Single-cell transcriptomics reveals a pivotal role of DOCK2 in Sjögren's disease.

Yiran Shen[1], Alexandria Voigt[1], Indraneel Bhattacharyya[2], Cuong Q. Nguyen[1,3,4*]

[1]Department of Infectious Diseases and Immunology, College of Veterinary Medicine, University of Florida. [2]Department of Oral and Maxillofacial Diagnostic Sciences, University of Florida College of Dentistry, [3]Department of Oral Biology, College of Dentistry, [4]Center of Orphaned Autoimmune Diseases, University of Florida, Gainesville, Florida, USA.

***Correspondence author**

Cuong Q. Nguyen, PhD

Department of Infectious Diseases and Immunology

PO Box 110880, College of Veterinary Medicine

University of Florida, Gainesville, Florida 32611-0880 USA

Telephone: 352-294-4180, Fax: 352-392-9704

nguyenc@ufl.edu

Methods

Mice

SjD-susceptible (SjD^s) C57BL/6 J.NOD-*Aec1/2* and non-SjD^s C57BL/6 J (B6) control mice were housed under specific pathogen-free conditions in the animal facilities of the University of Florida Animal Care Services. The breeding and use of animals described herein were approved by and conducted under the direction of the University of Florida Institutional Animal Care and Use Committee. All methods were performed per the relevant guidelines and regulations. The development of C57BL/6.NOD-*Aec1/2* mouse and its SjD-like disease phenotype are described previously (1,2). Briefly, the SjD^s mouse was developed by introducing two genetic regions, one on chromosome 1 (designated *Aec2*) and the second on chromosome 3 (designated *Aec1*) derived from the NOD/LtJ mouse into the B6 mouse. All animals were maintained on a 12-hour light-dark schedule, and food and acidified water were provided ad libitum. At times indicated in the study, mice were euthanized by cervical dislocation after deep anesthetization with isoflurane, and their organs and tissues were freshly harvested for analyses. Utilizing the therapeutic approach, mice aged 28 weeks were treated with a DOCK2 inhibitor, CPYPP (4-[3-(2-Chlorophenyl)-2-propen-1-ylidene]-1-phenyl-3,5-pyrazolidinedione, TOCRIS, Minneapolis, MN). CPYPP blocks DOCK2 by binding to DOCK2 DHR-2 (DOCK homology region 2) domain and inhibits its catalytic activity (1). The mice were chosen at 28 weeks of age due the fact that at this age, mice have developed advanced clinical signs of SjD. Mice were given an initial dose of 100 uL of either 50 mg/mL CPYPP in DMSO or DMSO alone as control via intraperitoneal (IP) injection. Three more DMSO or CPYPP IP injections were given on days 3, 9, and 12 before euthanasia on day 14.

Human samples

Immunofluorescent staining for CD8 and DOCK2 was performed on five sicca control and six SjD patients. Sicca control patients were defined as those with xerostomia but without meeting the criteria for an SjD diagnosis; they were referred to the Oral Medicine Clinic at the University of Florida. Biopsies were obtained as reviewed and approved by the University of Florida's Institutional Review Board. SjD patients were identified by a rheumatologist, having met the criteria outlined by the 2016 American College of Rheumatology/European League Against Rheumatism (3). In brief, the classification criteria are based on the weighted sum of 5 items: anti-SSA(Ro) antibody positivity and focal lymphocytic sialadenitis with a focus score $\geqslant$ 1 foci/mm^2, each scoring 3; an abnormal ocular staining score $\geqslant$ 5 (or van Bijsterveld score $\geqslant$ 4), a Schirmer test $\leqslant$ 5 mm/5 min, and an unstimulated salivary flow rate $\leqslant$ 0.1 mL/min, each scoring 1. Individuals with a total score $\geqslant$ 4 for 5 items meet the criteria for primary SjD. Paraffin-embedded labial salivary gland slides of primary SjD patients were generously provided by The SICCA Biorepository and Data Registry. Available clinical profiles were presented in **Table S1**.

Immunofluorescent staining

Salivary glands from DMSO and CPYPP-treated SjD^s mice were extracted and fixed in 10% phosphate-buffered formalin in a histology cassette for 24 hours. Glands were paraffin-embedded and sectioned at 10μm (Histology Tech Services, Gainesville, FL). Paraffin-embedded biopsy samples were pressure-cooked in Trilogy (Cell Marque, Rocklin, CA) for 5 minutes and 10 minutes for mouse salivary glands. After blocking with donkey sera (1 hour, room temperature), primary staining for human CD8 (Abcam, Cambridge, UK) or mouse CD8 (Santa Cruz Biotechnology, Dallas, TX) with DOCK2 (Bioss, Woburn, MA) was performed (4°C, overnight). The following secondary antibodies (Invitrogen, Waltham, MA) were used for humans: donkey anti-mouse AF594 and donkey anti-rabbit AF488. For mice, these secondary antibodies were used: donkey anti-rat AF594 and donkey anti-rabbit AF 488. Secondary antibodies were each incubated at room

temperature for 1 hour. Images were captured with a Nikon Ti-E fluorescent microscope at 400x magnification. Deconvolution was performed in Nikon NIS Elements. For the enumeration of CD8 T cells, a manual count was performed on a 100x magnification of the field containing an infiltrate, and then ROI intersectional thresholding was used to identify CD8+DOCK2+ cells.

10x Genomics single-cell sample processing and cDNA library preparation

Samples were prepared using the Chromium Next GEM Single Cell V(D)J Kit v1.1, Mouse (10x Genomics, Pleasanton, CA) following the manufacturer's instructions. In brief, sorted single cells of the salivary glands for each sample were resuspended in RPMI containing 10% FBS to a final concentration of 700-1200 cells/µl. A total of 8 samples were loaded onto a Chromium Next GEM Chip G, analyzed by the Chromium Controller (10x Genomics, Pleasanton, CA) for Gel Beads-in-emulsion (GEMs) generation and reverse transcription. The generated cDNA was purified with SPRIselect (Beckman Coulter Inc, Indianapolis, IN) and used for 5′ gene expression library construction. The cDNAs and libraries were examined for quality control using D5000 ScreenTape (Agilent Technologies, Waldbronn, Germany), and Qubit (Thermo Fisher Scientific, Waltham, MA) was used for quantification. To achieve 20,000 reads per cell for 5' gene expression libraries, the libraries were sequenced using Illumina NovaSeq6000 system (Illumina, San Diego, CA).

Measurement of saliva flow

Saliva flow rate (SFR) was recorded prior to drug injection (baseline), then every seven days. Briefly, mice were weighed and given an IP injection of 100 µl isoproterenol (0.2 mg/1 ml of PBS) and pilocarpine (0.05 mg/1 ml of PBS) to stimulate saliva production. Saliva was collected from the oral cavity with a pipet for ten minutes, with a one-minute break at the midpoint. Saliva was briefly centrifuged, and the SFR was calculated as the volume of saliva (uL) per gram (weight of mouse)

Pathological examination of the mouse salivary glands

Salivary glands were fixed in 10% phosphate-buffered formalin for 24 hours. The tissues were paraffin-embedded; sections were cut at a 5-um thickness and mounted onto slides, followed by hematoxylin and eosin (H&E) staining. Stained sections were observed at 200x magnification by using a Nikon Eclipse Ti-E inverted microscope (Nikon, Tokyo, Japan). Focus score were determined by enumerating lymphocytic aggregates of ≥ 50 leukocytes for a single whole salivary gland per mouse.

Detection of antinuclear antibodies

Sera of mice was analyzed for the presence of antinuclear antibodies (ANAs) per the manufacturer's instructions (Immuno Concepts, Sacramento, CA). Briefly, sera were evaluated at 1:40 in PBS and incubated on HEP-2 ANA slides for 30 minutes at room temperature. The secondary antibody, goat anti-mouse IgG AF488 (Invitrogen, Waltham, MA, A11001), was incubated at room temperature on the slide before sealing with Vectashield DAPI medium (Vector Laboratories, Burlingame, CA) and adding a glass coverslip. ANA staining pattern was observed at 400x with a Nikon Ti-E fluorescent microscope with an exposure of 200 ms (Nikon, Tokyo, Japan). Samples positive at 1:40 dilution were further titered for ANA analysis.

Analysis of tissues via flow cytometry

Salivary glands were excised, and single cells were isolated as previously described (4). Cells were rinsed, resuspended in FACS buffer, and stained (30 minutes, on ice) with Live/Dead Fixable Aqua Dead Cell Stain Kit, for 405 emission (Life Technologies, Carlsbad, CA) with either a B or T cell panel as follows: B cells: BV650 rat anti-mouse/human CD45R/B220 (Biolegend, Cat # 103241, San Diego, CA), FITC rat anti-mouse CD23 (Biolegend, Cat # 101605, San Diego, CA), PE rat anti-mouse CD21/CD35 (CR2/CR1) (Biolegend, Cat # 123419, San Diego, CA), AF700 rat anti-mouse IgD (Biolegend, Cat # 405729, San Diego, CA), BV421 rat anti-mouse IgM (Biolegend, Cat # 406517, San Diego, CA); T cells: BV 785 rat anti-mouse CD3 (Biolegend, Cat # 100355,

San Diego, CA), FITC Rat Anti-Mouse CD4 (Biolegend, Cat # 116004, San Diego, CA), APC rat anti-mouse IFN-γ (Biolegend, Cat # 505810, San Diego, CA), PE rat anti-mouse CD8 (BD Pharmingen, Cat # MCD0804, Franklin Lakes, NJ), PE/Cy7 rat anti-mouse IL-4 (Biolegend, Cat # 504118, San Diego, CA), BV 421 rat anti-mouse IL-17A (Biolegend, Cat # 506926, San Diego, CA), and APC-eF780 Mouse Anti-Mouse NK-1.1 (eBioscience, Cat # 47-5941-80, Franklin Lakes, NJ). Samples were run on a BD Fortessa flow cytometer, where 100,000 events were captured; in cases where a full 100,000 events were not available, the entire sample was run. Individual antibody compensations were performed using BD CompBeads (BD Biosciences, Franklin Lakes, NJ); in addition to negative compensation bead control, unstained salivary glands were also utilized to confirm gating strategy. Likewise, paired lymph nodes were used as a lymphocyte pure control to also confirm gating with a more robust cell density. Results were analyzed on FlowJo (FlowJo, Ashland, OR) prior to data processing with GraphPad Prism. For all samples, live lymphocyte populations were first selected. Then T cells were selected for either $CD4^+$ or $CD8^+$ for Th1 and Th17 ($CD4^+$) or Tc1 and Tc17 ($CD8^+$) subsets. FO I were $IgM^-IgD^+CD23^+$, FO II were $IgM^+IgD^+CD23^+$, and MZB were $IgM^+IgD^-CD21^-CD23^+$.

Tissue isolation and cell preparation

Salivary glands of C57BL/6.NOD-*Aec1/2* (51 weeks old, n=2 female, 2 male) and B6 (64 weeks old, n=2 female, 2 male) mice were explanted and digested in a buffer containing 1 mg/ml DNase (Sigma-Aldrich, St. Louis, MO) and 1 mg/ml Collagenase Type 4 (Worthington, Lakewood, NJ, USA) in RPMI (Lonza, Allendale, NJ) complete media (10% FBS, 2 mM L-glutamine, 0.05 mM β-mercaptoethanol). Tissues were placed in a MACS C tube (Miltenyi Biotec, San Diego, CA) for desiccation on GentleMACS V1.02 for a pulse of 38 seconds. After a 10-minute incubation at 37°C, the digest buffer was removed and placed into 4°C RPMI complete media. The process was repeated twice. Single-cell suspensions were centrifuged (2500 rpm, 10 min, 4°C) and resuspended in PBS for filtration through a 70-μm sterile cell strainer (Fisher, Pittsburgh, PA).

After a wash with PBS, cells were resuspended again in PBS for lymphocyte isolation with Lympholyte-M cell separation media (Cedar Lane, Burlington, Canada) per the manufacturer's instructions. Single-cell suspensions were stained for DAPI, and live cells were sorted with a sorter (SH800S, Sony, San Jose, CA) into RPMI containing 10% FBS on ice for single-cell sequencing library preparation.

Single-cell data preprocessing, gene expression quantification, and cell-type determination

The raw data from each sample were demultiplexed and aligned to the GRCm38 reference genome, and the UMI counts were quantified using the 10x Genomics Cell Ranger pipeline (v 7.0.1). Data analysis continued with the filtered barcode matrix files using the Seurat package (v 4.3.0.1). Cells with >200 detected features and <10% mitochondrial reads were considered valid. LogNormalize in Seurat was used for individual samples before merging for downstream analysis to prevent clusters from being biased by differential sequencing depth. FindVariableFeatures was applied to normalize and find variable features within the single-cell gene expression data, with 'vst' as method to choose top 2000 variable features. Clustering and differential expression analyses were performed using the R package Seurat with default parameters. Based on the ElbowPlot, the first 20 principal components (PCs) (1:20) were selected for the clustering analysis when that number reached the baseline of the standard deviation of the PCs. FindNeighbors uses the previously identified PCs was applied to calculate the distance between cells in the high-dimensional space. A resolution of 0.6 was applied in FindClusters function to obtain a meaningful number of subclusters within the major cell types, which was identified through the clustree function. Cell clusters were visualized using Uniform Manifold Approximation and Projection (UMAP). Cells were represented in a 2D UMAP plane with 23 distinct clusters, cells that are in the same large population and share the vast majority of classical immune cell signatures are categorized into five major (T cells (*Cd3d*, *Cd3e*, CD4, CD8a), *Cd68* (macrophage), *Ncr1* (NK),

and *Cd19* (B cells)) and one minor cluster, of which 23 sub-populations were identified and annotated within the macro-population according to known biological cell types using canonical marker genes or published reference gene signatures (5–9).

Differential gene expression was performed using model-based analysis of the single-cell transcriptomics (MAST) test (10) (log fold-change ≥0.25, minimum percentage 0.1, and minimum differential percentage >0.15) to select genes with an adjusted P value <0.01. UpSet (v 1.4.0) was used to make UpSet plots for showing matrix layout of all intersections of the comparison datasets. scRNAtoolVis (v 0.08) was used to make the volcano plot show the differentially expressed gene in certain subclusters of each sex and mouse strain. Pathway enrichment analysis was conducted using Metascape (http://metascape.org) for gene function annotation, and enrichment pathway analysis was used under the default setting. T-like cells ($CD3^+$) were extracted from the global data for downstream analyses to identify T cell subtypes. In the single-cluster enrichment analysis, the FindAllMarkers function in the Seurat package was performed to obtain the rank of all genes ('wilcox', log fold-change ≥0.25, minimum percentage 0.1, and minimum differential percentage >0.15). Then, the fgsea package (v1.17.1) was used to calculate GSEA enrichment scores and P values for each collection of gene sets. Signatures used for subset identity determination or phenotyping already published are referenced in each figure and were converted to corresponding mouse genes for analysis. All the analyses were conducted in the R environment (v.4.3.1).

Gene set variation analysis (GSVA)

Pathway analyses were performed on the 50 hallmark pathways annotated in the molecular signature database (11). The gene sets we used were from the database MSigDB (v7.4). The expression mean matrix between different subclusters was counted using the AverageExpression function as the input for package GSVA (v1.48.2) to calculate the enrich score. To compare the significance of each group, package limma (v 3.56.2) was used to construct a differential contract

matrix and to analyze differences in the enrichment scores for specific groups versus the remaining groups (e.g., group CD8_C4 vs. other CD8 cells).

Trajectory analysis

Package monocle3 (v 1.3.1) was used to estimate the pseudotime path of T cell differentiation. T cells were extracted for Trajectory analysis, and the subset of the data set obtained by Seurat analysis was imported to create a Monocle object. Pseudotime values are assigned to cells using order_cells based on the cell projection on the main graph learned by the learn_graph function and the location of the selected root state. Genes that vary over a trajectory between clusters were identified through graph-autocorrelation analysis [graph_test()] and genes expression trends of the top 5 differential genes were plotted using plot_genes_in_pseudotime, which is colored by subclusters in CD4 and CD8 T cells.

Statistical analysis

Statistical analyses were performed using Prism 8 software (GraphPad, La Jolla, CA). 2-way ANOVA, Welch's t-tests, or Mann-Whitney U tests were performed where indicated. In all cases, p values < 0.05 were considered significant. For the ANA staining, a Chi-squared test was performed.

Results

Table S1: Profile of diagnostic criteria for SjD and sicca controls. Provided here are the results of the autoantibody tests and focus scores performed on patients and controls. Here, pos indicates positive, neg indicates negative, and N/A indicates the test was not performed. All SjD cases met the 2016 ACR-EULAR classification criteria with positive anti-SS-A and focus scores. The one sicca control patient positive for anti-SS-A did not meet any other criteria for SjD.

SjD	Anti-SS-A	Anti-SS-B	ANA (Titer ≥320)	Focus Score
Neg	Pos	Neg	Neg	0
Neg	Neg	Neg	N/A	0
Neg	Neg	Neg	Pos	0
Neg	Neg	Neg	N/A	0
Neg	Neg	Neg	N/A	0
Pos	Pos	Neg	Pos	4
Pos	Pos	Pos	Pos	6
Pos	Pos	Pos	Pos	6
Pos	Pos	Pos	Pos	4
Pos	Pos	Pos	Pos	4
Pos	Pos	Neg	Neg	5

Table S2: Cell processing profiles for individual sample

Sample Name	Loaded Cells (n)	Processed Cells (n)	Median Genes per Cell
B6F1	7,483	7,417	1,302
B6F2	8,643	8,442	1,452
B6M1	1,316	1,155	1,690
B6M2	2,057	1,839	1,708
DCF1	2,528	2,287	1,520
DCF2	2,232	1,946	1,449
DCM1	6,131	5,774	1,538
DCM2	4,189	3,936	1,542

217 **Table S3: Differential gene expression of 20 immune-related and 3 unidentified subclusters**

cell cluster	gene
B_C1	*Ralgps2, Ighd, Fcrl1, Pxk, Pou2af1, B3gnt5, Rasgrp3, H2-Eb2, Vpreb3, Dipk1a, Treml2, Scd1, Cd22, Hs3st1, Snx8, Cpm, Lmo2, Siglecg, Ptp4a3, Snx9*
B_C2	*Gphn, Lars2, AY036118, Egr1, Bank1, Ddit3, Selenow, Ppp1r15a, Foxn3, Mef2c, Stt3b, Gm31243, Rpl36a, Plaur, Ebf1, Tmem123, Ccnl1, Dmxl1, Fcer2a, Cd79a*
B_C3	*Slc7a7, H2-Eb2, Gm10552, Nampt, Cacna1e, Siglecg, Spib, Fam43a, Pkig, AC125149.3, Fcrl5, Dok3, Cxcr5, Pou2af1, Il5ra, Rab30, Plcg2, B3gnt5, Pard3b, Sypl*
B_C4	*Ighm, Igkc, Mucl2, Igkv12-89, Hsp90b1, Plac8, Klf2, Hspa5, Gm10076, Cd69, Celf2, Napsa, mt-Co1, Ly6a, Mzb1, Sp140, Rpl10a, Txnip, Ly6d, Dnajc7*
B_C5	*Pde3b, Tuba1c, mt-Nd4, Foxo1, Malt1, Ago2, Arhgap31, Calm2, Itgb1, Luc7l2, Rpl9, 4930523C07Rik, Tut4, Dazap2, Ubl5, Stk17b, Kmt2e, Arpc3, Macf1, Tomm7*
B_C6	*Btbd9, Bach2, Snx29, Baz2b, Gm47782, Tcf12, Wasf2, Akt3, Kcnq5, Slc12a6, Tmem131l, Plekhm3, Prkce, Sipa1l1, Ptprj, Pax5, Cmip, Cyth1, Ppp3ca, Lrrk2*
Cd4_T_C1	*Cdk11b, Actn1, Nsg2, Sugct, Frat2, Rflnb, Trib2, A930005H10Rik, Ifngr2, Fam241a, Patj, Dgka, Galnt6, Stat5a, Rab3ip, Klk8, Acot2, Ccdc117, Zfand2a, Kif1b*
Cd4_T_C2	*Tspan13, Psme2, Ppp1cc, Fosl2, Atad2, Cxcr3, Icos, Tnfrsf1b, Cd4, Trat1, Arl4c, Sit1, Trac, Lcp2, Sh2b1, Pkp3, Hnrnpll, Fasl, 9-Sep, S100a13*
Cd4_T_C3	*Malat1, Actb, Ifi27l2a, Lgals1, S100a10, Bhlhe40, Mif, Pfn1, Tnfrsf4, Akap13, AU020206, Hspa8, Ppia, Rgs16, Hif1a, Sdf4, Inpp4b, Tox, Npm1, Atp5b*
Cd4_T_C4	*H3f3b, Hilpda, Areg, Gnas, Nfkb1, Rora, Nrip1, Cstb, Rdm1, Phlda1, Mgat5, Dusp5, Rgcc, Cdkn1a, Atp5md, Rgs2, Gm20186, Pim1, Samsn1, Tex14*
Cd8_T_C1	*Coro2a, Lax1, Pdcd1, Klrk1, Fasl, Padi2, Trgv2, Prkcz, Cdh1, Trac, Ifitm10, Cxcr6, Rinl, Itga1, Itgb2, Gm44174, Gimap7, Eif2s3y, Asap2, Sema4a*
Cd8_T_C2	*Ccl5, Ly6c2, Nkg7, Slc3a2, Ctla2a, Tomm5, Prdx6, Gzmk, Eomes, Grap2, Odc1, Smc4, Tigit, Sidt1, Cnn2, Atp5e, Ms4a4b, Simc1, Trbc2, Hnrnpa2b1*
Cd8_T_C3	*Eef1a1, Rps24, Rpl13, Rps15a, Rpl18, Rps7, Rps16, Uba52, Tmsb10, Rps5, Rpl9-ps6, Rps20, Rps4x, Rpsa, Rplp0, Rps3, Camk1d, Rpl30, Rps3a1, Rpl18a*
Cd8_T_C4	*Lncpint, Zeb1, Elmo1, Rabgap1l, Maml2, Esyt2, Themis, Dock2, Epb41l2, Iqgap2, Rras2, Ankrd44, Itpkb, Arid1b, Aopep, Smyd3, Fyn, Vps54, Skap1, Lrba*
Cd8_T_C5	*Ifit1, Rsad2, Isg15, Ifit3, Phf11b, Samhd1, Stat1, Ifi47, H2-T22, Zbp1, Ms4a4b, Smchd1, Isg20, Rtp4, Gbp6, Ifit3b, Ifi208, Zc3hav1, AW112010, Ifi203*

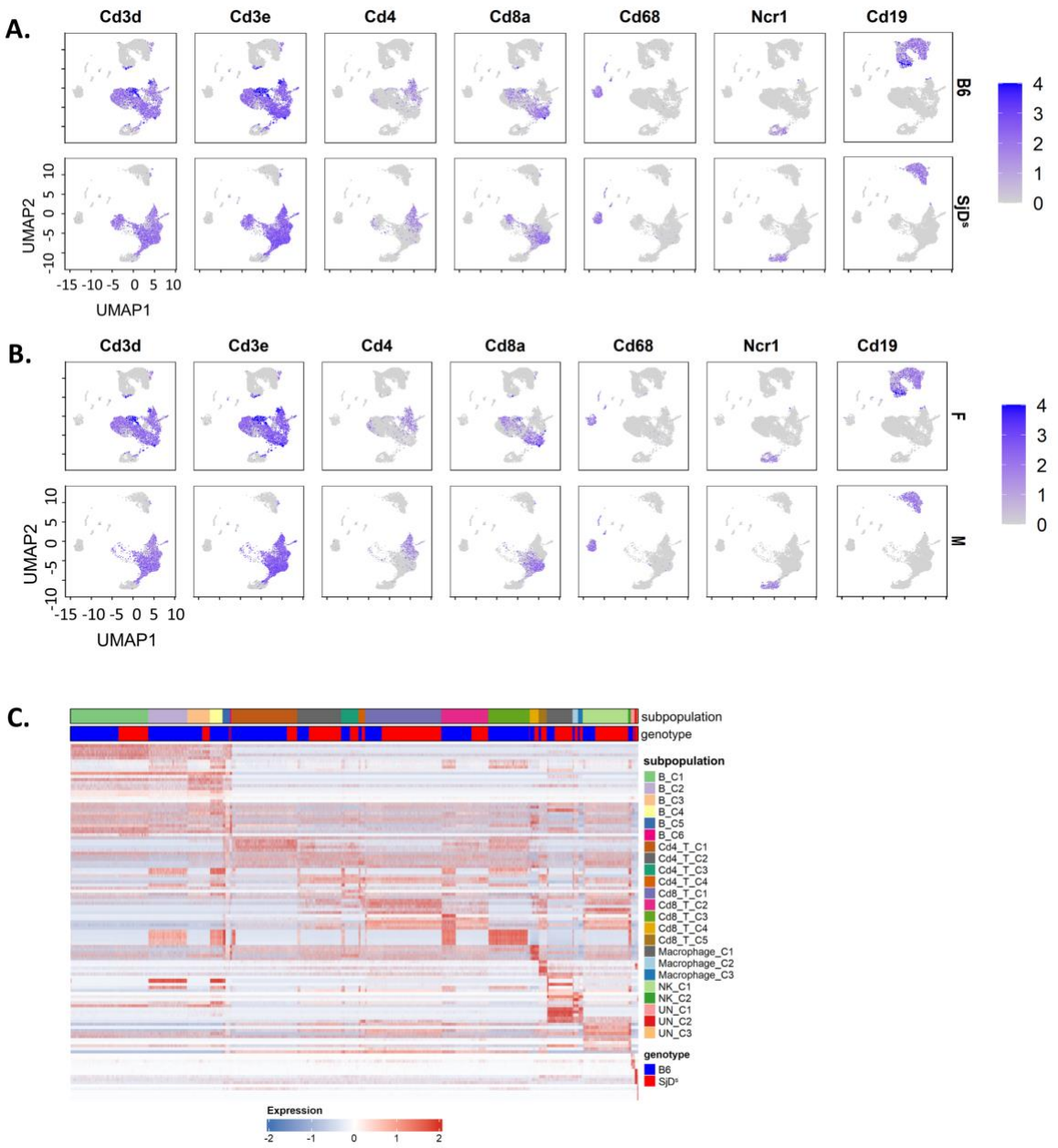
Macrophage_C1	*Ccl4, Cd74, Ccl3, Fth1, Rgs1, Apoe, Plau, Hspa1a, Atf3, C1qb, Ubc, Dusp1, Sat1, Kctd12, Fos, Cd14, Hexb, Dnajb1, Hsp90aa1, C1qc*
Macrophage_C2	*Lyz2, Cst3, Ifitm3, Vim, Ifi30, Tmsb4x, Bst2, Ifitm2, Lgals3, Fabp5, Grn, H2-Aa, Cebpb, Tyrobp, Ccl6, Gm2a, Nfkbia, Dleu2, Tmem176b, Tagln2*
Macrophage_C3	*Aif1, Cd68, Slamf9, Trf, Ckb, Cd300c2, Trem2, Pea15a, Axl, Fcgr3, Zmynd15, Fcgr1, Camk1, Tgfbi, Adgre1, ItgIfngr1, Klra4, Zfp36l2, Litaf, Isy1, Hcst, Klra9, Ncr1, Klrb1b, Klrb1c, Dnajb6, Styk1, Klri2, Clnk, Klrk1, Rin3, Chn2, Ctla2b, Arrdc4, Car2b5, Lilra5, Scimp, Tmem119, Timp2*
NK_C1	*Ifngr1, Klra4, Zfp36l2, Litaf, Isy1, Hcst, Klra9, Ncr1, Klrb1b, Klrb1c, Dnajb6, Styk1, Klri2, Clnk, Klrk1, Rin3, Chn2, Ctla2b, Arrdc4, Car2*
NK_C2	*Gzma, Gzmc, Zfp36, Gzmb, Cd3g, Xcl1, Cd7, Il2rb, H2-D1, Jun, Irf8, Fcer1g, Dusp2, Arf4, Junb, Ctsd, Ctsw, Klra8, Sh2d2a, Il21r*
UN_C1	*Klk1, Scgb2b27, Klk1b26, Fxyd2, Bglap3, Klk1b9, Klk1b5, Crisp3, S100a1, Chchd10, Tfcp2l1, Serp1, mt-Nd4l, Klk1b11, Phyh, mt-Atp8, Fxyd3, Cox7c, Klk1b22, Mdh1*
UN_C2	*Alas2, Hba-a1, Ube2l6, Fech, Bpgm, Snca, Apol11b, Gypa, Prdx2, Isg20, Car2, 2-Mar, Slc4a1, Prxl2a, Slc25a39, Ube2c, Blvrb, Epb41, Slc25a37, Bnip3l*
UN_C3	*Mgp, Ly6c1, Clu, Ltbp4, Tm4sf1, Ptprb, Egfl7, Eln, Pecam1, Cldn5, Ptprm, Slc9a3r2, Cdh13, Igfbp7, Cytl1, Ramp2, Ldb2, Timp3, Bcam, Fbln2*

Table S4: Top 5 differentially expressed genes within B, CD4$^+$, and CD8$^+$ T cell subsets

cell cluster	differentially expressed genes				
B_C1	*Fcer2a*	*Pxk*	*Pkib*	*Sell*	*Lrrk2*
B_C2	*Slc15a2*				
B_C3	*CD44*	*Atf3*	*A530032D15Rik*	*Arhgap24*	*Cd9*
B_C4	*Igkv12-89*	*Apoe*	*Vim*	*Odc1*	*Igkv4-63*
B_C5	*Ccl5*	*Cd3e*	*Itk*	*Lat*	*Emb*
B_C6	*Akt3*	*Arhgap26*	*Elmo1*	*Tcf12*	*Lncpint*
CD4_T_C1	*Lef1*	*Ccr7*	*Satb1*	*CD8b1*	*Igfbp4*
CD4_T_C2	*Ccl5*	*S100a6*	*Itgb1*	*Ahnak*	*S100a4*
CD4_T_C3	*Tnfsf8*	*Bhlhe40*	*Ctla4*	*Tnfrsf4*	*Rgs16*
CD4_T_C4	*Rora*	*Il17a*	*Tmem176b*	*Tmem176a*	*Fosb*
CD8_T_C1	*S100a6*	*Xcl1*	*Cxcr6*	*Litaf*	*Coro2a*
CD8_T_C2	*Ccl5*	*Ly6c2*	*Sidt1*	*Eomes*	*Pde2a*
CD8_T_C3	*Ccr7*	*Klf2*	*Lef1*	*Dusp10*	*Satb1*
CD8_T_C4	*Zswim6*	*Lncpint*	*Elmo1*	*Maml2*	*Slc9a9*
CD8_T_C5	*Ifit3*	*Isg15*	*Ifit1*	*Rtp4*	*Zbp1*

Table S5: Enumeration of infiltrating cell types via flow cytometry. After exporting counts from Flow Jo, GraphPad Prism v8 was utilized to calculate the average +/- SEM for each cell type in the DMSO and CPYPP treated SjD^S mice. *$p \leq 0.05$ by Mann-Whitney one-tailed t-tests.

	DMSO		CPYPP	
Cell type	**Avg ±**	**SEM**	**Avg ±**	**SEM**
CD4	530.4 ±	315.8	106.4 ±	84.7
CD8	181.4 ±	97.1	15.2 ±	4.3*
Th17	11.6 ±	5.5	3.4 ±	1.2*
Th1	2.4 ±	1.5	7 ±	5.1
Tc17	10.6 ±	5.7	1.8 ±	0.8
Tc1	25 ±	7.5	8.6 ±	3.2*
MZB	3.8 ±	1.7	1.6 ±	0.7
FO I	28.8 ±	5.2	29 ±	18.5
FO II	1.6 ±	0.5	0.8 ±	0.4*
FO	30.2 ±	5.3	29.8 ±	18.6



229

230 **Figure S1. Cellular composition differences by sex and disease phenotypes. (A, B)** UMAP
231 plots show the expression levels of selected marker genes in different clusters compared between
232 genotype **(A)** and sex **(B)**, with colors representing clusters expressing the genes and color
233 densities representing different levels of selected gene expression. **(C)** Comprehensive gene
234 expression heatmap across different subpopulations and with genotype distribution.

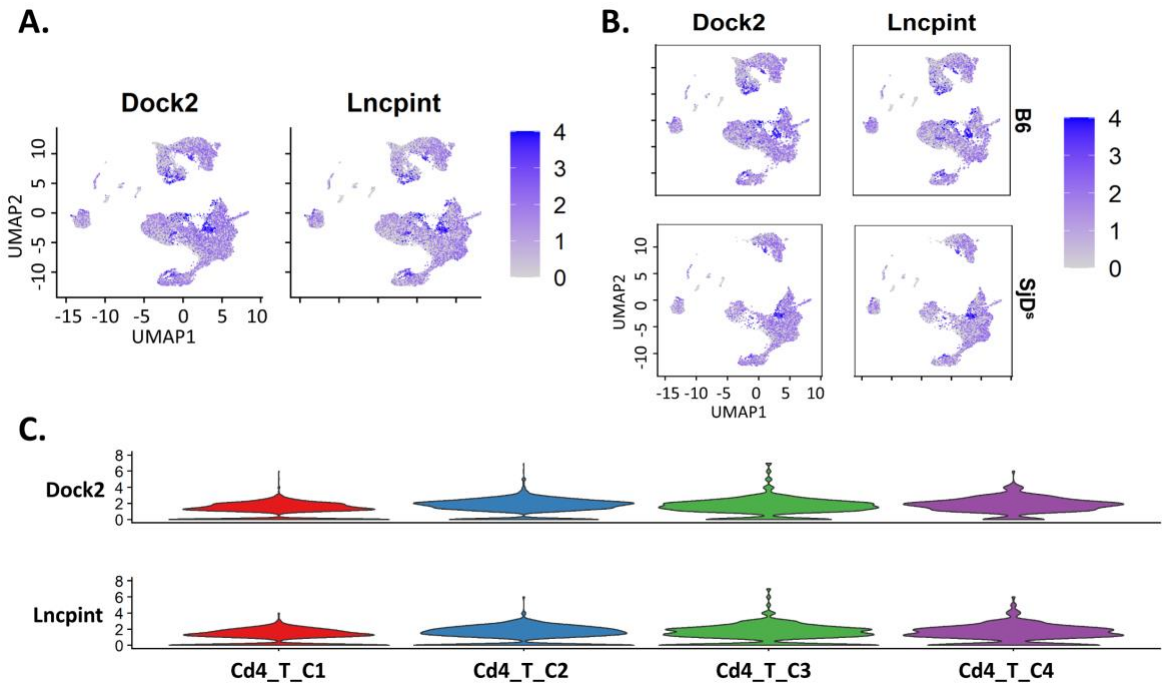
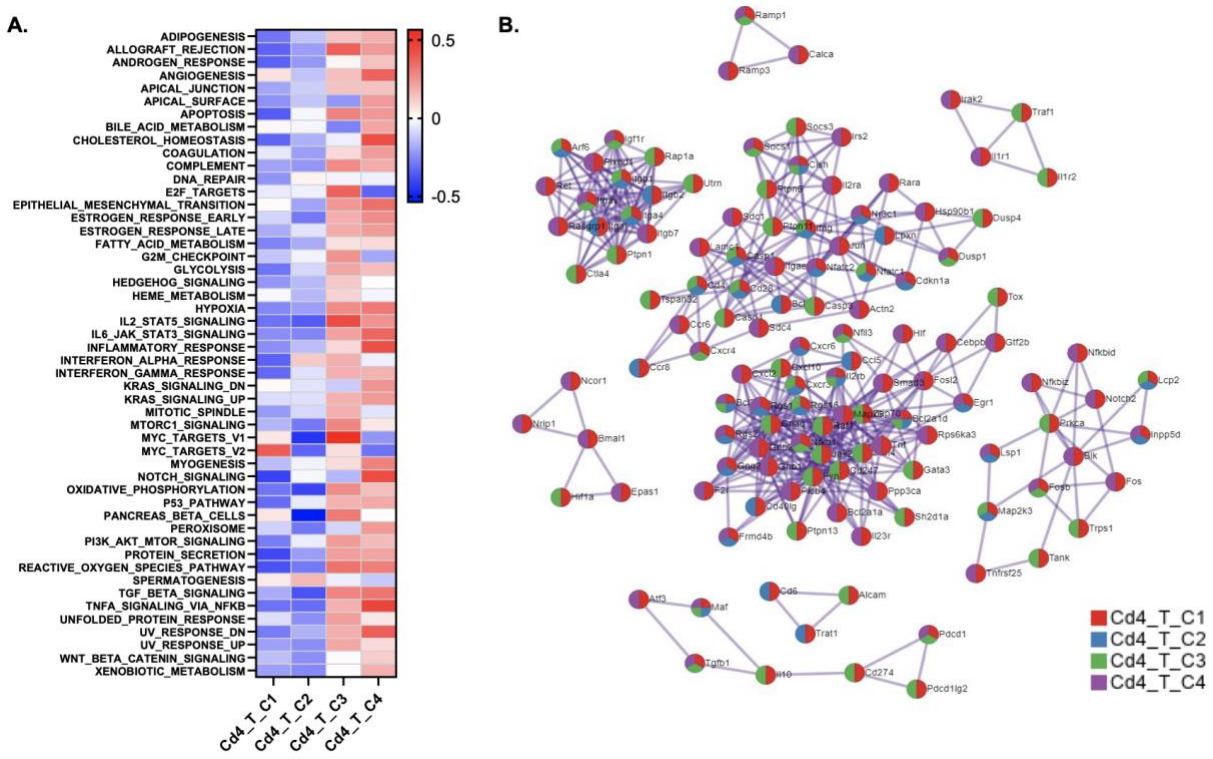


Figure S2. *Dock2* and *Lncpint* transcribed at varying levels across all immune cell subsets

(A, B) UMAP plots show the expression levels of *Dock2* and *Lncpint* in different clusters overall **(A)** and compared between genotype **(B)**, with colors representing clusters expressing the genes and color densities representing different levels of selected gene expression. (**C**) Violin plots showing expression levels of *Dock2* and *Lncpint* in four different subpopulations of $CD4^+$ T cells.



241

242 **Figure S3. Additional genetic networks analysis for CD4⁺ T cells.** (**A**) Genomic variation
243 analysis (GSVA) of the mouse MSigDB 50 hallmark gene sets in each Cd4 T cell subset by p-
244 value. (**B**) Protein-protein interaction networks (PPI) of differentially expressed genes (DEGs) in
245 each Cd4 T cell subcluster, to which the Molecular Complex Detection (MCODE) algorithm is then
246 applied to identify densely connected neighborhoods of proteins. Network nodes are shown as
247 pies. Different pie area colors represent individual gene lists.

248

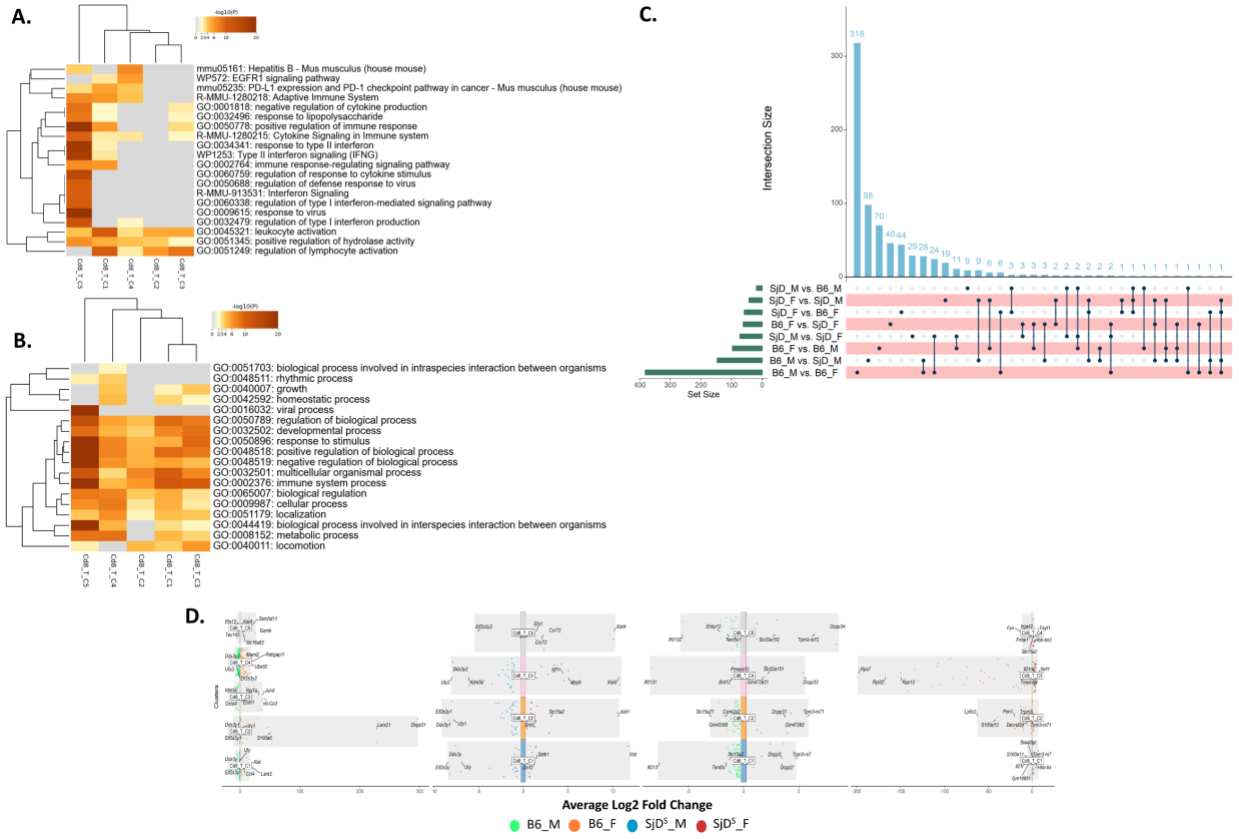


Fig. S4. Biological pathways of CD8$^+$ effector T cells in the salivary glands of SjD mice. (**A**) Significantly enriched pathway and (**B**) biological process of individual T cell subset of CD8$^+$ T cells. (**C**) UpSet plots show a matrix layout of all intersections of the eight comparison datasets in CD8$^+$ T cells (by each genotype and gender, separately) and sorted by size. The size of each comparison dataset was indicated as a green bar on the left, showing the number of up-regulated genes in the right dataset compared to the left dataset. Dark circles in the matrix indicate sets of genes with corresponding intersections, where the number of genes in each set is shown above the blue bar graph. The connecting lines indicate the comparison sets that share this gene set (two sets or more). In the B6 background, male mice have 318 upregulated DEGs compared to females and 98 upregulated DEGs compared to SjD^S genotype. 28 genes were shared among this two-comparison dataset, which means male B6 mice have a more distinct gene expression profile. (**D**) Volcano plot of DEGs between paired experimental groups of individual CD8$^+$ T cell

262 subpopulations. Different colored dots represent DEGs in each group and are indicated in the
263 legend below the figure. The top 3 genes in each group (if more than 3 genes were identified)
264 were labeled in the figure.

265

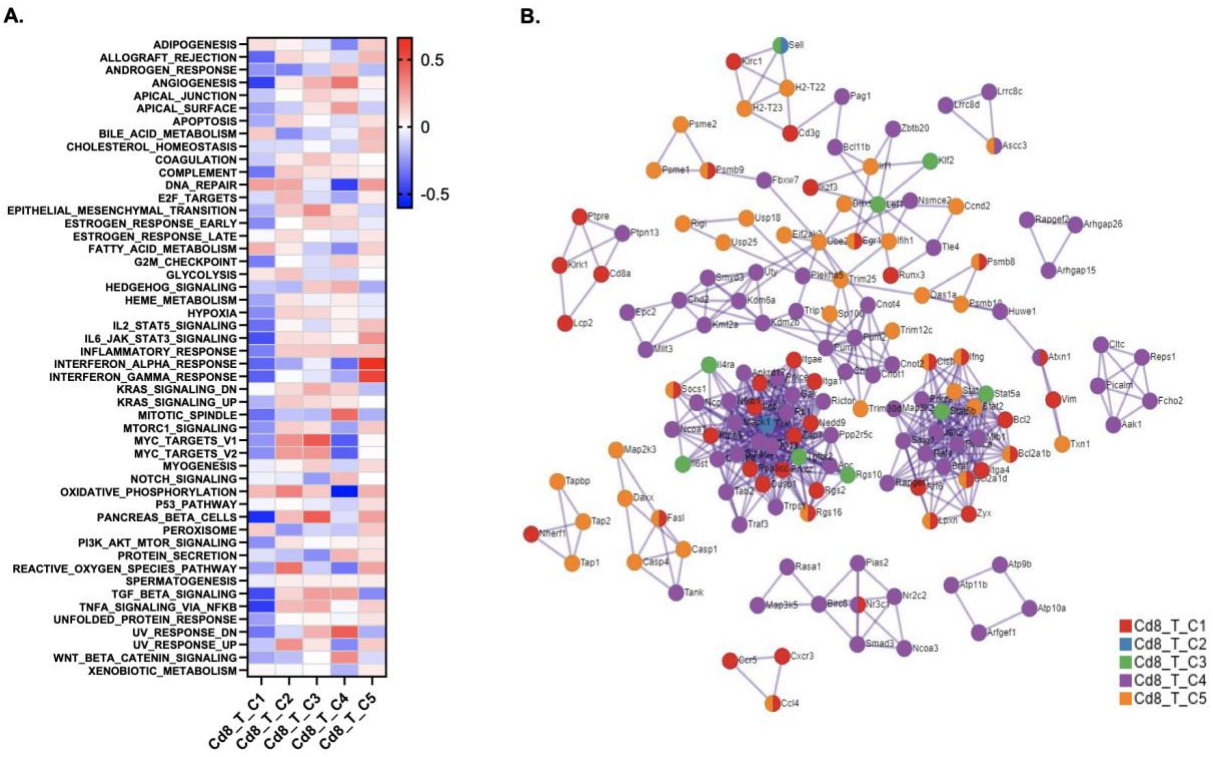
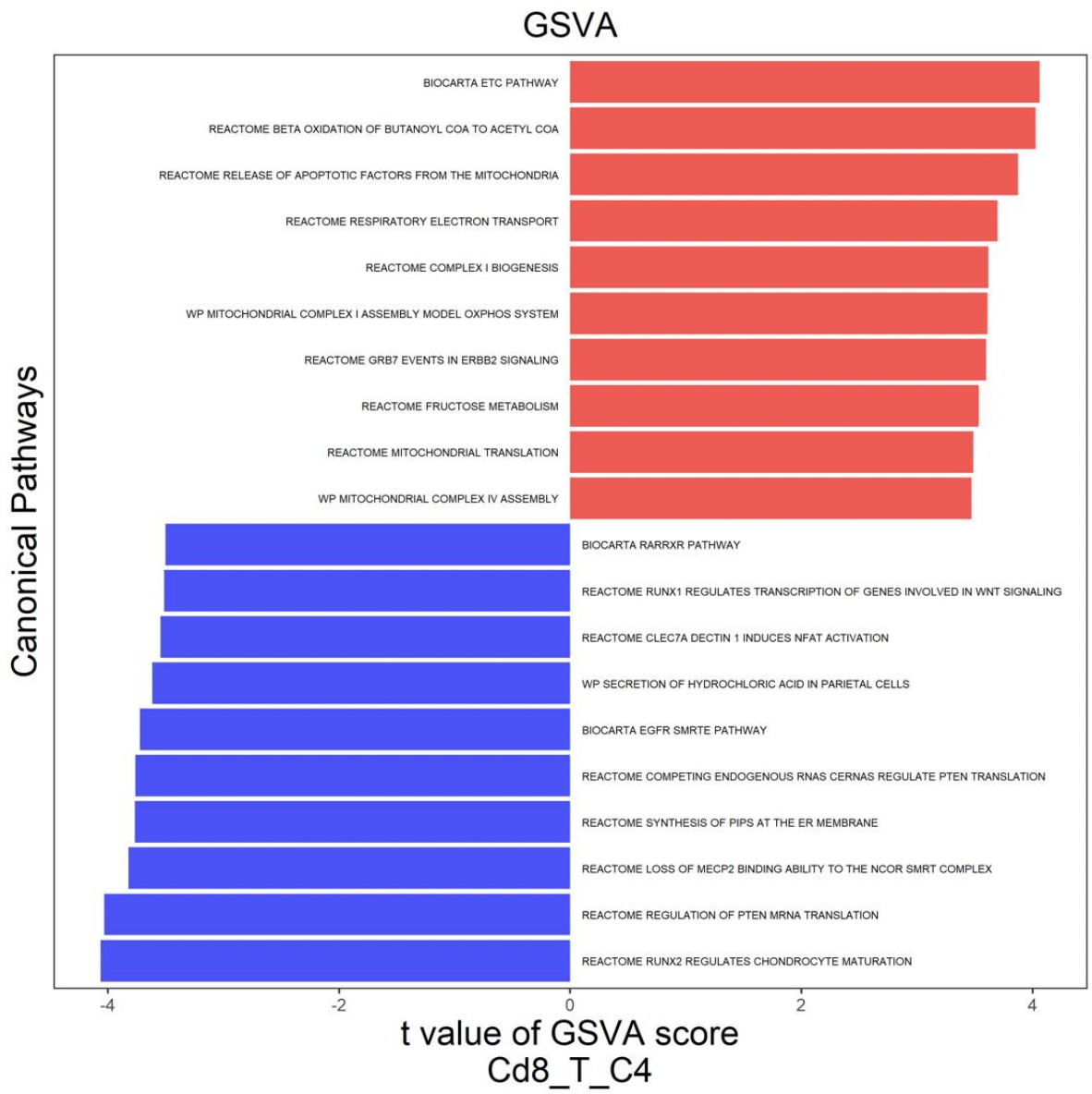


Figure S5. Additional genetic network analysis for CD8⁺ T cells. (**A**) Genomic variation analysis (GSVA) of the mouse MSigDB 50 hallmark gene sets in each Cd8 T cell subset by p-value. (**B**) Protein-protein interaction networks (PPI) of differentially expressed genes (DEGs) in each Cd8 T cell subcluster, to which the Molecular Complex Detection (MCODE) algorithm is then applied to identify densely connected neighborhoods of proteins. Network nodes are shown as pies. Different pie area colors represent individual gene lists.



275

276 **Figure S6**. **Genomic variation analysis (GSVA) of the unique Cd8_T_C4 subpopulation**

277 **compared to other CD8⁺ T cell subpopulations**.

278

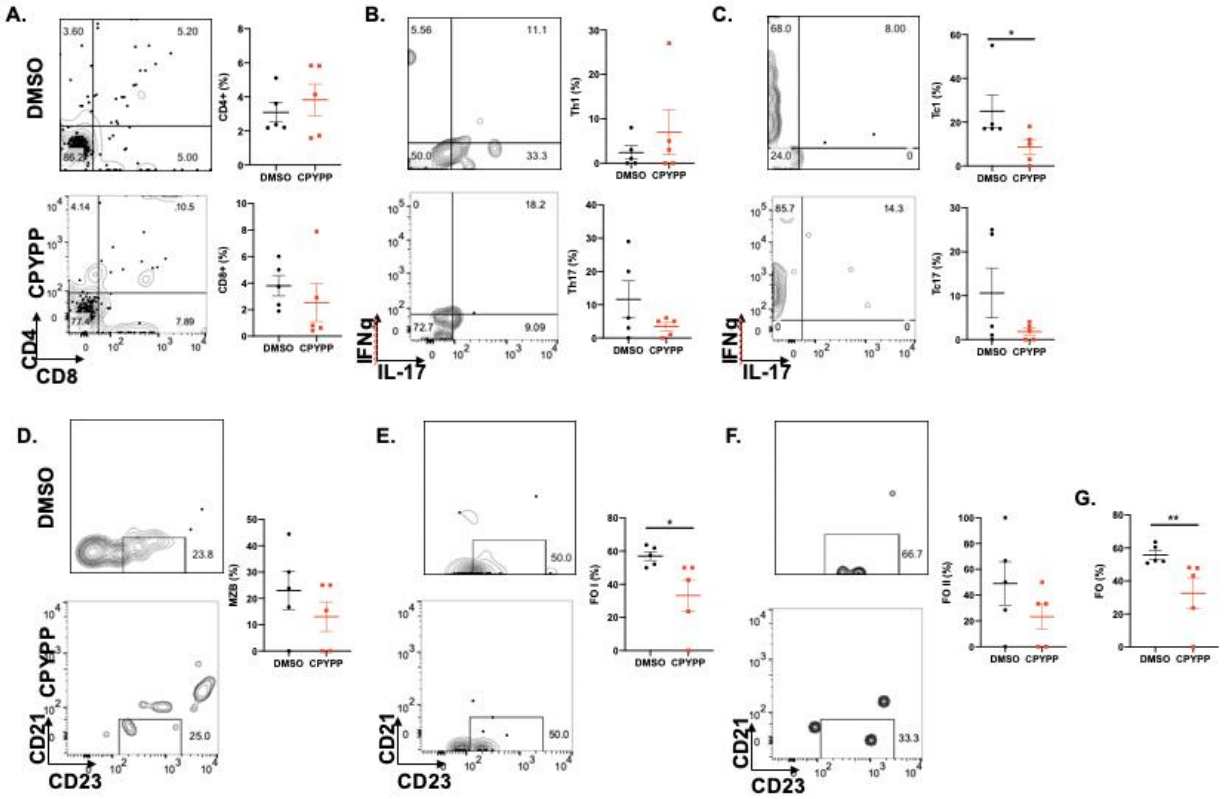


Figure S7. The effect of DOCK2 inhibitor on the infiltrating cells in the salivary gland. Flow cytometric analysis was performed for $CD4^+$ and $CD8^+$ T cells (**A**), Th1 and Th17 cells (**B**), Tc1 and Tc17 cells (**C**), MZB (**D**), follicular B cell type I (**E**), follicular B cell type II (**F**), and follicular B cell (**G**). SjD^s mice were treated with CPYPP at 28 weeks of age. Mice were given an initial dose of 100 uL of either 50 mg/mL CPYPP (n=5) or DMSO alone (n=5) as control via IP injection. Three more DMSO or CPYPP IP injections were given on days 3, 9, and 12. Mice were monitored for two weeks. Representative flow cytometric images for each analysis were shown. A one-tailed Mann-Whitney test was performed with error bars, *p< 0.05 and **p< 0.01.

References

1. Brayer J, Lowry J, Cha S, Robinson CP, Yamachika S, Peck AB, et al. Alleles from chromosomes 1 and 3 of NOD mice combine to influence Sjögren's syndrome-like autoimmune exocrinopathy. *J Rheumatol* 2000;27:1896–1904.

2. Cha S, Nagashima H, Peck AB, Humphreys-Beher MG. IDD3 and IDD5 alleles from nod mice mediate Sjögren's syndrome-like autoimmunity. *Adv Exp Med Biol* 2002;506:1035–1039.

3. Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, et al. 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: A consensus and data-driven methodology involving three international patient cohorts. *Ann Rheum Dis* 2017;76:9–16.

4. Wanchoo A, Voigt A, Sukumaran S, Stewart CM, Bhattacharya I, Nguyen CQ. Single-cell analysis reveals sexually dimorphic repertoires of Interferon-γ and IL-17A producing T cells in salivary glands of Sjögren's syndrome mice. *Sci Rep* 2017;7:12512.

5. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* 2018;564:268–272.

6. Leruste A, Tosello J, Ramos RN, Tauzième-Espariat A, Brohard S, Han Z-Y, et al. Clonally expanded T cells reveal immunogenicity of rhabdoid tumors. *Cancer Cell* 2019;36:597–612.e8.

7. Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ, et al. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell* 2019;176:775–789.e18.

8. Cano-Gamez E, Soskic B, Roumeliotis TI, So E, Smyth DJ, Baldrighi M, et al. Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nat Commun* 2020;11:1801.

9. Ren X, Wen W, Fan X, Hou W, Su B, Cai P, et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* 2021;184:1895–1913.e19.

10. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 2015;16:278.

11. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–15550.