malhotra R, Mayer G, Gorochov G, Miyara M. Human FOXP3+ T regulatory cell
541 heterogeneity [Internet]. Vol. 7, Clinical and Translational Immunology. Wiley-Blackwell; 2018
542 [cited 2020 Oct 14]. Available from: /pmc/articles/PMC5822410/?report=abstract

543 47. Okamura T, Fujio K, Shibuya M, Sumitomo S, Shoda H, Sakaguchi S, et al. CD4+CD25-LAG3+
544 regulatory T cells controlled by the transcription factor Egr-2. Proc Natl Acad Sci U S A [Internet].
545 2009 Aug 18 [cited 2020 Oct 14];106(33):13974–9. Available from:
546 www.pnas.org/cgi/content/full/

547 48. Weng NP, Araki Y, Subedi K. The molecular basis of the memory T cell response: Differential gene
548 expression and its epigenetic regulation [Internet]. Vol. 12, Nature Reviews Immunology. NIH
549 Public Access; 2012 [cited 2020 Oct 14]. p. 306–15. Available from:
550 /pmc/articles/PMC4686144/?report=abstract

551 49. Sichien D, Scott CL, Martens L, Vanderkerken M, Van Gassen S, Plantinga M, et al. IRF8
552 Transcription Factor Controls Survival and Function of Terminally Differentiated Conventional and
553 Plasmacytoid Dendritic Cells, Respectively. Immunity. 2016 Sep 20;45(3):626–40.

554 50. Gavin AL, Huang D, Huber C, Mårtensson A, Tardif V, Skog PD, et al. PLD3 and PLD4 are single-
555 stranded acid exonucleases that regulate endosomal nucleic-acid sensing. Nat Immunol

556 [Internet]. 2018 Sep 1 [cited 2020 Oct 15];19(9):942–53. Available from:
557 /pmc/articles/PMC6105523/?report=abstract

558 51. Collin M, Bigley V. Human dendritic cell subsets: an update [Internet]. Vol. 154, Immunology.
559 Blackwell Publishing Ltd; 2018 [cited 2020 Oct 15]. p. 3–20. Available from:
560 /pmc/articles/PMC5904714/?report=abstract

561 52. Camper N, Glasgow AMA, Osbourn M, Quinn DJ, Small DM, McLean DT, et al. A secretory
562 leukocyte protease inhibitor variant with improved activity against lung infection. Mucosal
563 Immunol [Internet]. 2016 May 1 [cited 2020 Oct 15];9(3):669–76. Available from:
564 www.nature.com/mi

565 53. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based
566 map of the human proteome. Science (80-) [Internet]. 2015 Jan 23 [cited 2020 Oct
567 15];347(6220):1260419–1260419. Available from:
568 https://www.sciencemag.org/lookup/doi/10.1126/science.1260419

569 54. Zietkiewicz E, Bukowy-Bieryllo Z, Rabiasz A, Daca-Roszak P, Wojda A, Voelkel K, et al. CFAP300:
570 Mutations in slavic patients with primary ciliary dyskinesia and a role in ciliary dynein arms
571 trafficking. Am J Respir Cell Mol Biol [Internet]. 2019 Oct 1 [cited 2020 Oct 15];61(4):400–49.
572 Available from: https://www.atsjournals.org/doi/10.1165/rcmb.2018-0260OC

573 55. Zhang P, Cao L, Zhou R, Yang X, Wu M. The lncRNA Neat1 promotes activation of inflammasomes
574 in macrophages. Nat Commun [Internet]. 2019 Dec 1 [cited 2020 Sep 20];10(1):1–17. Available
575 from: https://doi.org/10.1038/s41467-019-09482-6

576 56. Amodio N, Raimondi L, Juli G, Stamato MA, Caracciolo D, Tagliaferri P, et al. MALAT1: A druggable
577 long non-coding RNA for targeted anti-cancer approaches [Internet]. Vol. 11, Journal of
578 Hematology and Oncology. BioMed Central Ltd.; 2018 [cited 2020 Oct 14]. p. 1–19. Available

579 from: https://doi.org/10.1186/s13045-018-0606-4

580 57. Bittmann S, Weissenstein A, Villalon G, Moschuring-Alieva E, Luchter E. Simultaneous Treatment
581 of COVID-19 With Serine Protease Inhibitor Camostat and/or Cathepsin L Inhibitor? J Clin Med
582 Res [Internet]. 2020 [cited 2020 Oct 14];12(5):320–2. Available from:
583 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7239585/

584 58. Zmijewski JW, Pittet JF. Human Leukocyte Antigen-DR Deficiency and Immunosuppression-
585 Related End-Organ Failure in SARS-CoV2 Infection [Internet]. Anesthesia and Analgesia.
586 Lippincott Williams and Wilkins; 2020 [cited 2020 Oct 14]. p. 989–92. Available from:
587 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7386673/

588 59. Zheng M, Gao Y, Wang G, Song G, Liu S, Sun D, et al. Functional exhaustion of antiviral
589 lymphocytes in COVID-19 patients [Internet]. Vol. 17, Cellular and Molecular Immunology.
590 Springer Nature; 2020 [cited 2020 Oct 14]. p. 533–5. Available from:
591 https://doi.org/10.1038/s41423-020-0402-2

592 60. Mathew D, Giles JR, Baxter AE, Oldridge DA, Greenplate AR, Wu JE, et al. Deep immune profiling
593 of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science (80-)
594 [Internet]. 2020 Sep 4 [cited 2020 Oct 14];369(6508). Available from:
595 https://doi.org/10.1126/science.abc8511

596 61. Zhan J-F, Huang H-W, Huang C, Hu L-L, Xu W-W. Long Non-Coding RNA NEAT1 Regulates
597 Pyroptosis in Diabetic Nephropathy via Mediating the miR-34c/NLRP3 Axis. Kidney Blood Press
598 Res [Internet]. 2020 [cited 2020 Oct 15];45(4):589–602. Available from:
599 https://www.karger.com/Article/FullText/508372

600 62. Menon MP, Hua K-F. The Long Non-coding RNAs: Paramount Regulators of the NLRP3
601 Inflammasome. Front Immunol [Internet]. 2020 Sep 25 [cited 2020 Oct 15];11:569524. Available

602 from: www.frontiersin.org

603 63. Yu S yang, Dong B, Tang L, Zhou S hua. LncRNA MALAT1 sponges miR-133 to promote NLRP3
604 inflammasome expression in ischemia-reperfusion injured heart [Internet]. Vol. 254,
605 International Journal of Cardiology. Elsevier Ireland Ltd; 2018 [cited 2020 Oct 15]. p. 50. Available
606 from: https://pubmed.ncbi.nlm.nih.gov/29407129/

607 64. Cui H, Banerjee S, Guo S, Xie N, Ge J, Jiang D, et al. Long noncoding RNA Malat1 regulates
608 differential activation of macrophages and response to lung injury. JCI insight [Internet]. 2019
609 Feb 21 [cited 2020 Oct 15];4(4). Available from: /pmc/articles/PMC6478413/?report=abstract

610 65. Masoumi F, Ghorbani S, Talebi F, Branton WG, Rajaei S, Power C, et al. Malat1 long noncoding
611 RNA regulates inflammation and leukocyte differentiation in experimental autoimmune
612 encephalomyelitis. J Neuroimmunol [Internet]. 2019 Mar 15 [cited 2020 Sep 20];328:50–9.
613 Available from: https://pubmed.ncbi.nlm.nih.gov/30583215/

614 66. Hirano T, Murakami M. COVID-19: A New Virus, but a Familiar Receptor and Cytokine Release
615 Syndrome. Immunity. 2020 May 19;52(5):731–3.

616 67. Ali S, Hirschfeld AF, Mayer ML, Fortuno ES, Corbett N, Kaplan M, et al. Functional Genetic
617 Variation in NFKBIA and Susceptibility to Childhood Asthma, Bronchiolitis, and
618 Bronchopulmonary Dysplasia . J Immunol [Internet]. 2013 Apr 15 [cited 2020 Oct
619 15];190(8):3949–58. Available from: http://www.jimmunol.org/content/190/8/3949

620 68. Vogler M. BCL2A1: The underdog in the BCL2 family [Internet]. Vol. 19, Cell Death and
621 Differentiation. Nature Publishing Group; 2012 [cited 2020 Oct 15]. p. 67–74. Available from:
622 www.nature.com/cdd

623 69. Du C, Xie H, Zang R, Shen Z, Li H, Chen P, et al. Apoptotic neuron-secreted HN12 inhibits cell
624 apoptosis in Hirschsprung's disease. Int J Nanomedicine [Internet]. 2016 Nov 7 [cited 2020 Oct

625 15];11:5871–81. Available from: /pmc/articles/PMC5106231/?report=abstract

626 70. Santofimia-Castaño P, Lan W, Bintz J, Gayet O, Carrier A, Lomberk G, et al. Inactivation of NUPR1
627 promotes cell death by coupling ER-stress responses with necrosis. Sci Rep [Internet]. 2018 Dec 1
628 [cited 2020 Oct 15];8(1). Available from: /pmc/articles/PMC6242935/?report=abstract

629

630

631 **Supplementary Figures**

632 **S1 Figure: Filtering of genes with extremely high residual variance. A.** A plot of mean expression levels
633 versus residual variance for all genes detected in dataset. Genes with the highest residual variance are
634 nearly all immunoglobulin and ribosome genes. **B.** Plot of 100 genes with highest residual variance in
635 descending order. The rate of variance decrease stabilizes after the 21st gene. **C.** Table of top 21 genes
636 filtered out of downstream analysis due to very high residual variance.

637

638 **S2 Figure: Unbiased clustering of combined BAL and PBMC data reveals common cell types between**
639 **cohorts and sample conditions. A/B.** UMAP plots showing the distribution of all cells in both cohorts
640 within defined cell types. Plot A shows the 15 major cell groups we identified with labels over the cluster
641 centers. Plot B shows the subclusters making up those groups. The major groups are abbreviated as
642 prefixes with a letter suffix indicating subgroup. M1, M2 and Int are the macrophage/monocytes. NK is
643 natural killer cells, E/G is epithelial cells and granulocytes, and E/P/C is epithelia, pneumocyte, and ciliary
644 cells. **C.** Select markers for major cell groups plotted on the same UMAP projection. This illustrates the
645 specificity of these markers for different regions of the plot corresponding to our respective cell types.
646 **D.** Dot plot visualization of top markers utilized for identification of each cell cluster. The size of each dot
647 indicates the percentage of cells with detectable expression of each gene, with color indicating
648 expression level. **E.** Splitting the UMAP by cohort and by severity, with "H" indicating healthy control,
649 "M" indicating mild disease, and "S" indicating severe disease illustrates that cell clusters do not
650 organize according to sample type or patient condition, indicating successful integration of the datasets.

651 **S3 Figure: Transferred cell identities correspond to original cell identities from the nasopharyngeal**
652 **validation set.** Dot plot shows cells from the nasopharyngeal validation dataset, identified by their
653 transferred labels on the Y axis and their original labels on the X axis. Color of each dot indicates the

654 numeric log frequency of cells that fit each corresponding set of labels. The size of each dot represents
655 the percentage of each transferred cell types which is represented by each original cell type. From left to
656 right, the full cell type names from the nasopharyngeal dataset are: non-resident macrophage (nrMa),
657 monocyte-derived macrophage (MoD-Ma), monocyte-derived dendritic cell (moDC), resident
658 macrophage (rMa), cytotoxic T cell (CTL), regulatory T cell (Treg), natural killer T cell (NKT), proliferating
659 natural killer T cell (NKT-p), natural killer cells (NK), neutrophils (Neu), B cells, plasmacytoid dendritic cell
660 (pDC), basal cell, ciliated cell, differentiating ciliated cell, FOXN4+ epithelial cell, ionocyte, interferon
661 responsive cell (IRC), mast cell (MC), epithelial outliers, secretory, differentiating secretory, squamous,
662 and unknown epithelial.

663 **S4 Figure: Transferred cell identities correspond to original cell identities from the PBMC validation**
664 **set.** Dot plot shows cells from the PBMC validation set, identified by their transferred labels on the Y axis
665 and their original labels on the X axis. Color of each dot indicates the numeric log frequency of cells that
666 fit each corresponding set of labels. The size of each dot represents the percentage of each transferred
667 cell types which is represented by each original cell type. This dataset contained several cell types where
668 the same label was applied to more than one subcluster, resulting in numeric suffixes for similar cell
669 types. mDCs correspond to myeloid dendritic cells and pDCs correspond to plasmacytoid dendritic cells.

670

671 **Supplementary Excel Tables**

672 **S1 Table: Key characteristics of patients within each dataset**. Patient information from the BAL and
673 PBMC cohorts used in this analysis. Patients who were intubated or had PaO2/FiO2 ≤ 300 mmHg were
674 classified as severe. In BAL, patient "Mild 3" had only 369 cells recovered after filtering and was only
675 used for initial clustering. Most patients in the BAL cohort. Exact ages were not available in the PBMC
676 cohort. The first patient in this cohort was sampled twice, once while classified as a mild patient, and

677 once after their symptoms worsened and required mechanical ventilation. Several patients in the PBMC
678 cohort received azithromycin, which can have immunomodulatory effects before sample collection.

679 **S2 Table: Demographic characteristics of healthy subjects.** All healthy controls used from both the BAL
680 and PBMC cohorts are listed.

681

682 **S3 Table: Cell identities from original datasets versus new combined identification.** For BAL and
683 PBMCs, each cell's original cell identification and new identification are tabulated for the 26 clusters in
684 the subgroup sheets, and the 15 consolidated cell groups in the coarse sheets.

685 **S4 Table: P-value tables for cell proportion comparisons across each cell type.** Conditions compared
686 are listed in the first two columns, and FDR adjusted p-values are listed in the third column. Each sheet is
687 labeled by cell type.

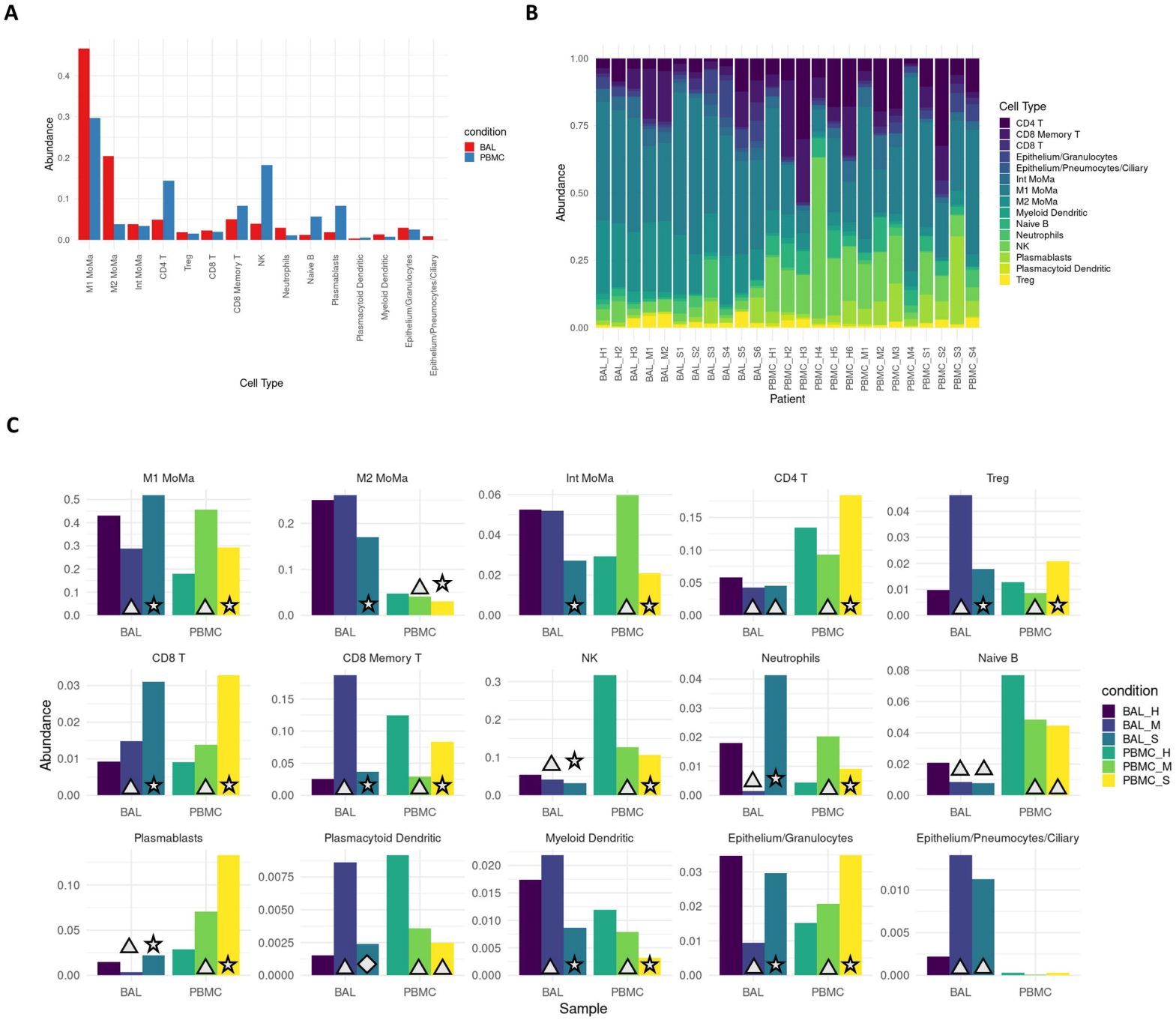
688 **S5 and S6 Tables: DEGs for BAL and PBMC samples respectively, separated by cell type, with raw p-**
689 **values as well as FDR adjusted p-values.** Each sheet is labeled by cell type. Column headings include
690 indicators for which conditions are being compared where applicable. The conditions are numbered:
691 "1=healthy control", "2=mild COVID-19 patient", "3=severe COVID-19 patient". For example, the prefix
692 "g2_1" indicates the comparison of mild patient expression levels minus healthy control expression
693 levels. Log-fold change is reported relative to the natural log. Columns labeled "pct.1", "pct.2", or
694 "pct.3" indicate the percentage of cells in the condition corresponding to that number with detectable
695 expression of a particular gene.

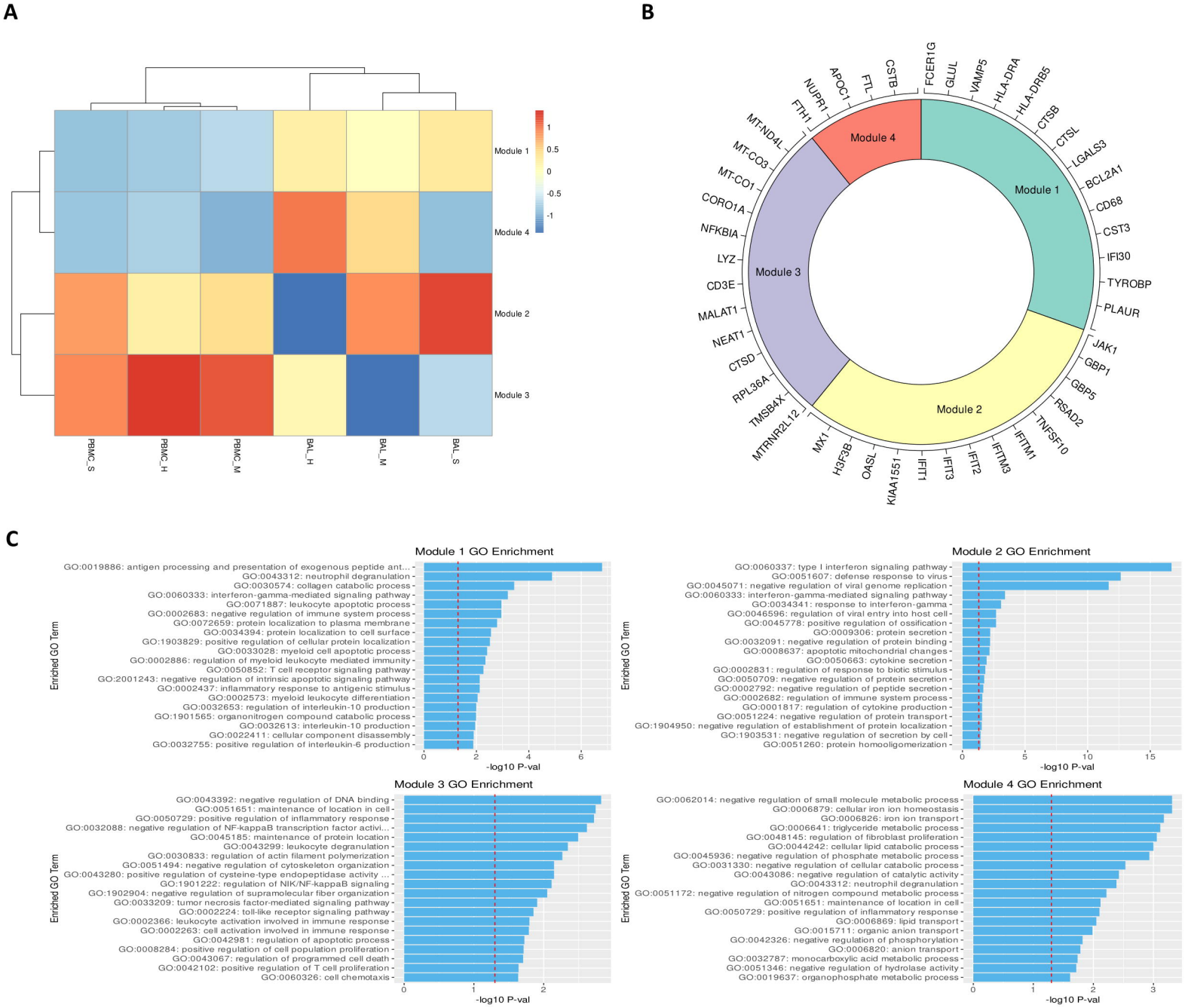
696 **S7 Table: rDEGs with their expression levels in each cell type where each rDEG's adjusted p-values**
697 **passed the p-value filter as defined in our Methods section.** Column headings include indicators for
698 which conditions are being compared where applicable. The conditions are numbered: "1=healthy

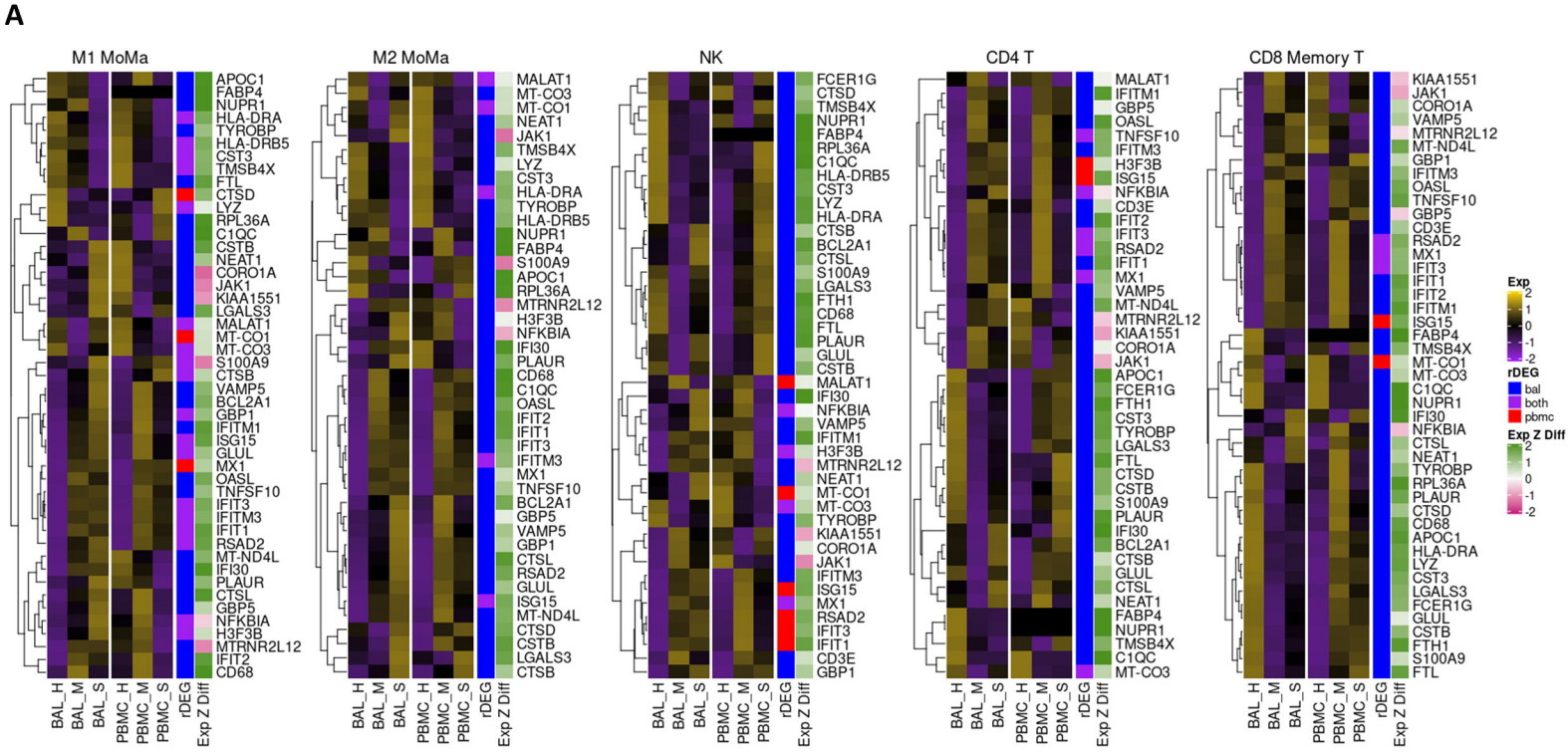
control", "2=mild COVID-19 patient", "3=severe COVID-19 patient". For example, the prefix "g2_1" indicates the comparison of mild patient expression levels minus healthy control expression levels. Log-fold change is reported relative to the natural log. Columns labeled "pct.1", "pct.2", or "pct.3" indicate the percentage of cells in the condition corresponding to that number with detectable expression of a particular gene. The "celltype" and "sample" columns indicate which cell type and which sample condition the rDEG passed filter in.

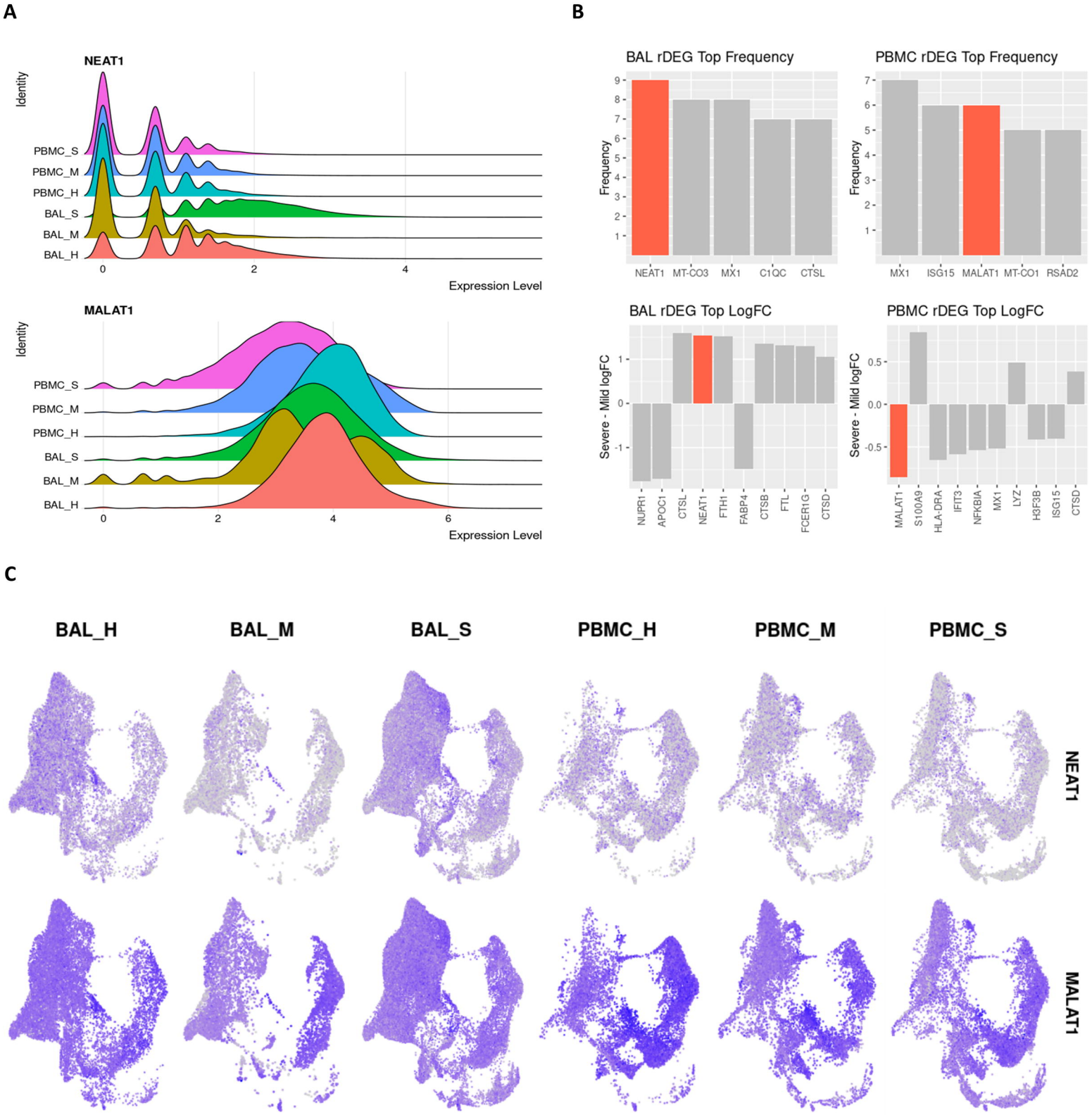
S8 Table: Comparisons between differentially expressed rDEGs discussed in our results shows strong agreement in validation datasets. A/B. Tables representing the tally of differential expression results of our discussed rDEGs which agreed or disagreed between analysis and validation groups based on the direction of detected differential expression. Tables are split by BAL/nasopharyngeal and PBMC groups. The first three columns correspond to cases where a comparison is not available due to a lack of differential expression detected in the original analysis (na.orig), the validation set (na.val), or both (na.all). The top half of each table reports the results for severe vs mild cases only (SvsM) while the bottom half reports results for all three comparisons: healthy vs mild, healthy vs severe, and severe vs mild.

S9 and S10 Tables: Per gene level validation tables for BAL and PBMC groups respectively. Each sheet name corresponds to the cell type presented. Row names indicate the gene being compared, and column names indicate the cohorts being compared: healthy (H), mild COVID (M), severe COVID (S). When differential expression is detected in both the original analysis and validation, the column is labled agree if the change occurred in the same direction and disagree if it is opposite. Other labels indicate where a comparison is not available due to a lack of differential expression detected in the original analysis (na.orig), the validation set (na.val), or both (na.all).


A
Abundance
Cell Type
condition
BAL
PBMC
M1 MoMa
M2 MoMa
Int MoMa
CD4 T
Treg
CD8 T
CD8 Memory T
NK
Neutrophils
Naive B
Plasmablasts
Plasmacytoid Dendritic
Myeloid Dendritic
Epithelium/Granulocytes
Epithelium/Pneumocytes/Ciliary
B
Abundance
Patient
Cell Type
CD4 T
CD8 Memory T
CD8 T
Epithelium/Granulocytes
Epithelium/Pneumocytes/Ciliary
Int MoMa
M1 MoMa
M2 MoMa
Myeloid Dendritic
Naive B
Neutrophils
NK
Plasmablasts
Plasmacytoid Dendritic
Treg
C
Abundance
Sample
condition
BAL_H
BAL_M
BAL_S
PBMC_H
PBMC_M
PBMC_S


A
Module 1
Module 4
Module 2
Module 3
PBMC_S
PBMC_H
PBMC_M
BAL_H
BAL_M
BAL_S
1
0.5
0
-0.5
-1
B
Module 1
Module 2
Module 3
Module 4
FCER1G
GLUL
VAMP5
HLA-DRA
HLA-DRB5
CTSB
CTSL
LGALS3
BCL2A1
CD68
CST3
IFI30
TYROBP
PLAUR
JAK1
GBP1
GBP5
RSAD2
TNFSF10
IFITM1
IFITM3
IFIT2
IFIT3
IFIT1
KIAA1551
OASL
H3F3B
MX1
MTRNR2L12
TMSB4X
RPL36A
CTSD
NEAT1
MALAT1
CD3E
LYZ
NFKBIA
CORO1A
MT-CO1
MT-CO3
MT-ND4L
FTH1
NUPR1
APOC1
FTL
CSTB
C
Module 1 GO Enrichment
GO:0019886: antigen processing and presentation of exogenous peptide ant...
GO:0043312: neutrophil degranulation
GO:0030574: collagen catabolic process
GO:0060333: interferon-gamma-mediated signaling pathway
GO:0071887: leukocyte apoptotic process
GO:0002683: negative regulation of immune system process
GO:0072659: protein localization to plasma membrane
GO:0034394: protein localization to cell surface
GO:1903829: positive regulation of cellular protein localization
GO:0033028: myeloid cell apoptotic process
GO:0002886: regulation of myeloid leukocyte mediated immunity
GO:0050852: T cell receptor signaling pathway
GO:2001243: negative regulation of intrinsic apoptotic signaling pathway
GO:0002437: inflammatory response to antigenic stimulus
GO:0002573: myeloid leukocyte differentiation
GO:0032653: regulation of interleukin-10 production
GO:1901565: organonitrogen compound catabolic process
GO:0032613: interleukin-10 production
GO:0022411: cellular component disassembly
GO:0032755: positive regulation of interleukin-6 production
Enriched GO Term
-log10 P-val
Module 2 GO Enrichment
GO:0060337: type I interferon signaling pathway
GO:0051607: defense response to virus
GO:0045071: negative regulation of viral genome replication
GO:0060333: interferon-gamma-mediated signaling pathway
GO:0034341: response to interferon-gamma
GO:0046596: regulation of viral entry into host cell
GO:0045778: positive regulation of ossification
GO:0009306: protein secretion
GO:0032091: negative regulation of protein binding
GO:0008637: apoptotic mitochondrial changes
GO:0050663: cytokine secretion
GO:0002831: regulation of response to biotic stimulus
GO:0050709: negative regulation of protein secretion
GO:0002792: negative regulation of peptide secretion
GO:0002682: regulation of immune system process
GO:0001817: regulation of cytokine production
GO:0051224: negative regulation of protein transport
GO:1904950: negative regulation of establishment of protein localization
GO:1903531: negative regulation of secretion by cell
GO:0051260: protein homooligomerization
Module 3 GO Enrichment
GO:0043392: negative regulation of DNA binding
GO:0051651: maintenance of location in cell
GO:0050729: positive regulation of inflammatory response
GO:0032088: negative regulation of NF-kappaB transcription factor activi...
GO:0045185: maintenance of protein location
GO:0043299: leukocyte degranulation
GO:0030833: regulation of actin filament polymerization
GO:0051494: negative regulation of cytoskeleton organization
GO:0043280: positive regulation of cysteine-type endopeptidase activity ...
GO:1901222: regulation of NIK/NF-kappaB signaling
GO:1902904: negative regulation of supramolecular fiber organization
GO:0033209: tumor necrosis factor-mediated signaling pathway
GO:0002224: toll-like receptor signaling pathway
GO:0002366: leukocyte activation involved in immune response
GO:0002263: cell activation involved in immune response
GO:0042981: regulation of apoptotic process
GO:0008284: positive regulation of cell population proliferation
GO:0043067: regulation of programmed cell death
GO:0042102: positive regulation of T cell proliferation
GO:0060326: cell chemotaxis
Module 4 GO Enrichment
GO:0062014: negative regulation of small molecule metabolic process
GO:0006879: cellular iron ion homeostasis
GO:0006826: iron ion transport
GO:0006641: triglyceride metabolic process
GO:0048145: regulation of fibroblast proliferation
GO:0044242: cellular lipid catabolic process
GO:0045936: negative regulation of phosphate metabolic process
GO:0031330: negative regulation of cellular catabolic process
GO:0043086: negative regulation of catalytic activity
GO:0043312: neutrophil degranulation
GO:0051172: negative regulation of nitrogen compound metabolic process
GO:0051651: maintenance of location in cell
GO:0050729: positive regulation of inflammatory response
GO:0006869: lipid transport
GO:0015711: organic anion transport
GO:0042326: negative regulation of phosphorylation
GO:0006820: anion transport
GO:0032787: monocarboxylic acid metabolic process
GO:0051346: negative regulation of hydrolase activity
GO:0019637: organophosphate metabolic process


A
M1 MoMa
APOC1
FABP4
NUPR1
HLA-DRA
TYROBP
HLA-DRB5
CST3
TMSB4X
FTL
CTSD
LYZ
RPL36A
C1QC
CSTB
NEAT1
CORO1A
JAK1
KIAA1551
LGALS3
MALAT1
MT-CO1
MT-CO3
S100A9
CTSB
VAMP5
BCL2A1
GBP1
IFITM1
ISG15
GLUL
MX1
OASL
TNFSF10
IFIT3
IFITM3
IFIT1
RSAD2
MT-ND4L
IFI30
PLAUR
CTSL
GBP5
NFKBIA
H3F3B
MTRNR2L12
IFIT2
CD68
M2 MoMa
MALAT1
MT-CO3
MT-CO1
NEAT1
JAK1
TMSB4X
LYZ
CST3
HLA-DRA
TYROBP
HLA-DRB5
NUPR1
FABP4
S100A9
APOC1
RPL36A
MTRNR2L12
H3F3B
NFKBIA
IFI30
PLAUR
CD68
C1QC
OASL
IFIT2
IFIT1
IFIT3
IFITM3
MX1
TNFSF10
BCL2A1
GBP5
VAMP5
GBP1
CTSL
RSAD2
GLUL
ISG15
MT-ND4L
CTSD
CSTB
LGALS3
CTSB
NK
FCER1G
CTSD
TMSB4X
NUPR1
FABP4
RPL36A
C1QC
HLA-DRB5
CST3
LYZ
HLA-DRA
CTSB
BCL2A1
CTSL
S100A9
LGALS3
FTH1
CD68
FTL
PLAUR
GLUL
CSTB
MALAT1
IFI30
NFKBIA
VAMP5
IFITM1
H3F3B
MTRNR2L12
NEAT1
MT-CO1
MT-CO3
TYROBP
KIAA1551
CORO1A
JAK1
IFITM3
ISG15
MX1
RSAD2
IFIT3
IFIT1
CD3E
GBP1
CD4 T
MALAT1
IFITM1
GBP5
OASL
TNFSF10
IFITM3
H3F3B
ISG15
NFKBIA
CD3E
IFIT2
IFIT3
RSAD2
IFIT1
MX1
VAMP5
MT-ND4L
MTRNR2L12
KIAA1551
CORO1A
JAK1
APOC1
FCER1G
FTH1
CST3
TYROBP
LGALS3
FTL
CTSD
CSTB
S100A9
PLAUR
IFI30
BCL2A1
CTSB
GLUL
CTSL
NEAT1
FABP4
NUPR1
TMSB4X
C1QC
MT-CO3
CD8 Memory T
KIAA1551
JAK1
CORO1A
VAMP5
MTRNR2L12
MT-ND4L
GBP1
IFITM3
OASL
TNFSF10
GBP5
CD3E
RSAD2
MX1
IFIT3
IFIT1
IFIT2
IFITM1
ISG15
FABP4
TMSB4X
MT-CO1
MT-CO3
C1QC
NUPR1
IFI30
NFKBIA
CTSL
NEAT1
TYROBP
RPL36A
PLAUR
CTSD
CD68
APOC1
HLA-DRA
LYZ
CST3
LGALS3
FCER1G
GLUL
CSTB
FTH1
S100A9
FTL
BAL_H
BAL_M
BAL_S
PBMC_H
PBMC_M
PBMC_S
rDEG
Exp Z Diff
Exp
2
1
0
-1
-2
rDEG
bal
both
pbmc
Exp Z Diff
2
1
0
-1
-2
B
BAL_H
BAL_M
BAL_S
PBMC_H
PBMC_M
PBMC_S
NFKBIA
NUPR1
BCL2A1
MTRNR2L12
CTSL
CTSB


A
NEAT1
Identity
PBMC_S
PBMC_M
PBMC_H
BAL_S
BAL_M
BAL_H
0
2
4
Expression Level
MALAT1
Identity
PBMC_S
PBMC_M
PBMC_H
BAL_S
BAL_M
BAL_H
0
2
4
6
Expression Level
B
BAL rDEG Top Frequency
Frequency
NEAT1
MT-CO3
MX1
C1QC
CTSL
PBMC rDEG Top Frequency
Frequency
MX1
ISG15
MALAT1
MT-CO1
RSAD2
BAL rDEG Top LogFC
Severe - Mild logFC
NUPR1
APOC1
CTSL
NEAT1
FTH1
FABP4
CTSB
FTL
FCER1G
CTSD
PBMC rDEG Top LogFC
Severe - Mild logFC
MALAT1
S100A9
HLA-DRA
IFIT3
NFKBIA
MX1
LYZ
H3F3B
ISG15
CTSD
C
BAL_H
BAL_M
BAL_S
PBMC_H
PBMC_M
PBMC_S
NEAT1
MALAT1