

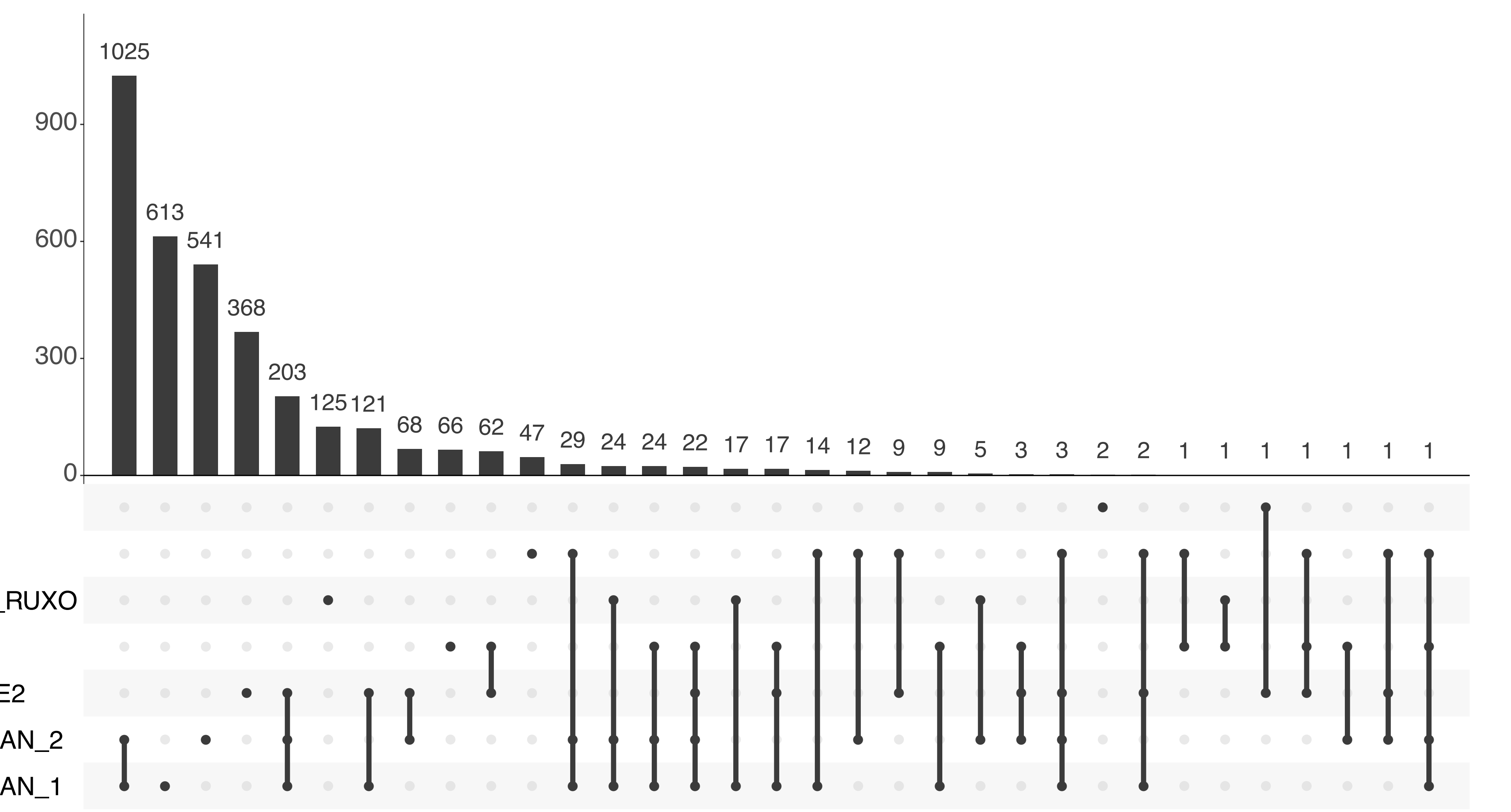
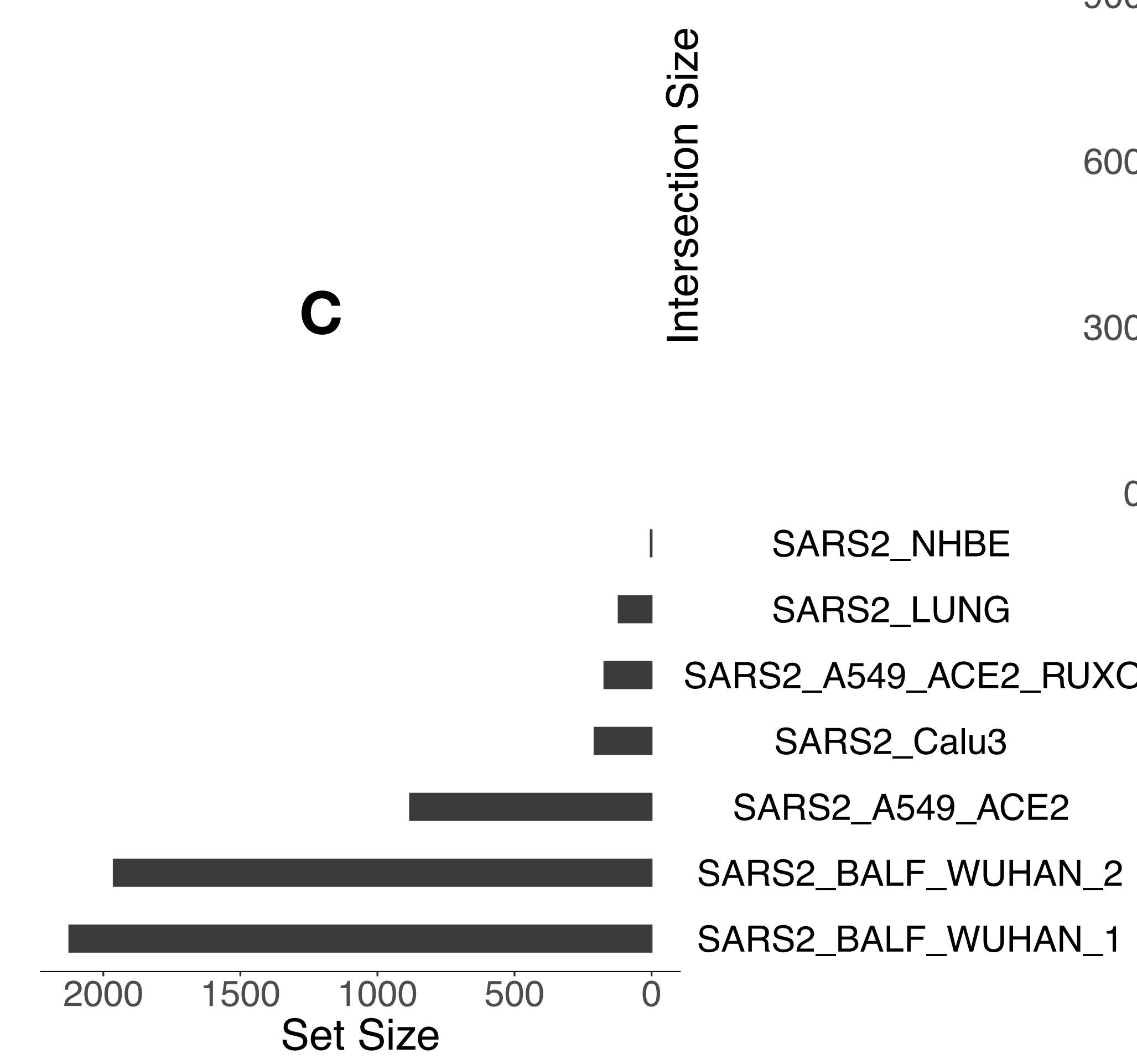
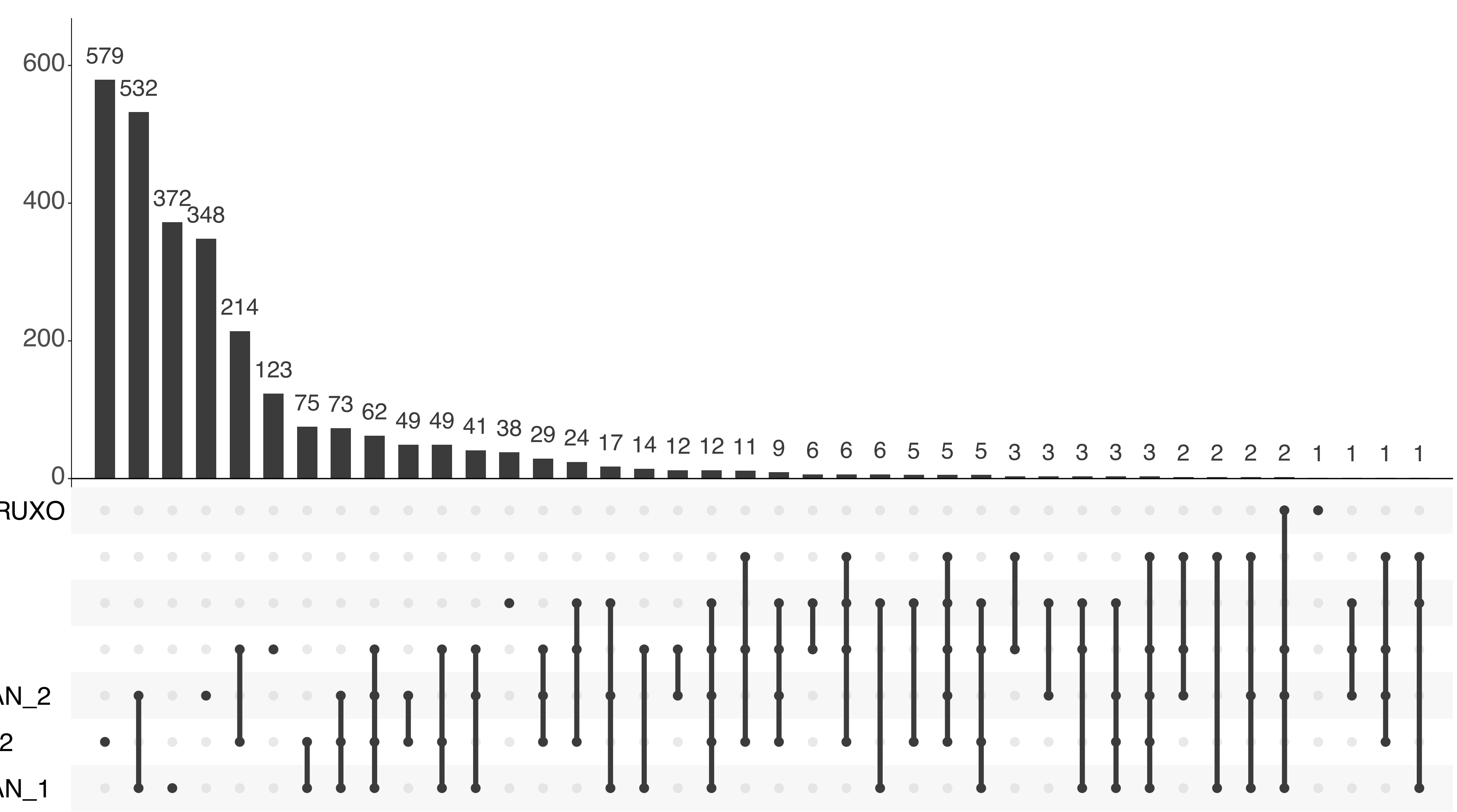
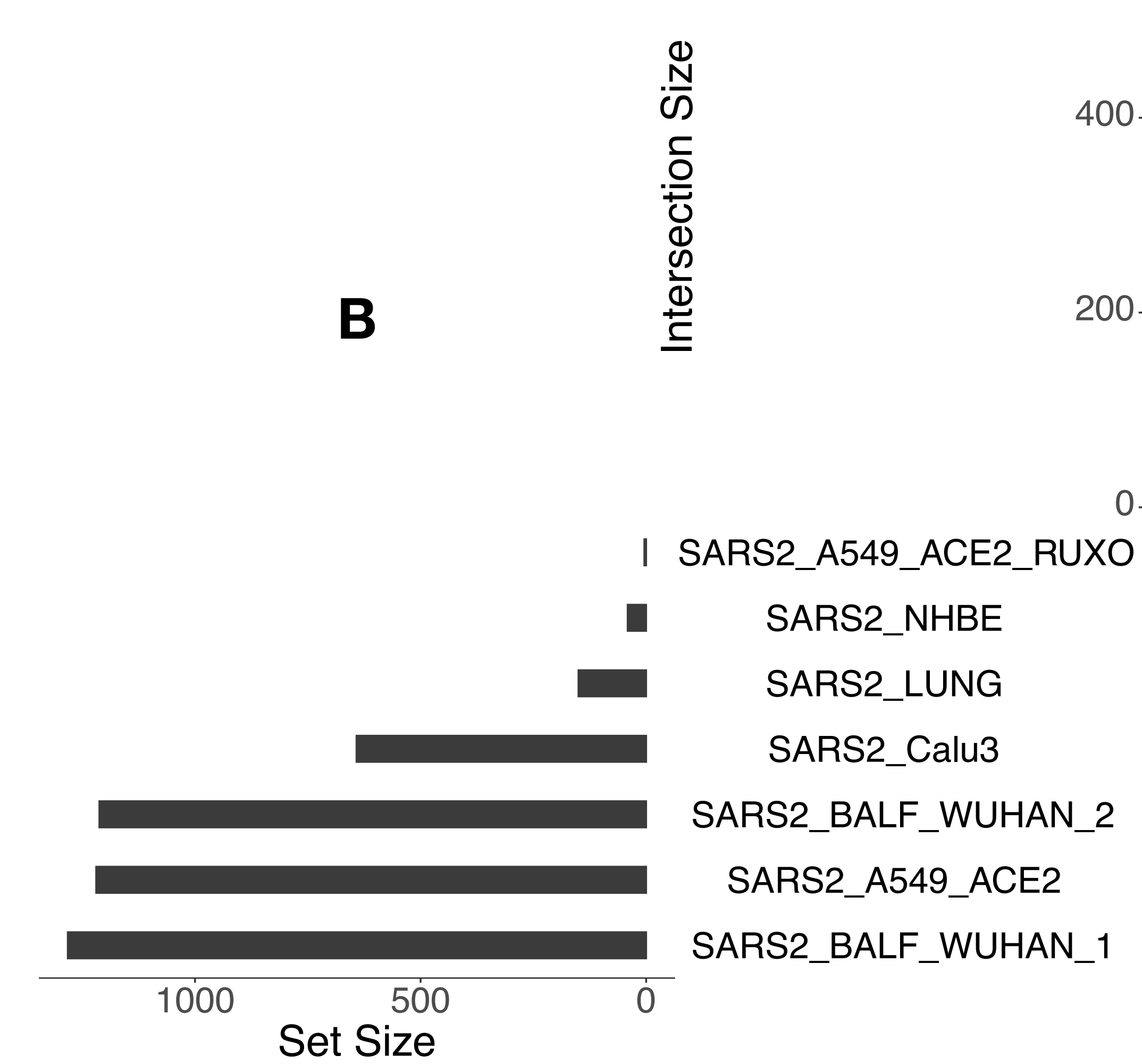
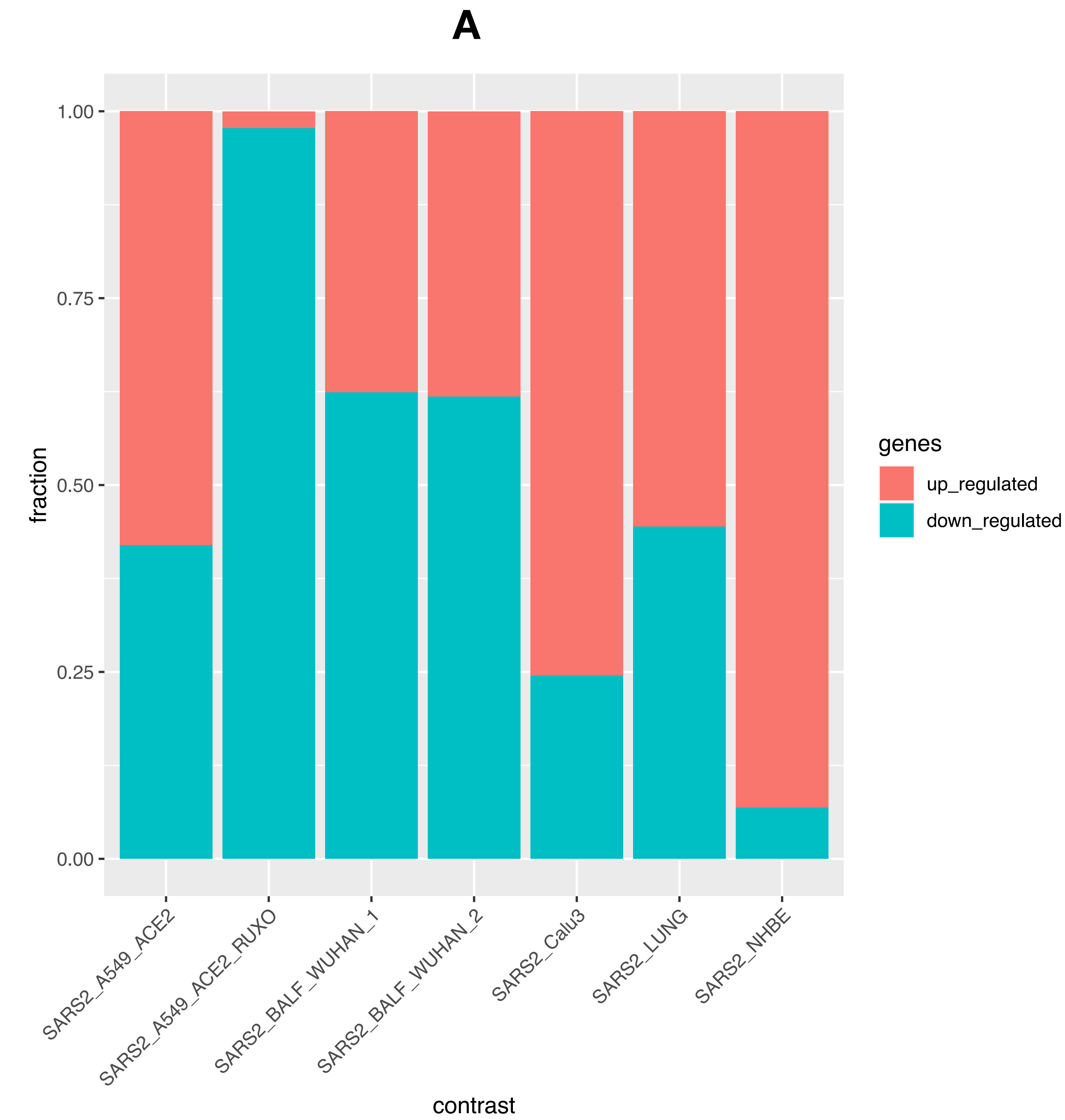
Supplemental Information

Integrative Transcriptome Analyses

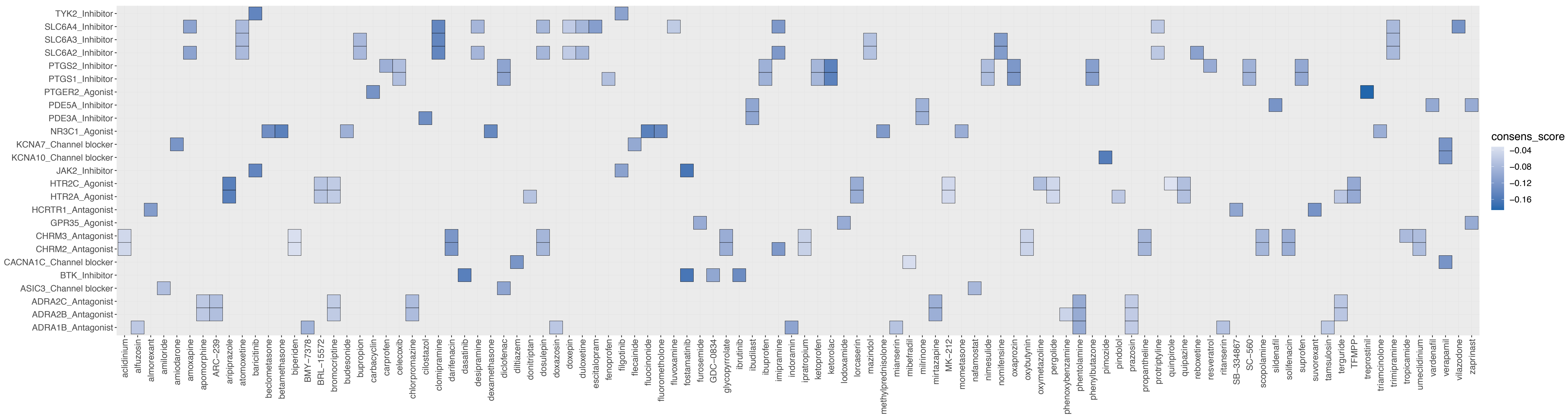
Empower the Anti-COVID-19 Drug Arsenal

Nehme El-Hachem, Edward Eid, Georges Nemer, Ghassan Dbaiho, Ossama Abbas, Nelly Rubeiz, Salah Zeineldine, Ghassan M. Matar, Jean-Pierre Bikorimana, Riam Shammaa, Benjamin Haibe-Kains, Mazen Kurban, and Moutih Rafei

Supplementary Figure 1. Summary of up- and down-regulated genes in each of the SARS-CoV2 contrasts and corresponding ratio of key differentially expressed. Related to Figure 2



Supplementary Figure S2. Top enriched drug-target associations with respect to the ranked cosine consensus drug score. Related to Figure 3



Supplementary Figure 1.

(A) Stacked barplot depicting fractions of up- and down-regulated genes in each of the contrasts. (B) UpSet plots to summarize the ratio of key differentially expressed (DE) genes. This panel shows the overlap of up-regulated genes across all comparisons. (C) Same as for (B) but shows the overlap of down-regulated genes across all comparisons. For all 3 panels, the FDR value of DE significance was set to 0.05 and a $|\log_2FC| > 1$.

Supplementary Figure S2.

Heatplot depicting top enriched drug-target associations against the ranked cosine consensus drug score (putative candidates to reverse SARS2-induced transcriptome perturbations in lung tissue)

Transparent Methods

SARS-CoV-2 datasets

We collected transcriptomic profiles from:

- 1) Cell lines infected with SARS-Cov-2 (Multiplicity of infection: 2) and COVID-19 patients from GEO: GSE147507 dataset. The in vitro setting includes A549 cells supplemented with a vector expressing ACE2, Calu-3 adenocarcinoma cells, and Human Bronchial Epithelial Cells (NHBE). Mock-treated cells were provided for each of the in vitro groups (N=3 per group). For the in vivo setting, two lung samples derived from COVID-19 patients were compared against two biopsied healthy lung tissues. The Ruxolitinib-treated group consisted of A549 cells overexpressing the ACE2 receptor and pre-treated with Ruxolitinib, a Janus kinase 2 inhibitor (JAK2i). A complete description of the dataset can be found in Blanco-Melo *et al.* (Blanco-Melo *et al.*, 2020)
- 2) Bronchoalveolar lavage fluid (BALF) samples from two COVID-19 patients (Xiong *et al.*, 2020). BALF healthy control samples corresponding to non-obese, non-asthmatic patients were downloaded from the SRA database with accession numbers: SRR10571724, SRR10571730, and SRR10571732 (Michalovich *et al.*, 2019)

Differential expression analyses

We conducted a differential expression analysis separately for each of the collected RNA-seq datasets using Deseq2 (Love, Anders and Huber, 2014). Genes with zero counts were removed and shrunk values of the fold change were computed for later use with the gene set enrichment tool (GSEA). A gene is considered differentially expressed between the SARS-CoV-2 setting and its corresponding control/mock group if its adjusted p-value (FDR) is below 0.05.

In vivo comparisons are referred to as SARS2_BALF_WUHAN1-2 and SARS2_LUNG whereas in vitro contrasts appear in the manuscript as SARS2_NHBE, SARS2_A549_ACE2, SARS2_Calu3, and SARS2_A549_ACE2_RUXO.

A SARS2 setting is the difference in relative expression between a model of SARS-CoV2 infection and its corresponding mock/control group.

The Genotype-Tissue Expression (GTEx) lung dataset

RNA-seq gene counts for 374 lung samples were downloaded from GTEx through the recount2 interface (Collado-Torres *et al.*, 2017). Raw data were normalized using the variance stabilizing transformation (VST) technique in DESeq2. Gene co-expression matrix (gene-gene correlation) was constructed using the Pearson correlation distance measure. Genes co-expressed with ACE2 in lung samples were ranked from -1 (opposite gene expression patterns) to +1 (identical expression patterns). This is referred to as ACE2_GTEX in the text.

Gene set enrichment analysis

Gene sets were collected from several sources: MsigDB hallmark set (Liberzon *et al.*, 2015), WikiPathways ('WikiPathways : An Experiment in the Community-based Curation of Biology', 2008) , Reactome (Joshi-Tope *et al.*, 2005), and a custom set associated with SARS-CoV and SARS-CoV-2 from the literature. This custom gene collection, manually curated, is referred here as geneshot_SARS_CoV downloaded from Geneshot (Lachmann *et al.*, 2019), respectively. We run GSEA (gene set enrichment analysis)(Subramanian *et al.*, 2005) with default settings against all lists of differentially expressed genes ranked by the shrunken log2 fold change, from *in vivo/vitro* SARS-CoV-2 and GTEx-ACE2 settings.

Drug-induced gene expression profiles and similarity search

Combined z-scores by biological replicates from LINCS L1000 Phase I (GSE92742) & Phase II (GSE70138) datasets (only small molecule perturbagen) were downloaded from (Qiu *et al.*, 2020). The pipeline processed raw data from L1000 based on the Bayesian approach and a probability-based z-score inference method that showed improved performance over the original L1000 data. We kept L1000 core cell lines (A375, A549, HA1E, HCC515, HT29, MCF7, PC3, VCAP) for further downstream analysis. Signatures corresponding to the same perturbagen but different conditions (cell line, duration, and dose), were averaged using the MODZ method to create a

consensus signature per chemical perturbation (Subramanian *et al.*, 2017). We applied the cosine similarity (Leydesdorff, 2005) to quantify the relationship between the drug-induced gene expression signature from L1000 and the ranked gene signatures from the SARS-CoV-2 groups to identify repurposing candidates. We assessed the significance of the cosine score from 10000 random permutations.

Drug-target databases

We collected and manually inspected drug-target interactions from several sources: DrugBank (Wishart *et al.*, 2008), IUPHAR/BPS Guide to PHARMACOLOGY (Wishart *et al.*, 2008; Southan *et al.*, 2016), and the repurposing hub (<https://clue.io/repurposing>). We conducted a drug set enrichment analysis (drug-target associations are used instead of genesets) to assess the enrichment for known drug-target interactions among the ranked list of drugs from the cosine similarity analysis. Protein targets with a minimum of three representative drugs were kept for the analysis. Drug families enriched for negative associations constitute potential repurposing candidates for COVID-19.

Software and Algorithms		
Deseq2	Love et al, 2014 Genome Biology	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
GSEA	Subramanian et al, 2005	https://www.gsea-msigdb.org/gsea/downloads.jsp