A blood atlas of COVID-19 defines hallmarks of disease severity and specificity: Associated data

This dataset contains raw and processed data from the COvid-19 Multi-omics Blood ATlas (COMBAT) consortium. Data are divided into 26 datasets representing anonymised raw and processed data from deep immune phenotyping of peripheral blood from COVID-19 patients.

In addition to the data listed below, some datasets are available through other repositories:

- Proteomics data (CBD-KEY-PROTEOMICS) is available at PRIDE
    - Accession number: PDX023175
    - Contact: Roman Fischer
- Genetic data and detailed clinical information are available via a data access agreement through EGA
    - Study accession: EGAS00001005493

For further information regarding specific datasets, please contact the individuals listed below at contact@combat.ox.ac.uk.

Dataset descriptions:

CBD-KEY-CLINVAR: Anonymised patient metadata and associated data dictionary

Contact: Alexander Mentzer

CBD-KEY/RAW-CYTOF-WB(-D): Single cell resolution mass cytometry (CyTOF) data generated from whole blood (WB) and granulocyte depleted whole blood (WB-D). “KEY” dataset contains expression matrices as SingleCellExpression objects, per sample cell counts and frequencies, and composition analysis (Differential abundance) results; “RAW” contains ungated, raw data files in fcs format and the list of antibodies used.

Contact: Giorgio Napolitani

CBD- KEY/RAW-CYTOF-MYELOID: Single cell resolution mass cytometry (CyTOF) data generated from whole blood and granulocyte depleted whole blood, with a focus on myeloid populations. “KEY” dataset contains the processed data files in fcs format and csv files of selected marker expression for monocytes and neutrophils; “RAW” contains ungated, raw data files in fcs format and the list of antibodies used.

Contact: Claudia Monaco; David Ahern

CBD- KEY/RAW-FACS: Flow cytometry data from PBMC cells. “KEY” dataset contains a summary of patient by run, frequencies of populations calculated according to the gating strategy depicted in the analysis workspace file (.wps), and linkers between original sample IDs and FACS file nomenclature; “RAW” dataset contains .fcs files and Flowjo workspaces from FACS data.

Contact: Mariolina Salio

CBD-KEY-RNASEQ-WB: Raw and log normalised gene counts derived from bulk RNA sequencing of whole blood.

Contact: Katie Burnham

CBD-KEY-RNASEQ-WGCNA: Weighted gene correlation network analyses (WGCNA) from whole blood RNAseq data. Contains eigengenes expression data and module gene membership

Contact: Daniel O'Connor

CBD-KEY-CITESEQ-ANNDATA: Contains two anndata objects: (1) "COMBAT-CITESeq-DATA": the raw and normalized gene expression and ADT data, cluster annotations, summary repertoire information and detailed metadata as reported and analysed in this study. (2) "COMBAT-CITESeq-EXPRESSION-ATLAS": raw and normalized gene expression data from an alternative mapping of the data to a transcriptome index that additionally included genes belonging to the "lncRNA" biotype. With ADT data, cluster annotations and a subset of clinical metadata. Suitable for visualization with cellxgene.

A limited version of the 'expression atlas' object is also provided at the CZI cellxgene portal: https://cellxgene.cziscience.com/collections/8f126edf-5405-4731-8374-b5ce11f53e82 for the purpose of online visualisation of the gene expression data.

Contact: Stephen Sansom

CBD-KEY-CITESEQ-GEX-COMPOSITION: Per sample cell counts and frequencies (at different levels of cell subpopulation resolution)

Contact: Calliope Dendrou

CBD-KEY-CITESEQ-GEX-PSEUDOBULKS: Raw and normalised pseudobulk counts generated for each combination of gene and sample at minor subset, major subset and cell type level by summing together the within-group gene counts

Contact: Stephen Sansom; Luke Jostins

CBD-KEY-CITESEQ-GEX-DIFFEXP: Differential expression and pathway enrichment results generated from the CITESeq pseudobulk data (CBD-KEY-CITESEQ-GEX-PSEUDOBULKS)

Contact: Luke Jostins

CBD-KEY-CITESEQ-GEX-WGCNA: Weighted gene correlation network analyses (WGCNA) of selected major cell types. Contains eigengenes expression data, module gene membership, and module names

Contact: Stephen Sansom

CBD-KEY-CITESEQ-VDJ-B: B-cell CITE-seq datasets. Contains receptor sequences (amino acid), summary of clones (by patient).

Contact: Rachael Bashford-Rogers

CBD-KEY-CITESEQ-VDJ-T: T-cell CITE-seq datasets. Contains receptor sequences (amino acid), summary of clones (by patient), diversity scores, kmers, cytotoxicity score.

Contact: Benjamin Fairfax

CBD-KEY-REPERTOIRE-B: B-cell bulk repertoire datasets. Contains B-cell repertoire feature tables including repertoire metrics and V gene usages

Contact: Rachael Bashford-Rogers

CBD-KEY-REPERTOIRE-T: T-cell bulk repertoire datasets. Contains T-cell repertoire feature tables including repertoire metrics and V gene usages

Contact: Benjamin Fairfax

CBD-KEY-ATAC: Peak locations, normalised enrichments, differential expression results, and motif enrichment results generated using single-cell ATAC-seq

Contact: Andew Kwok

CBD-KEY-LUMINEX: Contains raw and normalised expression matrices from Luminex-based assays of blood proteins

Contact: Luzheng Xue

CBD-KEY-SNF: Preprocessed/normalised inputs and results matrices from similarity network fusion (SNF) analysis

Contact: Alberto Santos Delgado

CBD-KEY-ML: Principal components and features for supervised machine learning to classify samples according to their WHO severity.

Contact: Fabian Ruehle

CBD-KEY-SDA: Input data and results from tensor and matrix decomposition (SDA) analyses.

Contact: Justin Whalley

CBD-KEY-SIMON: Processed datasets generated from key data modalities for machine learning analyses using SIMON to distinguish COVID-19 and sepsis

Contact: Adriana Tomic