

## **A DNA methylome biosignature in alveolar macrophages from TB-exposed individuals predicts exposure to mycobacteria**

Jyotirmoy Das<sup>1</sup>, Nina Idh<sup>1</sup>, Isabelle Pehrson<sup>1</sup>, Jakob Paues<sup>1,2</sup> and Maria Lerm<sup>1\*</sup>

<sup>1</sup>Division of Infection and Inflammation, <sup>2</sup>Division of Infectious Diseases

Department of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences, Linköping University, SE-58185, Linköping, Sweden.

*Corresponding author:*

Maria Lerm, Professor in Medical Microbiology

Div. of Infection and Inflammation

Lab 1, floor 12

Dept. of Biomedical and Clinical Sciences

Faculty of Medicine and Health Sciences

Linköping University

SE-58185 Linköping

Sweden

Phone: +46-732707786

E-mail: [maria.lerm@liu.se](mailto:maria.lerm@liu.se)

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## **Abstract**

Several studies have identified biomarkers for tuberculosis (TB) diagnosis based on blood cell transcriptomics. Here, we instead studied epigenomics in the lung compartment by obtaining induced sputum from subjects included in a TB contact tracing. CD3- and HLA-DR-positive cells were isolated from the collected sputum and DNA methylome analyses performed. Unsupervised cluster analysis revealed that DNA methylomes of cells from TB-exposed individuals and controls appeared as separate clusters and the numerous genes that were differentially methylated were functionally connected. The enriched pathways were strongly correlated to data from published work on protective heterologous immunity to TB. Taken together, our results demonstrate that epigenetic changes related to trained immunity occurs in the pulmonary immune cells of TB-exposed individuals and that a DNA methylation signature can be derived from the DNA methylome. Such a signature can be further developed for clinical use as a marker of TB exposure.

## Introduction

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis*, which is transmitted between individuals through the inhalation of aerosols generated by coughing <sup>1</sup>. The disease still claims more than one million human lives annually and an expansion of the current toolkit for diagnosis, prevention and treatment is critical for reaching the United Nations' Sustainable Development Goals for 2030 of ending the TB epidemic <sup>2</sup>. In response to the urgent need for new approaches to diagnose TB, recent studies identified TB-specific biosignatures based on RNA transcription profiles in peripheral blood of TB-infected individuals <sup>3,4</sup>. Similar TB biosignatures have been identified that could predict disease progression <sup>5,6</sup> and reflect treatment monitoring <sup>6,7</sup>. Biosignatures for the detection of TB-exposure in easily accessible clinical samples could provide a basis for the development of novel effective point-of-care diagnostic tools.

The only available TB vaccine is Bacillus Calmette Guérin (BCG), which is based on live attenuated *M. bovis* <sup>8</sup>. In a recent study, we showed that administration of the BCG vaccine to healthy subjects induced profound epigenetic alterations in immune cells, which correlated with enhanced anti-mycobacterial activity in macrophages isolated from the vaccinees<sup>9</sup>. The changes were reflected in the DNA methylome, with the strongest response being recorded within weeks after vaccination<sup>9</sup>. Our observation that BCG induces alterations of the DNA methylome of immune cells has later been confirmed by others <sup>10,11</sup>. Since BCG vaccination reflects an *in vivo* interaction between immune cells and viable mycobacteria, we hypothesized that natural exposure to *M. tuberculosis* would induce similar changes not only in TB patients, but also in individuals who have been exposed to TB. To enable recruitment of control subjects with very low likelihood of previous TB exposure, we performed the study in a low-endemic setting. We

established a protocol for the isolation of distinct cell populations from induced sputum<sup>12</sup>, from which DNA could be isolated. Analyses of DNA methylomes of immune cells isolated from lungs and peripheral blood allowed us to identify distinct DNA methylation signatures in TB-exposed individuals. The signature was most prominent in the lung-derived cell populations. Pathway analyses revealed strong overlaps with previous studies on BCG-induced epigenetic signatures that could be correlated with protection against *M. tuberculosis*. We also identified pathway overlap with previous work on trained immunity induced by  $\beta$ -glucan from *Candida albicans*. In conclusion, we found a distinct pattern of DNA methylome changes in immune cells isolated from lungs in individuals with documented exposure to TB. The alterations strongly overlap with pathways described as reflecting trained immunity and an enhanced anti-mycobacterial response. The identified signature has potential to be used as a tool to identify TB exposure in low-endemic settings.

## Results

### *Study design*

To determine epigenetic changes in the immune cells in TB-exposed individuals, we recruited subjects enrolled in a routinely performed TB contact tracing at Linköping University Hospital, Sweden. Age-matched individuals were included as controls (Table 1). The index case was diagnosed with drug-sensitive pulmonary TB and had completed two out of six months of standard treatment at the time of sample collection. All included subjects except one (a TB contact) were BCG-vaccinated (Table 1). Interferon-Gamma Release Assay (IGRA) status was determined and among the exposed individuals, two were positive (including the index case) and among the controls, one individual (C2) was classified as ‘borderline’- positive<sup>13</sup> (Table 1). From induced sputum, HLA-DR-positive (antigen-presenting cells, dominated by macrophages<sup>14</sup> and CD3-positive (T cell) populations were purified, whereas the PBMC fraction extracted from blood were kept as a mixed population (Fig. 1).

### *DNA methylome data from TB-exposed individuals form a separate cluster*

DNA isolation from the studied cell populations was followed by global DNA methylation analysis using the Illumina 450K protocol. After curation of the data<sup>15</sup>, the datasets were subjected to unsupervised hierarchical cluster analysis based on DNA CpG methylation  $\beta$ -values (see project work-flow, Fig. 1). This approach accurately clustered the participants into TB-exposed and controls based on the DNA methylome data derived from both HLA-DR- and CD3-positive cell populations. (Fig. 2a, b). On the other hand, in the PBMC-derived dataset, the TB index case appeared outside the clusters and two of the controls clustered with the other exposed individuals, one of them (“Con\_2”) being the individual identified as border line-positive in the IGRA test (Fig. 2c).

Next, we identified the differentially methylated CpG sites (DMCs) and DMGs by comparing the TB-exposed and controls groups for each cell population. To filter out the most significantly altered DMGs in the dataset, the stringency criteria of  $\log_2$  0.3 fold increased or decreased  $\beta$ -values and Benjamini-Hochberg (BH)-corrected  $p$ -value  $< 0.05$  (HLA-DR),  $< 0.1$  (CD3) and  $< 0.2$  (PBMC) were applied. The results are depicted as volcano plots, which show that the DNA methylomes of TB-exposed most strongly differ in the HLA-DR cells as compared to control subjects, followed by the CD3 population, whereas PBMC datasets revealed fewer DMGs (Fig. 2d, e, f, Table 2). To highlight the locus position of the DMGs, chromosome maps were constructed (Suppl. Fig. S1). Using the same stringency criteria as for the HLA-DR analysis, we tested whether DMGs would emerge when the datasets were arranged in other possible groups as derived from the demographics ( $>/<$  median age, sex, IGRA status), Neither age nor IGRA status generated any significant DMGs with these settings, and gender rendered only three (Table 2).

#### *Functional enrichment analysis reveals common and unique interactomes in the datasets*

Using the Panther Database, we investigated whether the identified DMGs were enriched in known pathways (Fig. 3a,b,c). The analysis revealed pathways with relevance for TB infection, including hypoxia-inducible factor (HIF)1- $\alpha$  activation, Vitamin D metabolism and p38, Wnt, Notch, interleukin, chemokine and cytokine signaling pathways<sup>16–23</sup>. Common pathways shared between at least two of the cell populations included B cell activation, glycolysis, angiotensin II signaling, and cholecystokinin signaling. Notably, several pathways being named after their known functions in the nervous system were enriched in the studied cell populations, including pathways involved in axon guidance and adrenaline, acetylcholine and glutamate

signaling. In the PBMC population but not in the lung cell populations, the interferon- $\gamma$  signaling pathway was identified among the enriched pathways.

*Comparisons across cell populations and species reveals the existence of a common DNA methylome-based biosignature in mycobacteria-exposed immune cells*

Given the fact that the interaction between mycobacteria and eukaryotes is evolutionary ancient, we predicted that highly conserved pathways exist that are common among the studied cell populations. Combining the identified DMGs from the HLA-DR, CD3 and PBMCs in a Venn analysis, we discovered 185 common DMGs (Fig. 4a). We expanded the Venn analyses to include data from our previous work on BCG vaccine-induced DMGs that correlated with enhanced mycobacterial control<sup>9</sup>, the rationale being that natural exposure to TB and BCG vaccination both represent *in vivo* encounters between mycobacteria and host immune cells. Even though the routes of mycobacterial exposure differ profoundly in these settings, a set of 151 DMGs could be identified that overlapped between our previous BCG study<sup>9</sup> and all cell populations studied here (Fig. 4b), suggesting that a highly conserved epigenetic response to mycobacterial challenge exists.

In 2018, Hasso-Agopsowicz *et al.* described alterations in DNA methylation patterns in PBMCs from BCG-vaccinated individuals, with concomitant enrichment in many immune-related pathways<sup>10</sup>. In order to compare that study with ours, we performed Panther analysis with the 185 common DMGs and matched the identified enriched pathways with those from that study, revealing that 75% of those pathways were the same as in the present study (Fig. 5a and Suppl. Table 1a), further corroborating the relationship of the altered DNA methylation patterns induced through TB exposure and BCG-induced changes. In a recent mouse study by Saeed *et al.* BCG-induced alterations of the epigenome was correlated to protection against *M.*

*tuberculosis* infection<sup>11</sup>, and to translate our human DNA methylome signature to the signature identified in the mouse study, we searched for pathway overlaps between the two studies. To allow comparison with that study, we performed a Gene Ontology (GO) enrichment analysis (Suppl. Fig. S2a-c). Figure 5b and Suppl. Table 1b demonstrates that for our PBMC data, the GO terms “biological processes” overlapped to 100% with the mouse study (same cell population) and to 31% and 65% for HLA-DR and CD3 cells respectively. In 2014, Saeed et al<sup>24</sup> demonstrated the induction of trained immunity pathways by another immune-training agent,  $\beta$ -glucan. We assessed possible pathway overlap with that study and although there were fewer overlaps as compared to the BCG-induced pathways described above, again the strongest correlation was found in the PBMC fraction, in this case in the GO terms “cellular components” (Fig. 5c and Table 1c).

Finally, we assessed how well the 284 CpG sites corresponding to the 185 overlapping DMGs performed in an unsupervised cluster analysis. To this end, we included one additional TB patient and two contacts, and collected HLA-DR cells from induced sputum, since the DNA methylome data this cell type was clearly outperforming the others with respect to accurate separation of the groups. Fig. 6 shows a  $k$  means-based dendrogram with a heatmap of the  $\beta$  values of the 284 CpG sites from the previous and the new subjects’ samples.

Together the results demonstrate that the identified biosignature with the strongly enriched pathways can be linked to modifications of immune cell functions that result in improved anti-mycobacterial defense, possibly through trained immunity.



## Discussion

In this study, we present data suggesting that exposure to TB generates a very distinct DNA methylation signature in pulmonary immune cells. The signature was found not only in those with active or latent TB infection but also in individuals who are exposed but IGRA-negative. The finding that also healthy, TB-exposed carry the signature opens up the possibility that the epigenetic alterations reflect a host-beneficial reprogramming of the immune mechanisms rather than being induced by *M. tuberculosis* as a step to evade the immune defense. This notion is supported by our observation that the DMGs identified in the present study strongly overlapped with the previously reported DNA methylation changes induced during BCG vaccination, which correlated with increased anti-mycobacterial capacity of macrophages<sup>9</sup>. In addition, we demonstrate that the GO data derived from our dataset display a strong overlap with data from a study on protective BCG vaccination in mice<sup>11</sup>.

BCG vaccination has convincingly been shown to induce heterologous immunity protecting against childhood mortality from other causes than TB<sup>25,26</sup>. Based on our finding that natural TB exposure and BCG vaccination trigger similar epigenetic changes we propose the hypothesis that a “beneficial exposure” to TB exists, which protects against other infections through heterologous immunity. Along the same line, it has been shown that a substantial fraction of individuals exposed to TB can be defined as ‘early clearers’, since they do not test positively in the tuberculin skin test (TST) or IGRA<sup>27</sup>, suggesting effective eradication of the infection<sup>27</sup>. Identifying these early clearers and understanding the biology behind their resistance to TB infection could move the field forward towards novel strategies of TB prevention.

The rationale of comparing overlaps in mycobacteria-induced DNA methylation changes between different immune cells is based on the accumulated evidence of co-evolution of mycobacteria and amoeba<sup>28,29</sup> and the origin of phagocytic immune defense in metazoans from

amoeboid cells<sup>30,31</sup>, predicting evolutionary conserved pathways to be engaged in anti-mycobacterial defense. In line with this prediction, we identified the ubiquitously expressed and evolutionary conserved Wnt signaling pathway, which is found in all metazoans<sup>32</sup> and with homologs in amoeba<sup>33</sup> to be strongly enriched across all cell populations, settings and species. The role for Wnt signaling in mycobacterial defense remains elusive, but many studies have ascribed Wnt pathways immunomodulating functions and induction during *M. tuberculosis* infection (reviewed in <sup>20</sup>).

In concordance with the macrophage being the main host cell for mycobacterial infection, the strongest enrichment of DNA methylation changes was observed in the HLA-DR-positive cell population, which is dominated by alveolar macrophages. The pathways identified to be enriched in the HLA-DR population have been described in the context of trained immunity, BCG exposure and TB. For example, activation of Hypoxia-Inducible Factor 1  $\alpha$  and glycolysis pathways (P00030 and P00024, respectively) are hallmarks of macrophages that have undergone the epigenetic changes that are reflective of trained immunity (reviewed in <sup>34,35</sup>), which is induced in myeloid cells upon BCG-stimulation<sup>36</sup>. VEGF-release (P00056) by macrophages has been shown to recruit immune cells during granuloma formation <sup>37</sup>. Further, vitamin D has been shown to strengthen the anti-mycobacterial activity of macrophages<sup>17,38</sup>, and upregulation of components of the vitamin D pathway is linked to the production of anti-microbial peptides<sup>18</sup>, providing a possible effector mechanism for mycobacterial control. Recent literature on immune regulation through T cell-derived acetylcholine<sup>39,40</sup> attributes relevance to the acetylcholine receptor pathway identified among the HLA-DR pathways.

Although macrophages and lymphocytes are not generally viewed as having many similarities, we found 34 of the identified pathways to overlap between HLA-DR and CD3. In data derived from the CD3 and PBMC populations, both of which represent lymphocytes, overlaps were

identified for glycolysis, glutamate receptor and angiotensin II pathways. Interestingly, a metabolic shift towards increased glycolysis, representative of the Warburg effect, has been strongly associated with trained immunity<sup>34</sup>. However, the literature is dominated by the view that this event takes place in trained myeloid cells, while we identified this circuit in CD3 cells (lymphocytes) and not in the HLA-DR cells (dominated by macrophages). The glutamate receptor is widely expressed on immune cells and have been described as having an important regulatory role in T cells, which can also produce and release glutamate<sup>41</sup>. The role for angiotensin II pathway in TB remains elusive, while Angiotensin II Converting Enzyme 2 is currently in the spotlight due to fact that the SARS-CoV2 virus utilizes it as a receptor for entry into host cells<sup>42</sup>. In the PBMC population, which over all showed a weaker epigenetic response, we found the interferon- $\gamma$  signaling pathway, which has a well-established role in anti-mycobacterial defense (reviewed in<sup>43</sup>), to be among the reprogramed pathways.

A weakness of our study is the small cohort, which warrants testing in larger cohorts performed in different clinical settings such as areas high- and low endemic for TB. Taken together, we present data supportive of DNA methylation changes that are induced through exposure to TB. The changes correlate with findings from studies on BCG vaccination including TB protection, heterologous and trained immunity.

## Methods

### *Study design and participants*

Patients with pulmonary TB, participants with occupational- or household-related TB-exposure and healthy controls, with an age ranging from 18 to 53 years, were enrolled at Linköping University Hospital and Linköping University, respectively. Included subjects (please see Table 1 for demographics) donated peripheral blood and induced sputum samples<sup>12</sup> following oral and written informed consent (ethical approval obtained from the regional ethical review board in Linköping, #2016/237-31). The study protocol included questionnaires on respiratory and overall health, the evaluation of IGRA-status and sputum samples for DNA extraction.

### *Induced sputum and pulmonary immune cell isolation*

Induced sputum is a well-tolerated, non-invasive method to collect cells from the surface of the bronchial airways after inhalation of a hypertonic saline solution. The procedure of sputum induction takes approximately 30 minutes and is both cost effective and safe with minimal clinical risks<sup>44</sup>. Sputum specimens were collected as described by Alexis et al<sup>45</sup>, with the following modifications: premedication with an adrenergic  $\beta$ 2-agonist, salbutamol (Ventoline, 1ml 1mg/ml) was administrated before the inhalation of hypertonic saline, using a nebulizer (eFlow, PARI). The subsequent steps of sputum processing were adopted from Alexis et al. (2005)<sup>46</sup> and Sikkeland et al. <sup>14</sup>. The HLA-DR and CD3-positive cells were isolated using superparamagnetic beads coupled with anti-human CD3 and Pan Mouse IgG antibodies and HLA-DR/human MHC class II antibodies (Invitrogen Dynabeads, ThermoFisher, cat no 11041 and 14-9956-82, respectively). An initial positive selection was done with CD3 beads followed by a positive HLA-DR selection. Bead-coating and cell isolation was performed according to manufacturer's protocol.

### *PBMC isolation*

Following venipuncture, PBMCs were isolated using SepMate -50 tubes (StemCell Technologies), as per the manufacturer's protocol. The blood from each subject was poured into 50 ml tubes (Falcon, ThermoFisher) and carefully diluted 1:1 with a Dulbecco's PBS (D-PBS, Gibco by Life Technologies) containing 2% fetal bovine serum solution. Then, 15 ml Lymphoprep (StemCell Technologies) was transferred into a SepMate -50 tube (StemCell Technologies) and the diluted sample subsequently added to the Lymphoprep-filled tube, by carefully pipetting down the side of the tube. After centrifugation (1 200 x g, 10 minutes, RT), the plasma layer was removed and the PBMCs poured off into a sterile 50 ml tube (Falcon, ThermoFisher) containing D-PBS (Gibco by Life Technologies). The tube was subsequently filled with cold D-PBS until a total volume of 50 ml and centrifuged (300 x g, 10 minutes, 4°C). The supernatant was removed, and the pellet gently resuspended in D-PBS (Gibco by Life Technologies), using a transfer pipette (Sarstedt). The sample was then filled up to 50 ml with cold D-PBS, centrifuged (220 x g, 5 minutes, 4°C), and the supernatant discarded. In parallel, IGRA status was determined on whole-blood samples using QuantiFERON-TB Gold (Cellestis) following the manufacturer's instructions.

### *DNA methylation data analysis*

DNA of the HLA-DR, CD3 and PBMC was extracted using the AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions. Genome-wide DNA methylation analysis was performed using the HumanMethylation450K BeadChip (Illumina, USA) array at the Bioinformatics and Expression Analysis Core Facility at Karolinska Institute, Stockholm. The methylation profiles for each cell type were analyzed from the raw IDAT files in R (v4.0.2) using the *minfi* (v1.36.0) with subset-quantile within array (SWAN)

normalization<sup>47,48</sup> and *ChAMP* (v2.19.3) with beta-mixture quantile normalization (BMIQ)<sup>49,50</sup> packages. The type 1 and type 2 probes were normalized using the quantile normalization method. Using the default setup of the *ChAMP* package, following probes were filtered out: i) probes below the detection p-value ( $>0.01$ ), ii) non-CpG probes, iii) multi-hit probes, and iv) all probes of X and Y chromosomes. Cell type heterogeneity was corrected for the PBMC cell types using the Houseman algorithm<sup>51</sup> and batch effects were fixed using *ComBat* from the *SVA* package (v3.38.0)<sup>52</sup> Differential methylation analysis were performed with the linear modeling (lmFit) using the *limma* package (v3.46.0)<sup>53</sup> in a contrast matrix of the TB-exposed and TB-non-exposed (Control) individuals. All Differentially methylated CpGs (DMCs) were considered significant at the Bonferroni-Hochberg (BH) corrected  $p$ -value  $< 0.05$  (for HLA-DR cell types),  $<0.1$  (for CD3 cell types) and  $<0.2$  (for PBMC cell types).

#### *Unsupervised cluster analysis*

Hierarchical clustering of the all TB-exposed and control individuals was performed with the normalized  $\beta$ -values obtained after the data filtration in each cell type individually. The distance was calculated using the Euclidean distance matrix. The *dendextend* (v1.14.0)<sup>54</sup> and *ape* (v5.4-1)<sup>55</sup> packages in R were used to construct the horizontal hierarchical plots from the three different cell populations using the *hclust* and *dendrogram* functions.

#### *Structural annotations*

The *EnhancedVolcano*<sup>56</sup> package (v1.8.0) was used to generate the individual volcano plots from all cell populations. The *ChromoMap*<sup>57</sup> package (v0.3) was used to annotate and visualize the genome-wide chromosomal distribution of the DMGs. The interactive plots were generated using the *plotly* (v4.9.3) package<sup>58</sup>.

The heatmaps were generated from the filtered DMGs with their respective CpGs for each cell type using the *ComplexHeatmap* (v2.6.2) package. The clustering dendrogram in heatmaps were plotted using the Euclidean distance matrix.

#### *Pathway and functional enrichment analyses*

We also used Panther database (PantherDB v15)<sup>59</sup> to identify the enriched pathways related to our identified DMGs. In addition, to assess functional enrichment, we used the *ReactomePA* (v1.34.0)<sup>60</sup> package with 1000 permutations and the BH-corrected *p*-values. Within the package, GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) were used and using *clusterProfiler* (v3.18.1)<sup>61</sup>, we performed KEGG pathway enrichment analysis (data not shown). To enhance the visualization and better understanding of the enrichment result, *GPlot* (v1.0.2)<sup>62</sup>, another R package was used. The pathway enrichment were also calculated using the topology-based ontology methods using *RontoTools*<sup>63</sup> (v2.18.0)<sup>64</sup> (v2.40.1), *SPIA*<sup>65</sup> (v2.42.0) and *pathview*<sup>66</sup> (v1.30.1) was used to visualize the related pathways with the KEGG pathway maps (data not shown).

#### *Venn and overlap analyses*

Venn analyses were performed in order to detect the DMGs overlapping between cell populations and between studies. We constructed the Venn diagrams by using *matplotlib-venn* package (<https://github.com/konstantint/matplotlib-venn>) using in-house python script. The overlap analyses were calculated and plotted using the *go.Sunburst* function from plotly using an in-house python script.

### *Statistical analyses*

All differences with a  $p$ -value  $< 0.05$  were considered significant if not otherwise stated. We calculated family-wise error rate (FWER) using the BH correction method. All analyses were performed in R (v4.0.2) with the mentioned packages.

### **Acknowledgements**

This study was funded through generous grants from Forskningsrådet Sydöstra Sverige (FORSS-932096), the Swedish Research Council (2015-02593 and 2018-02961) and the Swedish Heart Lung Foundation (20150709 and 20180613). J.D is a postdoctoral fellow supported through the Medical Infection and Inflammation Center (MIIC) at Linköping University. We direct our gratitude to the staff at Linköping University Hospital and the Vrinnevi Hospital in Norrköping for assistance in sample collection and all the subjects for donating samples. The DNA methylome data were generated at the Bioinformatics and Expression Analysis Core Facility at the Department of Biosciences and Nutrition, which is supported by the Board of Research at the Karolinska Institute, Stockholm. The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Linköping University campus partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

### **Author contributions**

M.L., N.I. and J.P. conceived and designed the study, N.I. and I.P. collected the patient samples and compiled the demographics tables, J.D. and M.L. designed and performed the bioinformatic analysis of the data, J.D. wrote the code and created the figures. N.I, I.P. J.D and M.L wrote the manuscript. All authors declare no conflict of interest.



## References

1. Donald, P. R. *et al.* Droplets, dust and Guinea pigs: An historical review of tuberculosis transmission research, 1878-1940. *Int. J. Tuberc. Lung Dis.* **22**, 972–982 (2018).
2. WHO. *The End TB strategy.* (2015).
3. Berry, M. P. R. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* (2010) doi:10.1038/nature09247.
4. Bloom, C. I. *et al.* Transcriptional Blood Signatures Distinguish Pulmonary Tuberculosis, Pulmonary Sarcoidosis, Pneumonias and Lung Cancers. *PLoS One* (2013) doi:10.1371/journal.pone.0070630.
5. Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* (2016) doi:10.1016/S0140-6736(15)01316-1.
6. Suliman, S. *et al.* Four-gene pan-African blood signature predicts progression to tuberculosis. *Am. J. Respir. Crit. Care Med.* (2018) doi:10.1164/rccm.201711-2340OC.
7. Cliff, J. M. *et al.* Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J. Infect. Dis.* (2013) doi:10.1093/infdis/jis499.
8. Lerm, M. & Netea, M. G. Trained immunity: A new avenue for tuberculosis vaccine development. *Journal of Internal Medicine* (2016) doi:10.1111/joim.12449.
9. Verma, D. *et al.* Anti-mycobacterial activity correlates with altered DNA methylation pattern in immune cells from BCG-vaccinated subjects. *Sci. Rep.* **7**, (2017).
10. Hasso-Agopsowicz, M., Scriba, T. J., Hanekom, W. A., Dockrell, H. M. & Smith, S. G. Differential DNA methylation of potassium channel KCa3.1 and immune signalling pathways is associated with infant immune responses following BCG vaccination. *Sci. Rep.* (2018) doi:10.1038/s41598-018-31537-9.

11. Kaufmann, E. *et al.* BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. *Cell* **172**, 176-190.e19 (2018).
12. Das, J., Idh, N., Sikkeland, L. I. B., Paues, J. & Lerm, M. DNA methylome-based validation of induced sputum as an effective protocol to study lung immunity: construction of a classifier of pulmonary cell types. *bioRxiv* (2021) doi:10.1101/2021.03.12.435086.
13. Jonsson, J. *et al.* A borderline range for Quantiferon Gold In-Tube results. *PLoS One* **12**, e0187313 (2017).
14. Sikkeland, L. I. B., Kongerud, J., Stangeland, A. M., Haug, T. & Alexis, N. E. Macrophage enrichment from induced sputum [3]. *Thorax* (2007) doi:10.1136/thx.2006.073544.
15. Das, J., Verma, D., Gustafsson, M. & Lerm, M. Identification of DNA methylation patterns predisposing for an efficient response to BCG vaccination in healthy BCG-naïve subjects. *Epigenetics* 1–13 (2019) doi:10.1080/15592294.2019.1603963.
16. Ogryzko, N. V. *et al.* Hif-1 $\alpha$ -Induced Expression of Il-1 $\beta$  Protects against Mycobacterial Infection in Zebrafish. *J. Immunol.* **202**, (2019).
17. Eklund, D. *et al.* Vitamin D enhances IL-1 $\beta$  secretion and restricts growth of Mycobacterium tuberculosis in macrophages from TB patients. *Int. J. Mycobacteriology* **2**, (2013).
18. Liu, P. T. *et al.* Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science (80-. )*. **311**, 1770–1773 (2006).
19. Hölscher, C. *et al.* Chemical p38 MAP kinase inhibition constrains tissue inflammation and improves antibiotic activity in Mycobacterium tuberculosis-infected mice. *Sci. Rep.* **10**, (2020).
20. Brandenburg, J. & Reiling, N. The Wnt blows: On the functional role of Wnt signaling

- in mycobacterium tuberculosis infection and beyond. *Frontiers in Immunology* (2016)  
doi:10.3389/fimmu.2016.00635.
21. Castro, R. C. *et al.* NOTCH1 and DLL4 are involved in the human tuberculosis progression and immune response activation. *Tuberculosis* **124**, (2020).
  22. Eklund, D. *et al.* Human gene variants linked to enhanced NLRP3 activity limit intramacrophage growth of mycobacterium tuberculosis. in *Journal of Infectious Diseases* (2014). doi:10.1093/infdis/jit572.
  23. Raffetseder, J. *et al.* Replication rates of mycobacterium tuberculosis in human macrophages do not correlate with mycobacterial antibiotic susceptibility. *PLoS One* (2014) doi:10.1371/journal.pone.0112426.
  24. Saeed, S. *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* (80-. ). **345**, (2014).
  25. Benn, C. S., Fisker, A. B., Whittle, H. C. & Aaby, P. Revaccination with Live Attenuated Vaccines Confer Additional Beneficial Nonspecific Effects on Overall Survival: A Review. *EBioMedicine* **10**, 312–317 (2016).
  26. Thyssen, S. M. *et al.* Neonatal BCG vaccination and child survival in TB-exposed and TB-unexposed children: A prospective cohort study. *BMJ Open* **10**, (2020).
  27. Verrall, A. J. *et al.* Early clearance of Mycobacterium tuberculosis is associated with increased innate immune responses . *J. Infect. Dis.* (2019) doi:10.1093/infdis/jiz147.
  28. Dunn, J. D. *et al.* Eat prey, live: Dictyostelium discoideum as a model for cell-autonomous defenses. *Frontiers in Immunology* vol. 8 (2018).
  29. Chen, G., Zhuchenko, O. & Kuspa, A. Immune-like phagocyte activity in the social amoeba. *Science* (80-. ). **317**, 678–681 (2007).
  30. Zhang, X., Zhuchenko, O., Kuspa, A. & Soldati, T. Social amoebae trap and kill bacteria by casting DNA nets. *Nat. Commun.* **7**, (2016).

31. Zhang, X. & Soldati, T. Of amoebae and men: Extracellular DNA traps as an ancient cell-intrinsic defense mechanism. *Front. Immunol.* **7**, (2016).
32. Holstein, T. W. The evolution of the wnt pathway. *Cold Spring Harb. Perspect. Biol.* **4**, 1–17 (2012).
33. Harwood, A. J. Dictyostelium development: A prototypic Wnt pathway? *Methods Mol. Biol.* **469**, 21–32 (2008).
34. Arts, R. J. W., Joosten, L. A. B. & Netea, M. G. Immunometabolic circuits in trained immunity. *Seminars in Immunology* vol. 28 425–430 (2016).
35. Kumar, R. *et al.* Immunometabolism of Phagocytes During Mycobacterium tuberculosis Infection. *Frontiers in Molecular Biosciences* vol. 6 (2019).
36. Kleinnijenhuis, J. *et al.* Bacille Calmette-Guerin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc. Natl. Acad. Sci.* **109**, 17537–17542 (2012).
37. Harding, J. S. *et al.* VEGF-A from Granuloma Macrophages Regulates Granulomatous Inflammation by a Non-angiogenic Pathway during Mycobacterial Infection. *Cell Rep.* **27**, 2119-2131.e6 (2019).
38. Rao Muvva, J., Parasa, V. R., Lerm, M., Svensson, M. & Brighenti, S. Polarization of Human Monocyte-Derived Cells With Vitamin D Promotes Control of Mycobacterium tuberculosis Infection. *Front. Immunol.* **10**, (2020).
39. Cox, M. A. *et al.* Beyond neurotransmission: acetylcholine in immunity and inflammation. *Journal of Internal Medicine* vol. 287 120–133 (2020).
40. Cox, M. A. *et al.* Choline acetyltransferase-expressing T cells are required to control chronic viral infection. *Science (80-. ).* **363**, (2019).
41. Ganor, Y. & Levite, M. The neurotransmitter glutamate and human T cells: Glutamate receptors and glutamate-induced direct and potent effects on normal human T cells,

- cancerous human leukemia and lymphoma T cells, and autoimmune human T cells. *J. Neural Transm.* **121**, 983–1006 (2014).
42. Kuhn, J. H., Li, W., Radoshitzky, S. R., Choe, H. & Farzan, M. Severe acute respiratory syndrome coronavirus entry as a target of antiviral therapies. *Antiviral Therapy* vol. 12 639–650 (2007).
43. Lake, M. A., Ambrose, L. R., Lipman, M. C. I. & Lowe, D. M. ’ ‘Why me, why now?’ Using clinical immunology and epidemiology to explain who gets nontuberculous mycobacterial infection. *BMC Med.* **14**, (2016).
44. Ugarte-Gil, C., Elkington, P. T., Gotuzzo, E., Friedland, J. S. & Moore, D. A. J. Induced sputum is safe and well-tolerated for TB diagnosis in a resource-poor primary healthcare setting. *Am. J. Trop. Med. Hyg.* **92**, 633–635 (2015).
45. Alexis, N., Soukup, J., Ghio, A. & Becker, S. Sputum Phagocytes from Healthy Individuals Are Functional and Activated: A Flow Cytometric Comparison with Cells in Bronchoalveolar Lavage and Peripheral Blood. *Clin. Immunol.* **97**, 21–32 (2000).
46. Alexis, N. E. *et al.* In Vivo Particle Uptake by Airway Macrophages in Healthy Volunteers. doi:10.1165/rcmb.2005-0373OC.
47. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* (2012) doi:10.1186/gb-2012-13-6-r44.
48. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
49. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* (2013) doi:10.1093/bioinformatics/bts680.

50. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430 (2014).
51. Yuan, L. & De-Shuang Huang, &. A Network-guided Association Mapping Approach from DNA Methylation to Disease. doi:10.1038/s41598-019-42010-6.
52. Leek, J. T. *et al.* sva: Surrogate Variable Analysis. *R package version 3.20.0*. (2016).
53. Ritchie, M. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies | Nucleic Acids Research | Oxford Academic. *limma powers Differ. Expr. Anal. RNA-sequencing microarray Stud.* (2015).
54. Galili, T. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv428.
55. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* (2019) doi:10.1093/bioinformatics/bty633.
56. Blighe, K. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. (2019).
57. Anand, L. chromoMap: An R package for Interactive Visualization and Annotation of Chromosomes. doi:10.1101/605600.
58. Sievert, C. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. *Interactive Web-Based Data Visualization with R, plotly, and shiny* (2020). doi:10.1201/9780429447273.
59. Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
60. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. BioSyst.* **12**, 477–479 (2016).
61. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

62. Walter, W., Sánchez-Cabo, F. & Ricote, M. GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv300.
63. Voichita, C., Ansari, S. & Draghici, S. *ROntoTools: The R Onto-Tools suite*. <https://bioconductor.riken.jp/packages/3.10/bioc/vignettes/ROntoTools/inst/doc/rontotools.pdf> (2019).
64. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* (2009) doi:10.1186/1471-2105-10-161.
65. Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* (2009) doi:10.1093/bioinformatics/btn577.
66. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).

## Tables

Characteristics	Exposed (n = 4)	Controls (n = 6)
Mean age (year) <sup>†</sup>	36±12	27±6
Mean height (cm) <sup>†</sup>	173±4	176±7
Mean weight (kg) <sup>†</sup>	69±8	77±16
Mean Body Mass Index (BMI) <sup>†</sup>	23±3	24±5
Sex (male/female)	1/3	4/2
Smoking (current/previous/never)	0/1/3	0/0/6
BCG (yes/no)	3/1	6/0
IGRA-positive/IGRA-negative	2/2	1*/5

Table 1: Demographic data of the participants.

<sup>†</sup>The standard deviation of the mean values is added to the age, height, weight and BMI.

\* Borderline-positive.



Characteristics		Number of Differentially Methylated Genes (DMGs)	Gene Symbol of DMGs	Level of Significance (adjusted <i>p</i> - value)	Cell population
TB Exposure	exposed vs. controls	9,266	(not shown)	0.05	HLA-DR
		4,131	(not shown)		CD3
		537	(not shown)		PBMC
Age	younger (< 30) vs. older (≥ 30)	0		0.05	CD3
		0			HLA-DR
		0			PBMC
Gender	male vs. female	3	GALNTL5; MIR3201; OPN3		CD3
		3	COL9A1; ACTG1; REAT1E		HLA-DR
		0			PBMC
IGRA	positive vs. negative	0			CD3
		0			HLA-DR
		0			PBMC
BMI	high (> 25) vs low (≤ 25)	0		CD3	
		0		HLA-DR	
		0		PBMC	

Table 2: Differentially methylated genes emerging through comparison of different characteristics (TB-exposure, age, gender, IGRA status and BMI) of the participants in the three different cell populations, CD3, HLA-DR and PBMC.

<b>Characteristics</b>	<b>TB patients (n = 2)</b>	<b>TB-exposed (n = 2)</b>
Mean age (year) <sup>†</sup>	19.5±1.5 <sup>†</sup>	35.5±17.5
Mean height (cm) <sup>†</sup>	176±7	171±3
Mean weight (kg) <sup>†</sup>	62.5±6.5	80±10
Mean Body Mass Index (BMI) <sup>†</sup>	19.9±0.6	27.3±2.5
Sex (male/female)	2/0	1/1
Smoking (current/previous/never)	2/0/0	1/0/1
BCG (yes/no)	0/2	2/0

Table 3: Demographic data of the participants in the second recruitment (“test dataset”)

<sup>†</sup>The standard deviation of the mean values of age, height, weight and BMI.

## Figure legends

Figure 1: Flow chart of the project workflow.

Figure 2: DNA methylome analyses. a, b, c) Dendrograms of the unsupervised hierarchical clustering of the DNA methylation  $\beta$ -values of from HLA-DR, CD3 and PBMCs, respectively. “Con”: green=controls, “Exp”: purple=TB-exposed, red=TB index case. The scale defines the clustering Euclidean distance. d, e, f) Volcano plots of DMGs from HLA-DR, CD3 and PBMCs. Red dots represent DMGs above cut-offs ( $\pm 0.3$  Log<sub>2</sub> fold change and BH-corrected  $p$ -value  $< 0.05$ ,  $< 0.1$  or  $0.2$  as indicated).

Figure 3. Panther pathway analysis of the identified DMGs with the cut-offs for the different cell populations given in Figure 2. Bubble charts show the gene ratio, gene counts and FDR-corrected  $p$ -value for a) HLA-DR (top 20 pathways), b) CD3 (totally 17 pathways) and c) PBMC (top 20 pathways).

Figure 4: Venn analyses comparing DMGs, pathways and GO terms between different datasets a) Overlapping DMGs derived from the HLA-DR, CD3 and PBMC DNA methylomes, b) Overlapping DMGs from this study and from our previous work (Verma et al).

Figure 5. Pathway overlap with other studies' results. a) Venn diagram describing the number of Panther pathways overlapping between the ones derived from the 185 common DMGs in this study (dark green) and Hasso-Agopsowicz et al (human BCG vaccine study, light green)<sup>10</sup>, b) Sunburst Plot describing the overlap of enriched GO biological processes emerging from a comparison between the GO data derived from the 185 common DMGs (Figure 4a) and

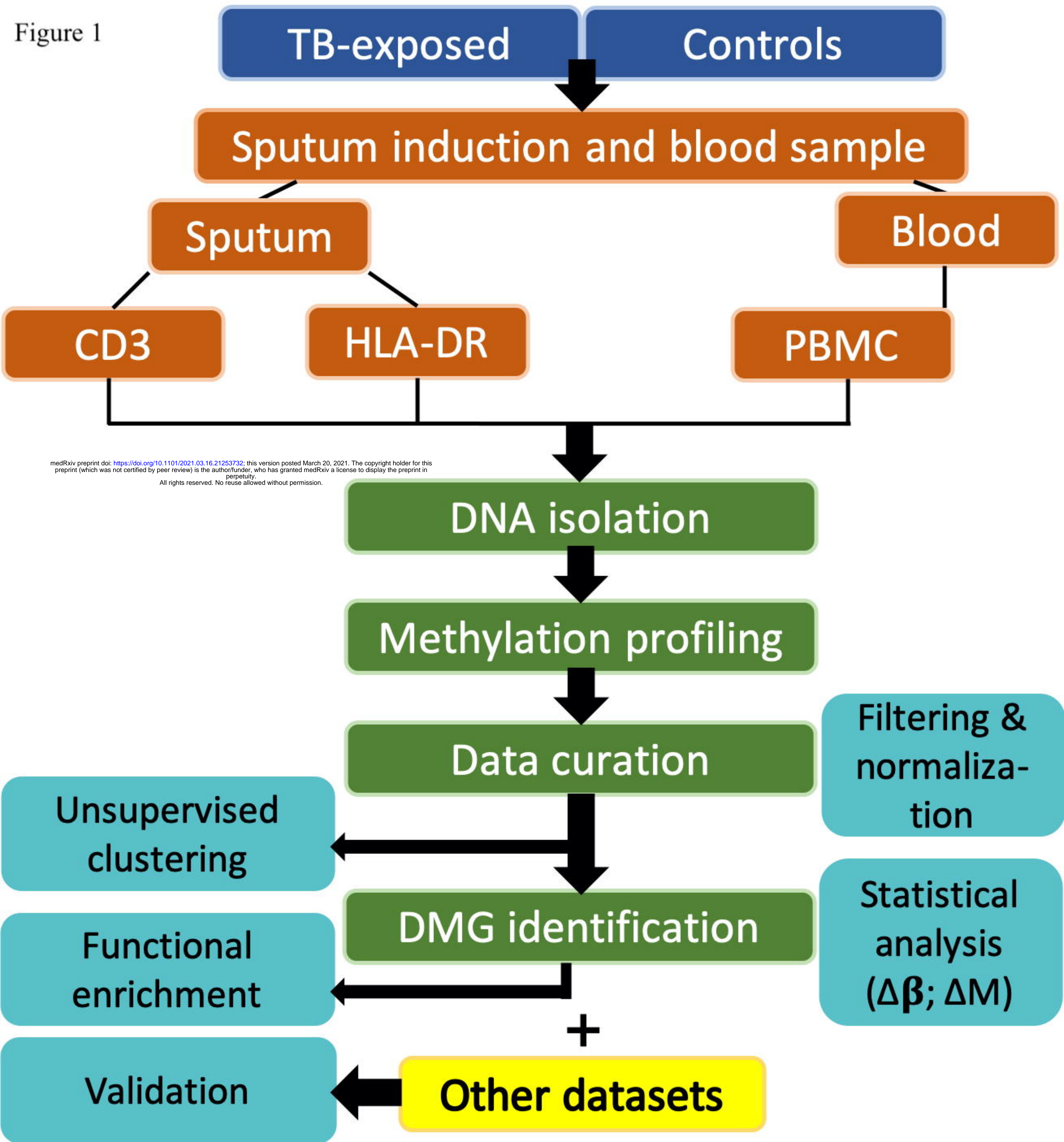
Kaufmann et al (BCG study performed in mouse PBMCs<sup>11</sup>, c) Sunburst Plot describing the overlap of enriched GO biological processes emerging from a comparison between the GO data derived from the 185 common DMGs and Saeed et al (study on trained immunity induced by  $\beta$ -glucan<sup>24</sup>).

Figure 6. Heatmap of the HLA-DR-derived  $\beta$  values of the signature's 284 CpG sites of the 6 initial subjects (Exp\_1-4 and Con\_1-6) and the three additional exposed subjects (Exp\_6-8). Purple=exposed, red=TB index case, green=controls.

Supplementary Figure S1. Chromosome maps showing the chromosomal distribution of all DMCs in 22 autosomes (hypomethylated regions in blue and hypermethylated regions in red). a) HLA-DR, b) CD3 and c) PBMC.

Supplementary Figure S2. Gene Ontology (GO) enrichment functional analysis a) HLA-DR, b) CD3, c) PBMC and d) the list of functional categories identified in the dataset. The colour bar shows the logFC values ranging from -1 (blue) to 1 (red).

Figure 1



medRxiv preprint doi: <https://doi.org/10.1101/2021.03.16.21253732>; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 2a

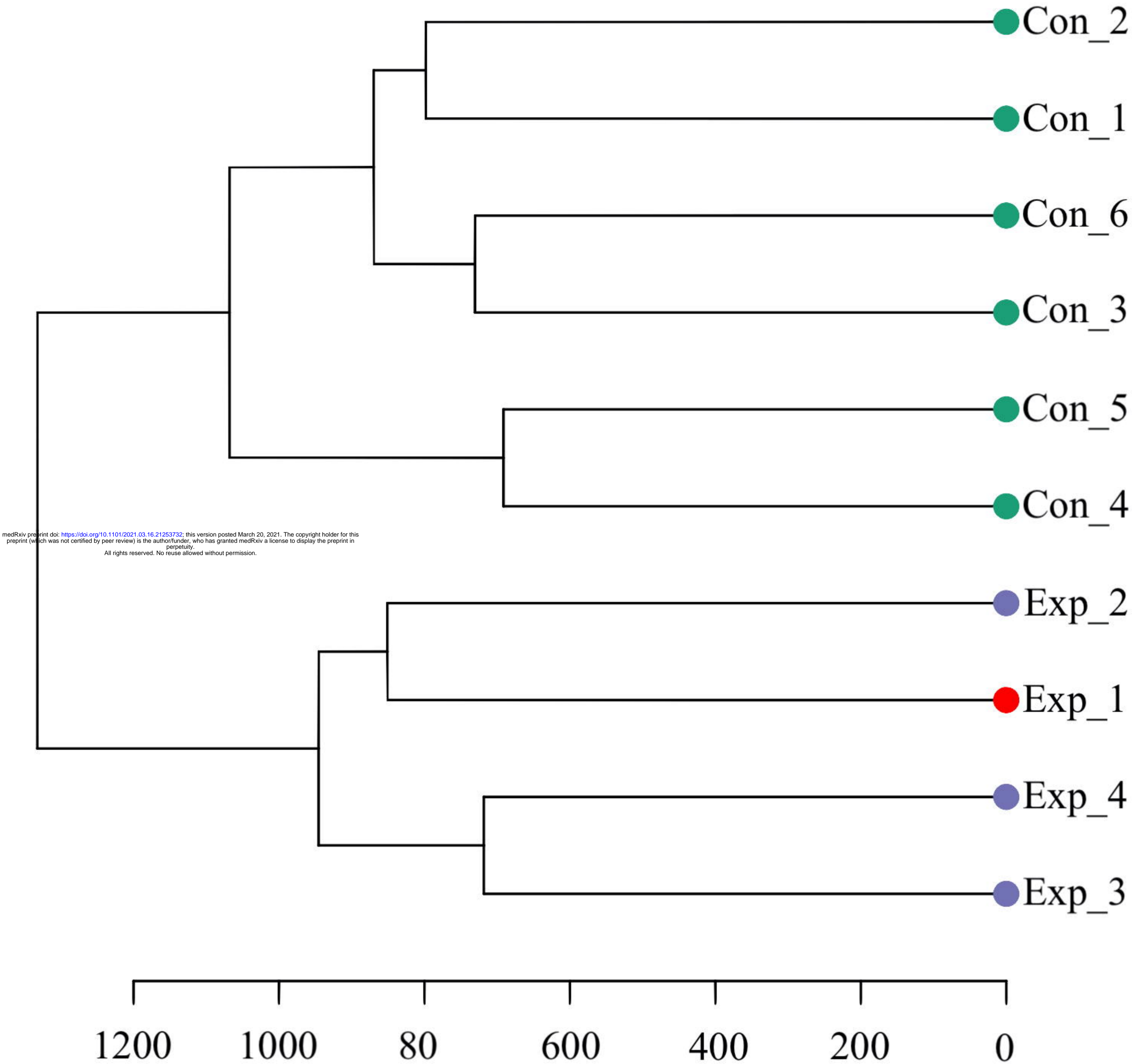


Figure 2b

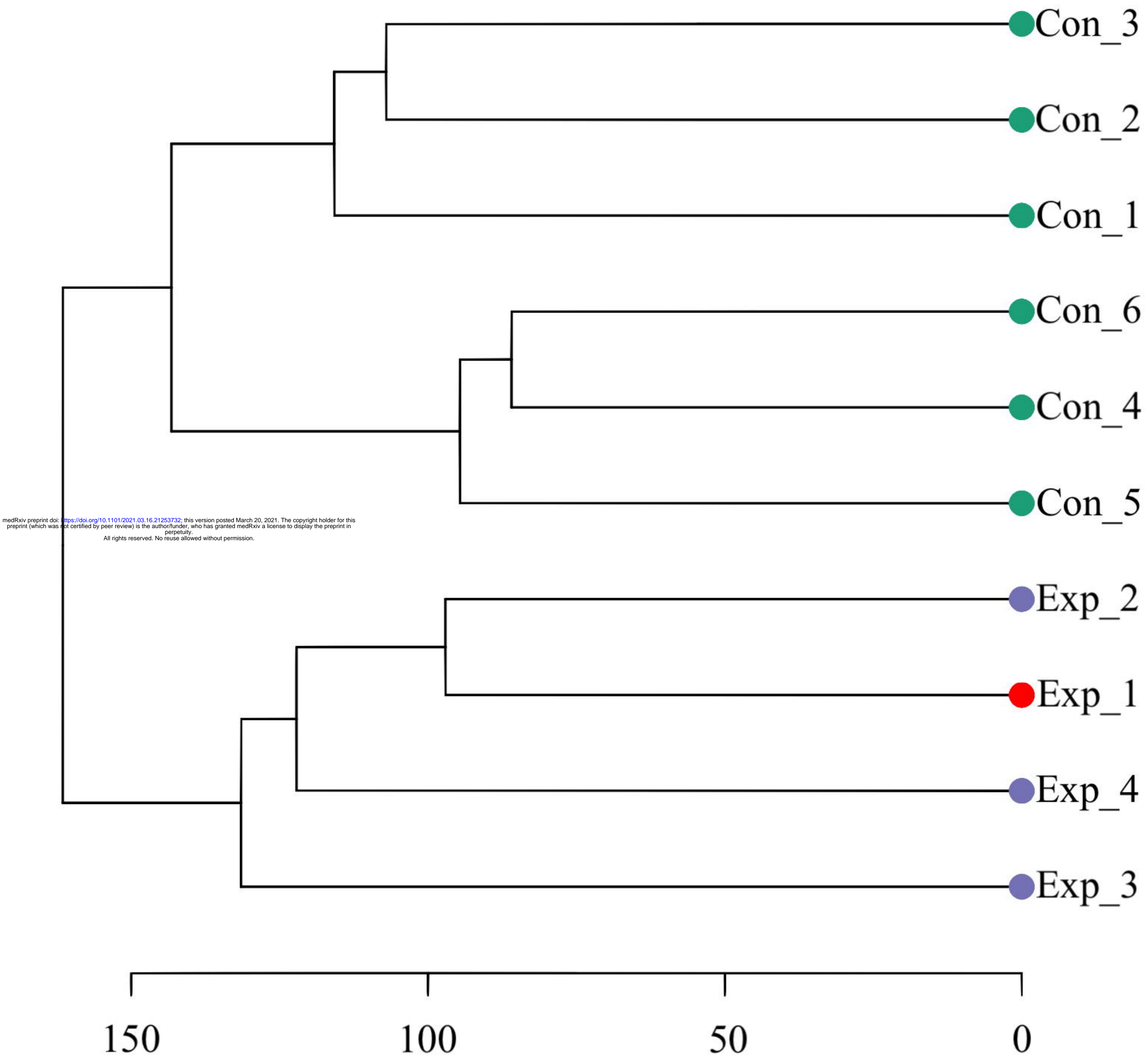


Figure 2c

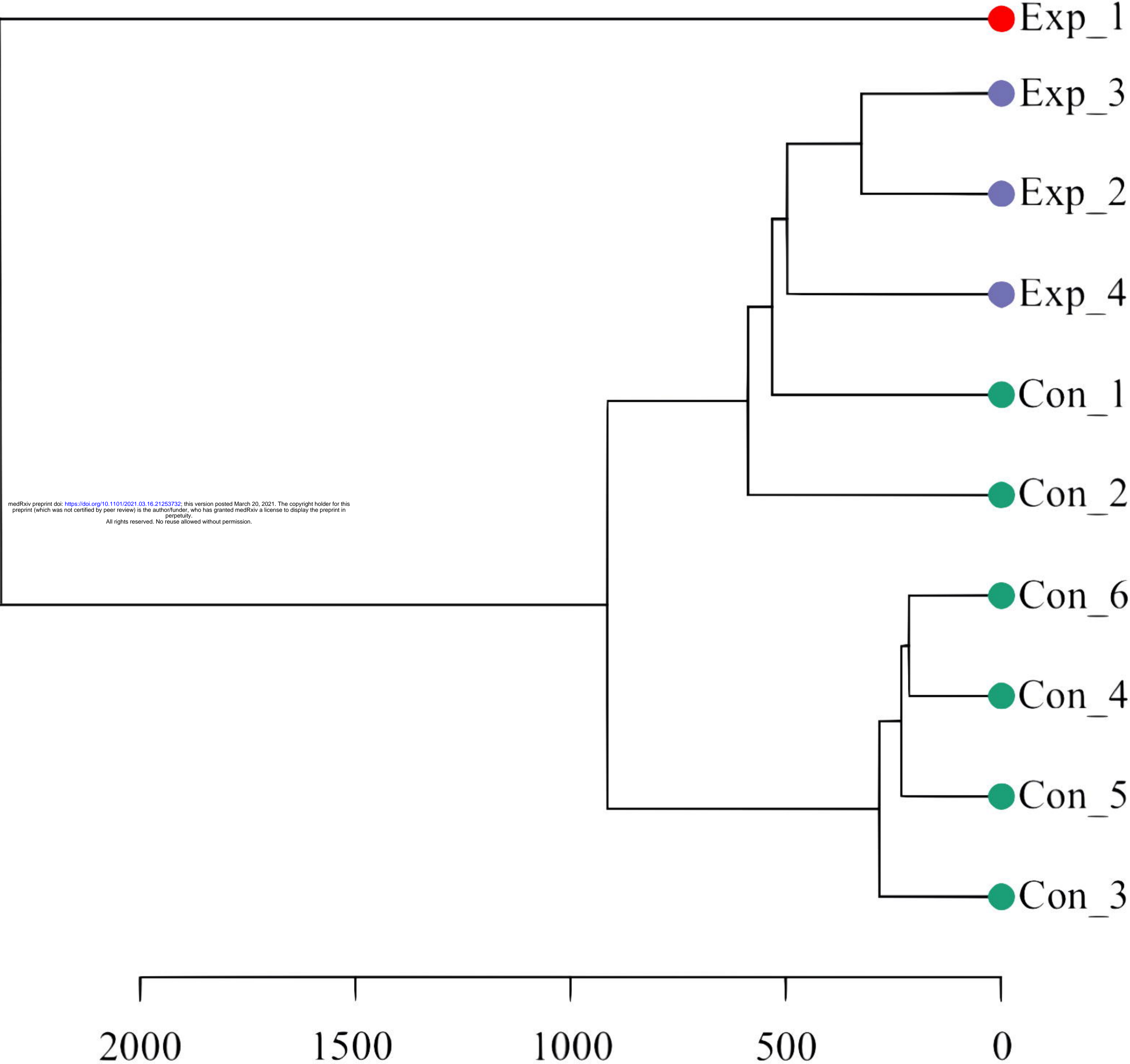




Figure 2d

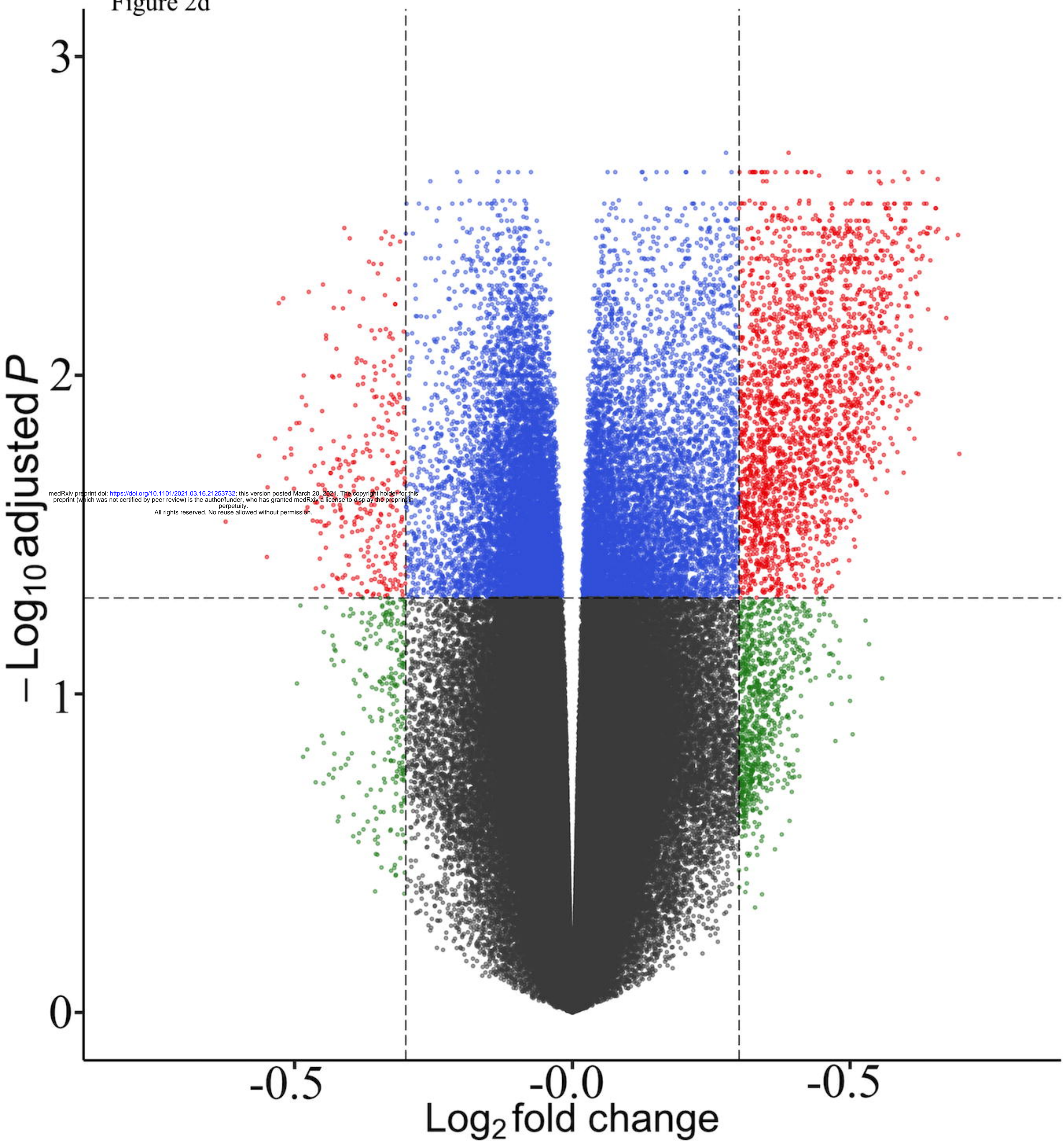


Figure 2e

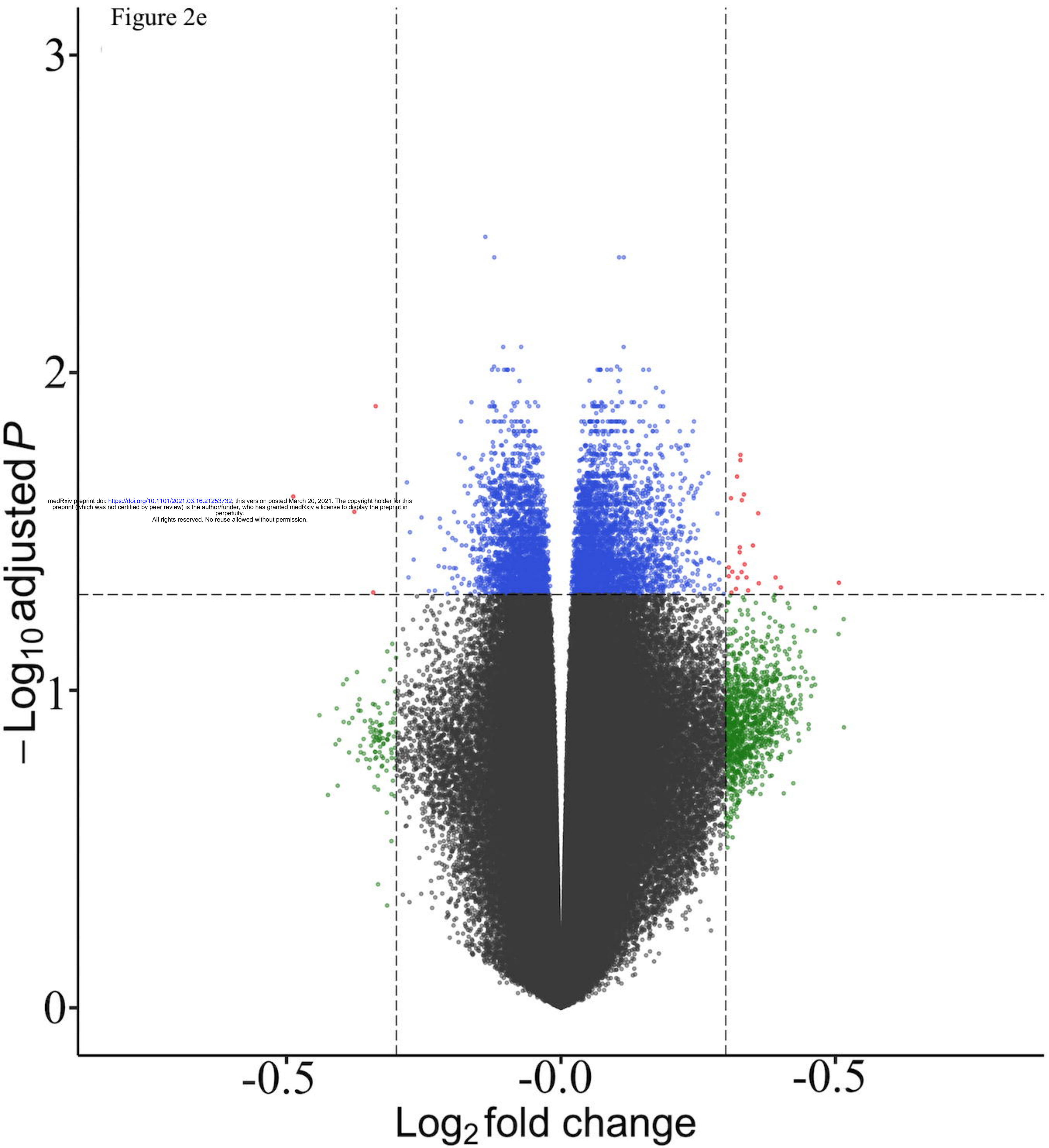


Figure 2f

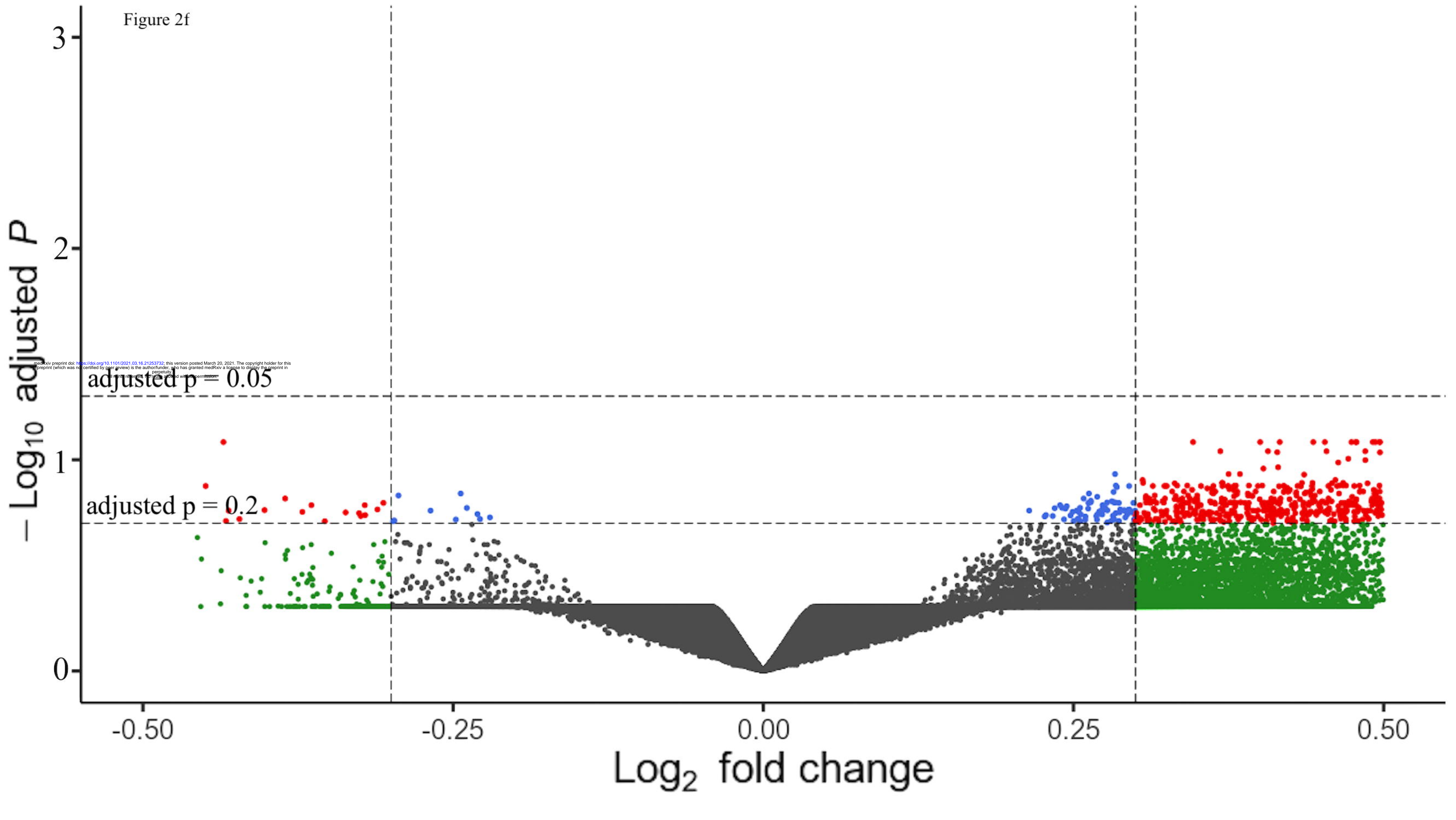


Figure 3a

Vitamin D metabolism and pathway -

Toll receptor signaling pathway -

Thyrotropin-releasing hormone receptor signaling pathway -

PI3 kinase pathway -

Notch signaling pathway -

Nicotinic acetylcholine receptor signaling pathway -

medRxiv preprint doi: <https://doi.org/10.1101/2021.03.16.21253732>; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Interleukin signaling pathway -

Hypoxia response via HIF activation -

Heme biosynthesis -

FAS signaling pathway -

Endothelin signaling pathway -

CCKR signaling map -

Cadherin signaling pathway -

Beta1 adrenergic receptor signaling pathway -

B cell activation -

Axon guidance mediated by semaphorins -

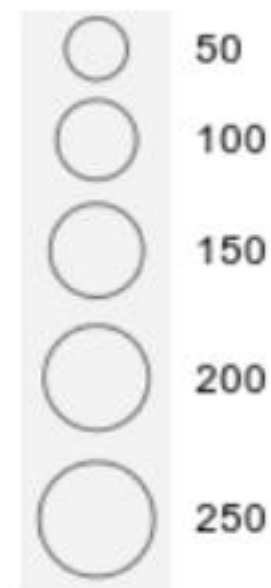
Axon guidance mediated by netrin -

Apoptosis signaling pathway -

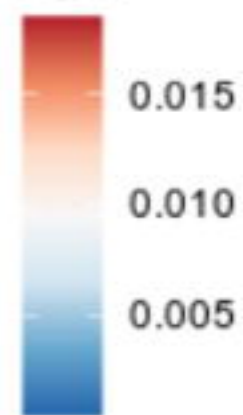
Angiogenesis -

5HT2 type receptor mediated signaling pathway -

Gene.Count



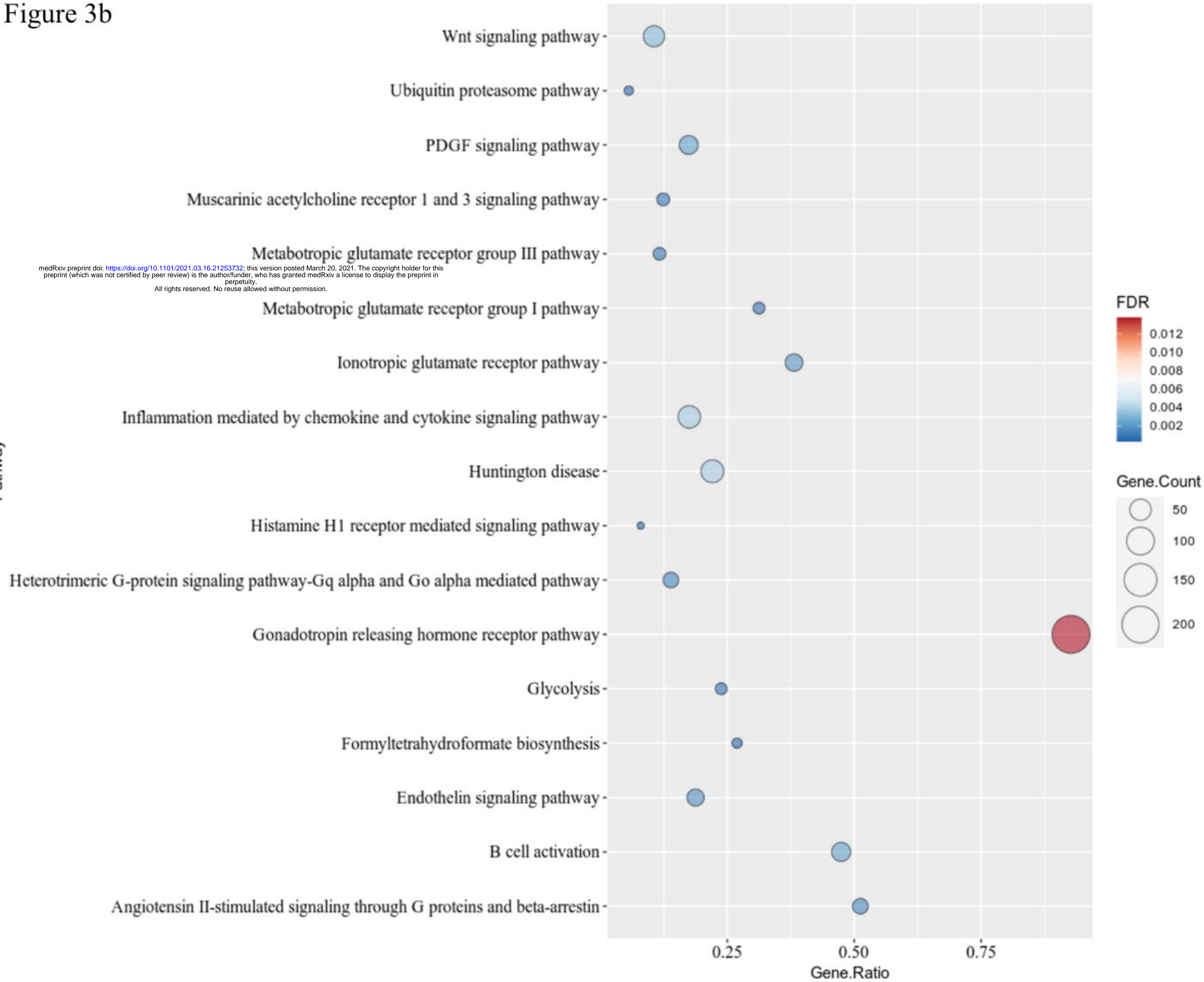
FDR



Gene.Ratio

Pathway

Figure 3b



# Figure 3c

Transcription regulation by bZIP transcription factor

Toll receptor signaling pathway

Salivary gland development  
medRxiv preprint doi: <https://doi.org/10.1101/2021.03.16.21253732>; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.  
All rights reserved. No reuse allowed without permission.

Ras Pathway

PI3 kinase pathway

p53 pathway by glucose deprivation

Notch signaling pathway

Integrin signalling pathway

Insulin/IGF pathway-protein kinase B signaling cascade

Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade

Inflammation mediated by chemokine and cytokine signaling pathway

Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway

General transcription regulation

Fructose galactose metabolism

FAS signaling pathway

Dopamine receptor mediated signaling pathway

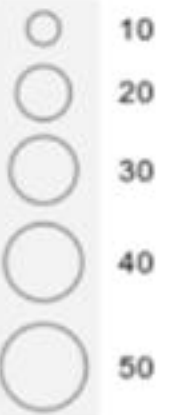
B cell activation

Axon guidance mediated by Slit/Robo

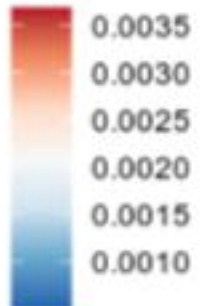
Axon guidance mediated by semaphorins

Axon guidance mediated by netrin

Gene.Count



FDR



Gene.Ratio

Pathway

# Figure 4a

medRxiv preprint doi: <https://doi.org/10.1101/2021.03.16.21253732>; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.  
All rights reserved. No reuse allowed without permission.

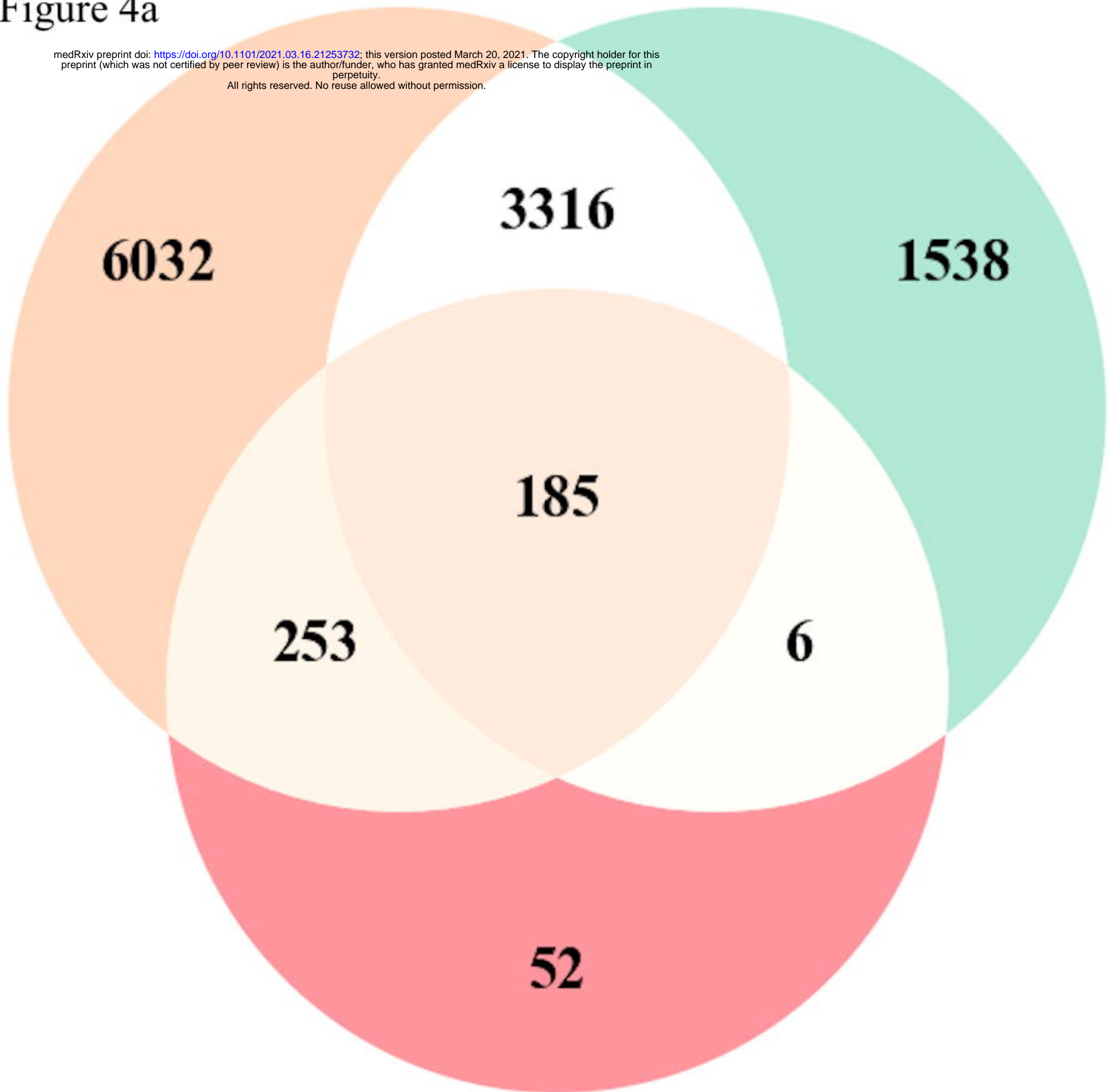
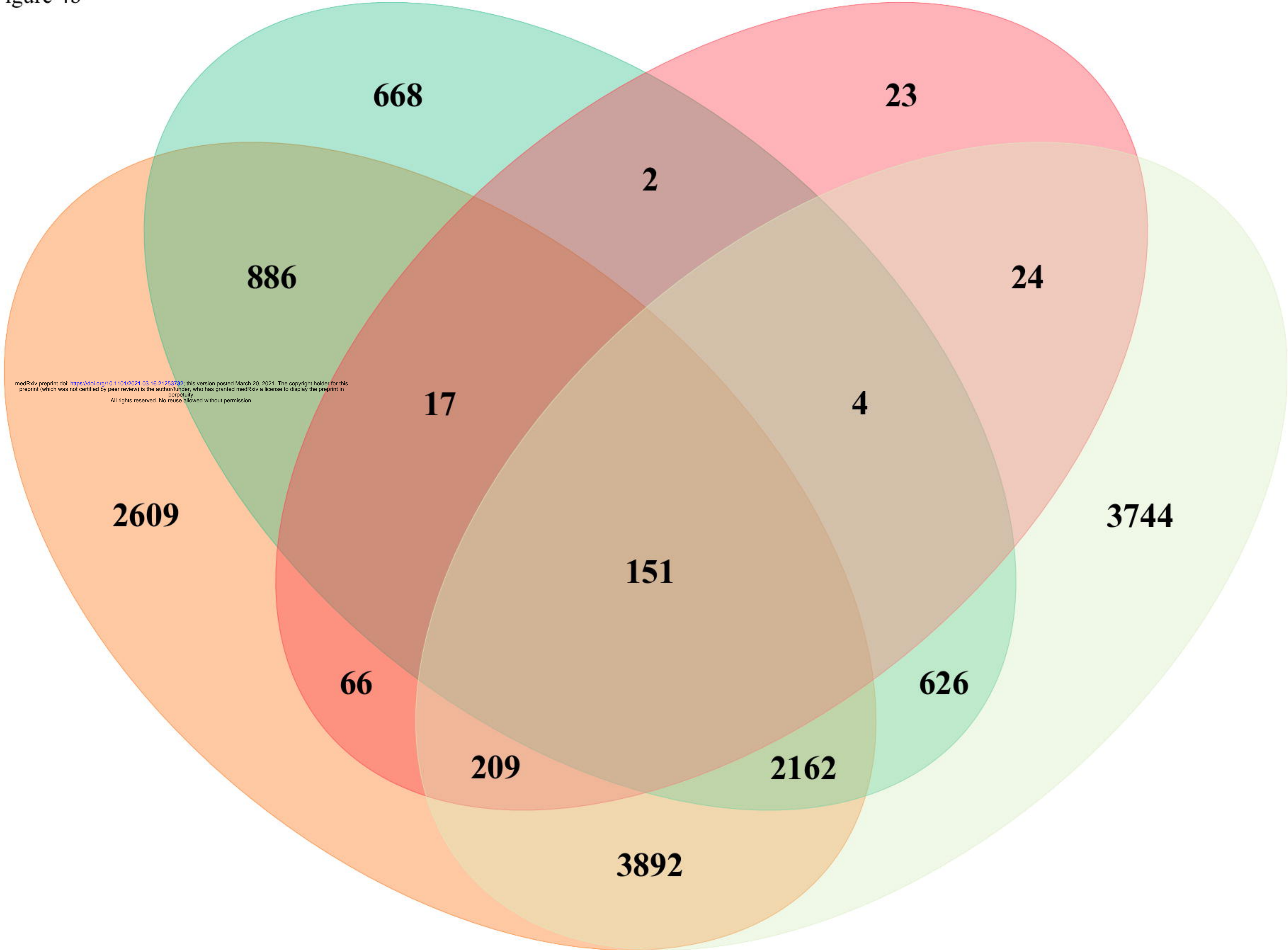


Figure 4b



medRxiv preprint doi: <https://doi.org/10.1101/2021.03.16.21253732>; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

All rights reserved. No reuse allowed without permission.



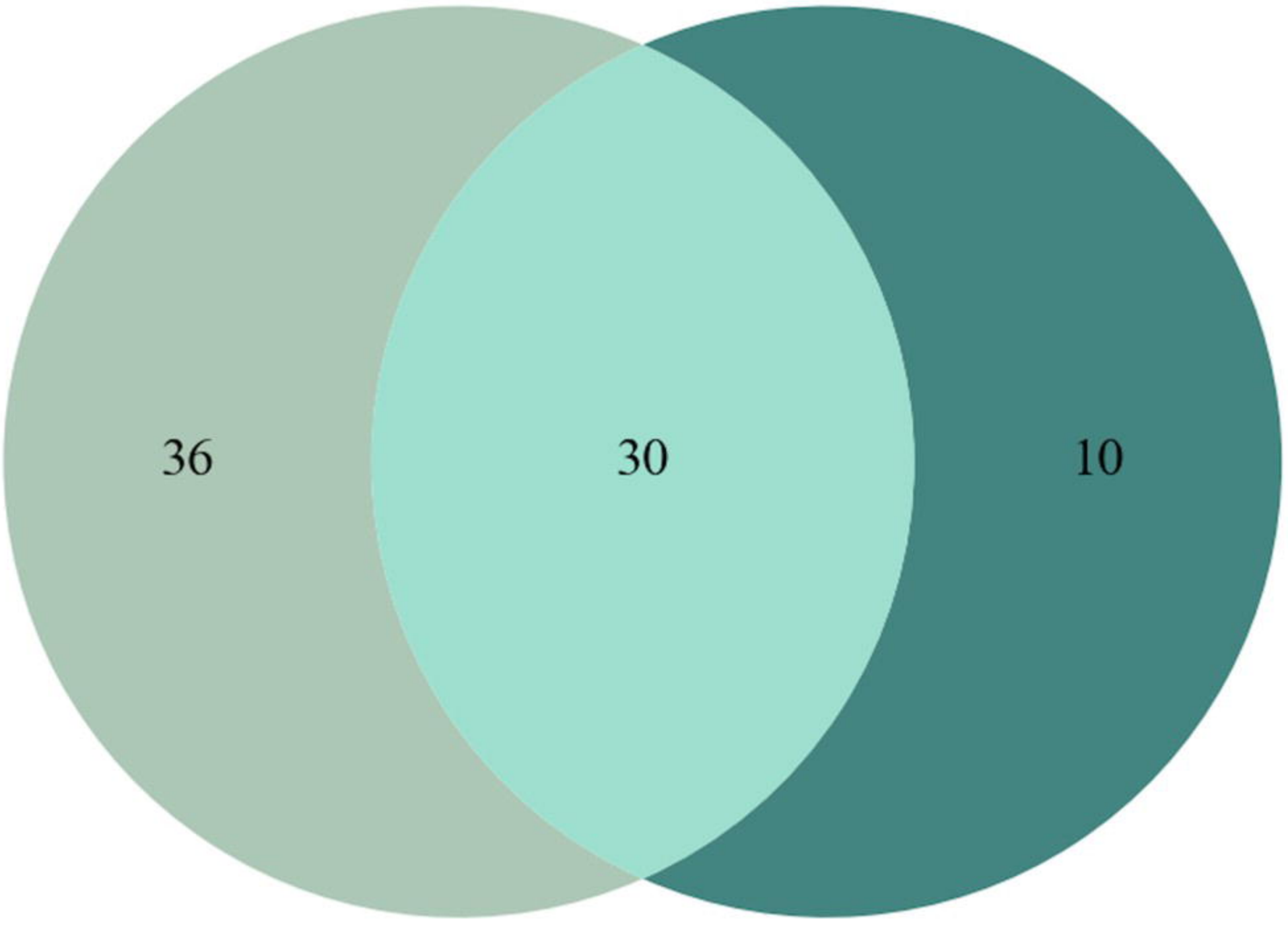






Figure 6

