1 **No evidence that plasmablasts transdifferentiate into developing neutrophils in severe**

2 **COVID-19 disease**

3

4 José Alquicira-Hernandez[1], Joseph E Powell[1,2]*, Tri Giang Phan[1,3]*

5

6 [1] Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst NSW 2010, Sydney, Australia

7 [2] UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, Australia

8 [3] St Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, 384 Victoria St, Darlinghurst

9 NSW 2010, Sydney, Australia

10 *equal contribution

11 Email: j.powell@garvan.org.au or t.phan@garvan.org.au

12

13

14 **Abstract**

15 A recent study by Wilk et al. of the transcriptome of peripheral blood mononuclear cells (PBMCs)

16 in seven patients hospitalized with COVID-19 described a population of "developing neutrophils"

17 that were "phenotypically related by dimensionality reduction" to plasmablasts, and that these two

18 cell populations represent a "linear continuum of cellular phenotype"[1]. The authors suggest that, in

19 the setting of acute respiratory distress syndrome (ARDS) secondary to severe COVID-19, a

20 "differentiation bridge from plasmablasts to developing neutrophils" connected these distantly

21 related cell types. This conclusion is controversial as it appears to violate several basic principles in

22 cell biology relating to cell lineage identity and fidelity. Correctly classifying cells and their

23 developmental history is an important issue in cell biology and we suggest that this conclusion is

24 not supported by the data as we show here that: (1) regressing out covariates such as unique

25 molecular identifiers (UMIs) can lead to overfitting; and (2) that UMAP embeddings may reflect

26 the expression of similar genes but not necessarily direct cell lineage relationships.

27

28 **Main**

29 Infection-induced plasmablasts are proliferating plasma cells derived from B cells of the lymphoid

30 cell lineage that secrete antibodies against invading pathogens[2]. In contrast, neutrophils are cells

31 derived from the myeloid cell lineage that trap, phagocytose and kill pathogenic organisms by

32 generating reactive oxygen species, secreting proteases and degradative enzymes and casting

33 neutrophil extracellular traps (NETs)[3]. Accordingly, the lineage conversion of plasmablasts into

34 developing neutrophils requires a massive re-organization of the cell at the level of the

35 transcriptome, epigenome, and proteome. While the forced transdifferentiation of B cell lines, B

36 cell progenitors and naïve B cells into macrophages have been described, notably via the ectopic

37 expression of CCAAT-enhancer-binding protein (C/EBP) family of transcription factors[4,5], the

38 spontaneous transdifferentiation of terminally differentiated plasma cells is an entirely different

39 hypothesis that would require compelling evidence, which was not presented in the paper.

40

41 The recent introduction of high-throughput single cell RNA-sequencing (scRNA-seq) has created

42 new opportunities for developmental biologists to map the developmental history of differentiated

43 cell types[6]. These technologies assay the expression of genes in a large number of cells – their cell

44 transcriptional state – and a standard computational approach is to flatten this high-dimensional data

45 into a two-dimensional (2D) Euclidean space as state manifolds. In these projections, the proximity

46 of one cell type to another may denote similarities in the cell transcriptional state. A major

47 challenge in the analysis of scRNA-seq data is to remove noise from technical variation while

48 preserving biological heterogeneity[7]. To understand how this trade-off is balanced, we first used

49 their code to reproduce figure 1c from their paper (**Fig. 1a**). In doing so, we noted that, in addition

50 to regressing out mitochondrial genes, ribosomal RNA and ribosomal genes, the authors also

51 regressed out the number of UMIs (nCount_RNA) and the number of expressed genes

52 (nFeature_RNA) from the gene expression data in their analysis. This step is unnecessary in Seurat

53 when using the SCTransform normalization method as the number of UMIs is explicitly modelled

54 using a regularised negative binomial regression[7]. Regressing out the number of UMIs using a

55 standard generalised linear model (GLM) is discouraged by the developers due to overfitting[7].

56 Moreover, the number of expressed genes in each cell is correlated to its number of UMIs (Pearson

57 correlation = 0.9249). We therefore re-analysed the data in three ways: without regressing out any

58 covariates; regressing out only mitochondrial genes (**Fig. 1b**); and regressing out mitochondrial and

59 ribosomal genes (**Fig. 1c**). This exercise shows that the reported relationship between developing

60 neutrophils and plasmablasts breaks down when technical noise is not removed at the expense of

61 biological variability and the data is not overfitted.

62

63 Accordingly, we believe that orthogonal approaches, such as single cell DNA sequencing to detect

64 rearranged immunoglobulin heavy and light chain variable genes in the developing neutrophils, are

65 needed to provide evidence of their B cell origin and support the conclusion that terminally

66 differentiated plasma cells transdifferentiate into developing neutrophils. Furthermore, unless the

67 data visualization parameters are carefully chosen, the 2D UMAP embeddings may give the

68 misleading impression that cell clusters are closer or further than they actually are. For example, in

69 figure 1c of Wilk et al., developing neutrophils appear to occupy a similar manifold space as

70 plasmablasts. However, from this viewpoint the plasmablasts appear to be distantly related to the B

71 cells from which they are derived. In addition, these types of analyses make the *a priori* assumption

72 that all the cell types being studied are developmentally related and linked by a cell lineage tree.

73 This may hold true when analysing *in vitro* differentiated cells that share a common ancestor cell of

74 origin that is present in the cell culture, but may not apply to PBMCs that contain a heterogeneous

75 collection of cell types that have arisen from different committed progenitors.

76

77 Since plasmablasts are derived from B cells, we decided to subset neutrophils, developing

78 neutrophils, B cells and plasmablasts to further explore any similarities between this cell clusters.

79 This analysis revealed that, counter-intuitively, neutrophils were not related to developing

80 neutrophils, and B cells were not related to plasmablasts (**Fig. 1d**). This exercise suggests instead

81 that the developing neutrophils and plasmablasts, in responding to severe COVID-19 disease, are

82 activated cell types that may share the expression of a number of genes and gene modules that have

83 led to their misclassification as related cell types. These gene modules may result from

84 hyperactivation of the immune system by SARS-CoV-2 and reflect the heightened state of cellular

85 activation. Indeed, in their analysis the authors point out that the developing neutrophils were

86 unlikely to be doublets and also unlikely to be granulocytes that had phagocytosed B cells as there

87 were no clinical features of hemophagocytic lymphohistiocytosis (HLH). However, HLH is often

88 difficult to diagnose clinically and severe COVID-19 disease has been linked in some cases with

89 secondary HLH[8,9].

90

91 Nevertheless, the paper by Wilk et al. is a timely and valuable contribution to the understanding of

92 how severe COVID-19 disease reshapes the distribution and activation state of cell clusters in

93 PBMCs. Neutrophils have a high density and are not normally present in Ficoll preparations of

94 blood used to isolate PBMCs. The developing neutrophils may therefore be similar to the pro-

95 inflammatory low-density granulocytes that have been described in a number of autoimmune

96 diseases[10]. These low-density neutrophils consists of hyposegmented immature neutrophils (band

97 forms) that are also produced during emergency granulopoiesis[11]. Thus, it is more plausible that the

98 developing neutrophils are produced in the bone marrow from myeloid precursors in response to

99 overwhelming infection with SARS-CoV-2.

100

101 In summary, scRNA-seq is becoming an increasingly popular tool for dissecting cellular

102 heterogeneity in complex biological systems. However, it does have its limitations and, in the

103 example shown here, may mislead and generate interesting hypotheses that risk being erroneously

104 accepted as conclusions without further cross-validation. Our analysis of the data from Wilk et al.

105   indicate that there is not enough evidence to suggest that developing neutrophils transdifferentiate

106   from plasmablasts.

107

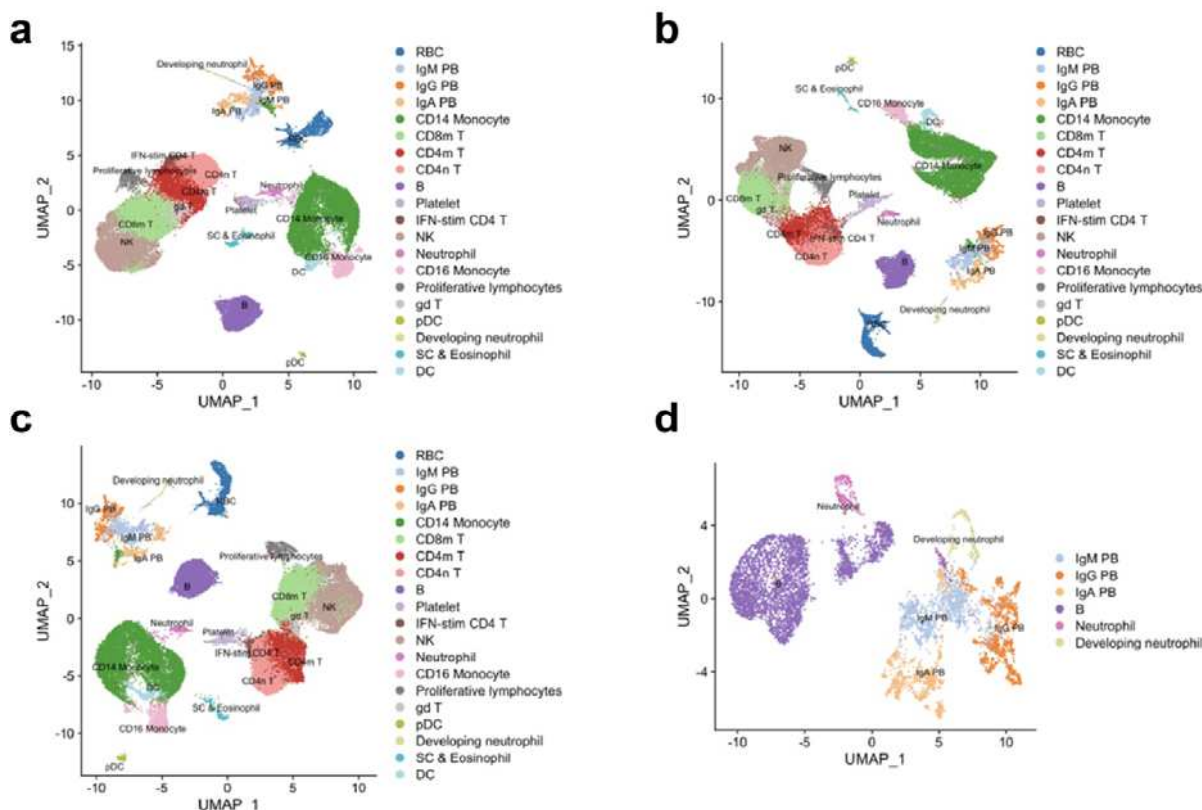108   **Figure 1**

109

110   **Figure 1 | Re-analysis of scRNA-seq data from Wilk et al. show developing neutrophils do not**

111   **transdifferentiate from plasmablasts. a.** Reconstruction of the UMAP embedding from figure 1c

112   of Wilk et al. **b.** UMAP embedding showing effect of regressing out mitochondrial genes. **c.** UMAP

113   embedding showing effect of regressing out mitochondrial genes and ribosomal genes. **d.** UMAP

114   embedding of neutrophils, developing neutrophils, B cells and plasma cells.

115

116   **Competing interests**

117   The authors declare no competing interests.

118

119   **Author contributions**

120   JAH performed the bioinformatic analysis; JAH, JEP and TGP wrote the manuscript.

121

122   **Acknowledgments**

125

126 **References**

127

128 1.   Wilk, A.J.*, et al.* A single-cell atlas of the peripheral immune response in patients with
129   severe COVID-19. *Nat Med* (2020).

130 2.   Fink, K. Origin and Function of Circulating Plasmablasts during Acute Viral Infections.
131   *Front Immunol* **3**, 78 (2012).

132 3.   Nauseef, W.M. & Borregaard, N. Neutrophils at work. *Nat Immunol* **15**, 602-611 (2014).

133 4.   Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages.
134   *Cell* **117**, 663-676 (2004).

135 5.   Cirovic, B.*, et al.* C/EBP-Induced Transdifferentiation Reveals Granulocyte-Macrophage
136   Precursor-like Plasticity of B Cells. *Stem Cell Reports* **8**, 346-359 (2017).

137 6.   Wagner, D.E. & Klein, A.M. Lineage tracing meets single-cell omics: opportunities and
138   challenges. *Nat Rev Genet* **21**, 410-427 (2020).

139 7.   Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-
140   seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).

141 8.   Mehta, P.*, et al.* COVID-19: consider cytokine storm syndromes and immunosuppression.
142   *Lancet* **395**, 1033-1034 (2020).

143 9.   Dimopoulos, G.*, et al.* Favorable Anakinra Responses in Severe Covid-19 Patients with
144   Secondary Hemophagocytic Lymphohistiocytosis. *Cell Host Microbe* **28**, 117-123 e111
145   (2020).

146 10.   Beyrau, M., Bodkin, J.V. & Nourshargh, S. Neutrophil heterogeneity in health and disease:
147   a revitalized avenue in inflammation and immunity. *Open Biol* **2**, 120134 (2012).

148 11.   Manz, M.G. & Boettcher, S. Emergency granulopoiesis. *Nat Rev Immunol* **14**, 302-314
149   (2014).

150