Supplemental Information for:

Early microbial exposure shapes adult immunity by altering CD8+ T cell development

Cybelle Tabilas, David S. lu, Ciarán W.P. Daly, Kristel J. Yee Mon, Arnold Reynaldi, Samantha P. Wesnak, Jennifer K. Grenier, Miles P. Davenport, Norah L. Smith, Andrew Grimson, and Brian D. Rudd*

*Corresponding author

This PDF file includes:

Supplemental Materials and Methods

Mice

Unless otherwise stated, mice were used at 8 weeks of age. All mice that survived the four-week microbial exposure period were used for timed matings. Clean mice were maintained under specific pathogen-free conditions at the College of Veterinary Medicine and mice exposed to pet-shop material were maintained in a separate facility.

Timestamping method

Timestamp mice were generated by crossing a TCRδCre-ERT2 mouse with a ZsGreen reporter mouse. To activate fluorescent reporter expression in fetal-derived T cells, 2.5 mg of tamoxifen was administered to dams by oral gavage three times in 12-hour intervals over a 24-hour period and 0-1d old pups received tamoxifen through lactation. To activate fluorescent reporter expression in adult-derived T cells, 2.5 mg was administered to mice by oral gavage in 24-hour intervals for two days.

Thymic transplants

Thymic transplants were performed as previously described[1]. Briefly, thymus lobes were isolated from 0-1d old timestamp mice. The thymus lobes were separated into individual lobes and were placed under the kidney capsule of a 6-week-old mouse. To mark thymocytes from the donor thymus lobe, 5 mg of tamoxifen was administered through oral gavage from 0-3d post-transplantation.

Manipulating developmental layers

To prevent the formation of the adult layer, 2.5-3 wk old mice were anesthetized using 1.2 mg of avertin and thymus lobes were removed through aspiration. A sham thymectomy was a complete surgery without thymic removal. At 8 wks of age, mice were visually inspected for the presence of thymic lobes and mice that had complete removal of both thymic lobes were included in

experiments. To 'delete' the fetal layer, 2 wk old mice received 0.1 mg of a CD8a mAb (BioX Cell) intraperitoneally every 24 hrs for two days.

Adoptive transfer and pathogen burden

CD8+ T cells from the spleens were purified using positive selection with CD8a microbeads (Miltenyi). $5x10^6$ cells were injected (i.v.) into a TCRα KO recipient. The next day, mice were infected with $1x10^4$ CFU of wild-type (WT) Listeria monocytogenes expressing the gB-8p peptide (Lm-gB) as previously described[2]. At 3 days post-infection (peak of bacterial growth), lysed tissue homogenate from the spleen and liver were serially plated and colony growth was enumerated. 3 days post-infection was chosen because any changes in the layering or behavior of CD8+ T cells would likely have a significant impact on bacterial growth at this time point.

Infections

Mice were directly infected with either $1x10^4$ CFU of wt LM-gB I.V. or $1.5x10^6$ PFU of Vaccinia-gB I.P., as previously described[2]. To assess antigen-specific CD8+ T cell responses, we looked at 5 days post-infection (d.p.i). Our group has previously reported that 5 d.p.i. is the peak of the fetal-derived CD8+ T cell response[1], making it a suitable timepoint to observe differences in the behavior and phenotype. Listeria was chosen because it is a well-characterized pathogen that requires a CD8+ T cell response for clearance. Vaccinia was chosen because it is a large DNA virus that requires a different route of inoculation.

Proliferation Index

To quantify the number of divisions that each experimental group underwent following *in vitro* stimulation with CD3/28, we utilized the proliferation index in FlowJo which allowed us to control for the undivided peak within each group (Tree Star, Ashland, OR).

Flow cytometry

For flow cytometry, cells were stained with antibodies against CD4 (Gk1.5), CD44 (IM7), CD122 (TM-B1), CD49d (R1-2), CD62L (ML-14), CD8a (53-6.7), TCR g/d (GL3). All antibodies were

purchased from eBioscience, BioLegend, Invitrogen, or BD Biosciences. Concentrations were used as recommended by the manufacturer. Flow cytofluorimetric data were acquired using FACSDiva software from a BD FACSymphony equipped with five lasers (BD Biosciences). Analysis was performed with FlowJo (Tree Star, Ashland, OR).

CD8+ T cell purification and fluorescent activated cell sorting

CD8+ T cells from the spleens were purified using positive selection with CD8a microbeads (Miltenyi). Cells were stained with antibodies against CD4 (Gk1.5), CD44 (IM7), CD122 (TM-B1), CD49d (R1-2), and CD8a (53-6.7). TN cells were sorted for $CD4^{-}CD8^{+}CD49d^{lo}CD44^{lo}$ VM cells were sorted for $CD4^{-}CD8^{+}CD49d^{lo}CD44^{hi}$. All samples were sorted to >95% purity as determined by a post-purity sort check.

RNA preparation and sequencing

Splenic CD8+ T cells were isolated and purified using positive selection with CD8a microbeads (Miltenyi). Purified CD8+ T cells were stained using antibodies against CD4 (Gk1.5), CD44 (IM7), CD122 (TM-B1), CD49d (R1-2), and CD8a (53-6.7). 'Bulk stamped' CD8+ populations were sorted for $CD4^{-}CD8^{+}CD49d^{lo}GFP^{+}$, 'stamped VMs' populations were sorted for $CD4^{-}CD8^{+}CD49d^{lo}GFP^{+}CD44^{hi}$, and 'stamped TN' populations were sorted for $CD4^{-}CD8^{+}CD49d^{lo}GFP^{+}CD44^{lo}$. Samples were sorted using a BD FACS Aria Fusion II equipped with four lasers. Samples were sorted to >95% cell purity as determined by a post-sort check. Following cell counting, 10,000 sorted cells were isolated then lysed in 1 mL Trizol according to the manufacturer's instructions (Thermo Fisher) with the addition of a second chloroform extraction and Glycoblue carrier (Thermo Fisher) before precipitation (1 hr at 4°C) and a second wash of the pelleted RNA in 70% ethanol. The remaining cells were used to generate ATAC seq libraries. RNA integrity was confirmed with a Femto Pulse Fragment Analyzer (Agilent). RNAseq libraries were prepared by the Transcriptional Regulation and Expression (TREx) Facility using the NEBNext Ultra II RNA Library Prep Kit (New England Biolabs). All RNA-seq libraries were sequenced with 2x150 paired-end reads on a HiSeqX (Illumina) at Novogene (Sacramento, CA).

ATAC-seq library preparation and sequencing

Splenic CD8+ T cells were isolated and purified using positive selection with CD8a microbeads (Miltenyi). Purified CD8+ T cells were stained using antibodies against CD4 (Gk1.5), CD44 (IM7), CD122 (TM-B1), CD49d (R1-2), and CD8a (53-6.7). 'Bulk stamped' CD8+ populations were sorted for $CD4^{-}CD8^{+}CD49d^{lo}GFP^{+}$, 'stamped $CD44^{hi}$' populations were sorted for $CD4^{-}CD8^{+}CD49d^{lo}GFP^{+}CD44^{hi}$, and 'stamped $CD44^{lo}$' populations were sorted for $CD4^{-}CD8^{+}CD49d^{lo}GFP^{+}CD44^{lo}$. Samples were sorted using a BD FACS Aria Fusion II equipped with four lasers. Samples were sorted using a BD Aria Fusion II equipped with four lasers. Samples were sorted to >95% cell purity as determined by a post-sort check. After isolating 10,000 cells for RNA-seq, the 25,000 cells were permeabilized as previously described[2]. The permeabilized cells were counted, snap-frozen in liquid nitrogen, and stored at -80°C. ATAC-seq libraries were prepared by the TREx Facility from 25,000 permeabilized using half-reaction volumes for the transposition reaction as previously described[3,4]. Briefly, frozen permeabilized cells were thawed, washed twice in 1 mL cold ATAC-RSB buffer containing 0.1% Tween20, resuspended in 25 µL Omni-ATAC Transposition Mix (10 mM Tris pH 8, 5 mM $MgCl_2$, 10% DMF, 0.1% Tween20, 0.01% digitonin, 1.25 µL TDE1 enzyme (Illumina), and 8.25 µL PBS in final 25 µL volume per sample), incubated for 30 minutes at 37°C with shaking (900 rpm), with immediate addition of Zymo binding buffer. After Zymo column cleanup, libraries were amplified using Nextera PCR oligos (Nextera Index Kit, Illumina) and 2x HiFi Master Mix (New England Biolabs) with an initial incubation at 72°C for 5 minutes to extend the segmented ends and 11 PCR amplification cycles. The libraries were cleaned up twice (in series) with a 2:1 ratio of SPRIselect beads (Beckman Coulter). ATACseq libraries were sequenced with 2x150 paired-end reads on a HiSeqX (Illumina) at Novogene (Sacramento, CA).

RNA-seq analysis

Raw reads were evaluated, trimmed, and filtered using trim galore (v0.6.5; -j 8 —quality 20 -gzip —length 50 –paired –fastqc) and aligned to the mouse reference genome (mm10) using STAR (v2.7.0e) with default parameters[5]. Quality control metrics were calculated using multiqc (v1.8)[6]. Mapped reads were assigned to genes using featureCounts under Rsubread (v2.4.3; minMQS=30,

countMultiMappingReads=FALSE, isPairedEnd=TRUE)[7]. DESeq2 (v1.30.1) was then used to perform differential expression analysis and statistical testing of expressed genes (more than 5 overlapping reads in at least 2 samples) [8]. For gene set enrichment analysis, genes were first ranked by $\log_2$(clean/dirty). Gene sets comparing naïve, effector, and memory $CD8^+$ T cells [9] were downloaded from the molecular signature database (MSigDB), then fgsea (v1.16.0) was used with default parameters except eps=0 to calculate normalized enrichment scores and p-values, followed by Benjamini-Hochberg correction.

ATAC-seq analysis

Raw reads were evaluated, trimmed, and filtered using trim galore (v0.6.5; -j 8 —quality 20 -gzip —length 50 –paired –fastqc) and aligned to the mouse reference genome (mm10) using bwa mem (-M -R)[10], then processed using samtools view (v1.9; -b -h -F 0x0100). The reads were further sorted and indexed using samtools sort and samtools index[11]. Mitochondrial reads were filtered out using a custom shell script, and then picard MarkDuplicates was used to deduplicate the reads. Peaks were called using MACS3 (v3.0.0a6; callpeak -f BAMPE -g mm -B -q 0.01)[12], then filtered through Irreproducibility Discovery Rate (IDR) analysis, as previously described by the ENCODE Consortium[13]. A unified peak list was generated by merging reproducible peaks from each cell type using bedtools (v2.29.2)[14], resulting in a total of 30,830 peaks. Peaks that matched blacklist of artifactual regions in mm10 were filtered out. The peaks were associated with their nearest genes based on the shortest distance between the peak and the gene's promoter, as annotated by GENCODE (release M27, mapped back from mm39 to mm10) and defined as 1kb upstream and 500 bp downstream of the annotation transcription start site. featureCounts under Rsubread (v2.4.3) was used to calculate raw read counts per peak, which were then aggregated per gene for differential analysis. Raw reads were normalized to counts per million (CPM), then the EdgeR (v3.32.1) framework was used to determine differentially accessible genes across all genes accessible (cpm $\geq$ 0.5) in at least one sample[15]. Gene set enrichment analysis was performed as described above.

Poised Gene Analysis

Poised genes were defined as genes that have at least one associated peak with significantly (q-value < 0.05) higher accessibility in one sample compared to another per ATAC-seq signal, but do not have significant upregulation in transcripts per RNA- seq. Gene ontology on these genes were performed on ShinyGO (v0.66) with all annotated genes as background[16]. After Benjamini-Hochberg correction, the significant terms (q-value < 0.05) were visualized using Cytoscape.

Track Visualization and Normalization

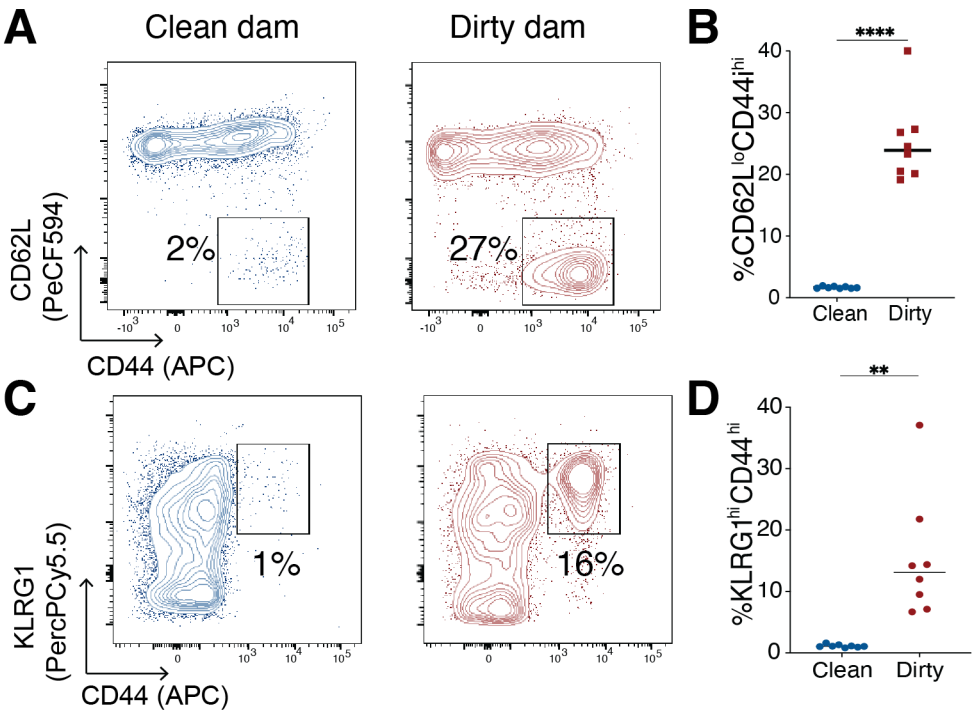
Replicates were combined using samtools merge (v1.9), indexed, then normalized to CPM using bamCoverage from deep tools (v3.5) to generate bigwig files used for visualization. Tracks were visualized with igv (v2.9.1).

Infectious agent screening

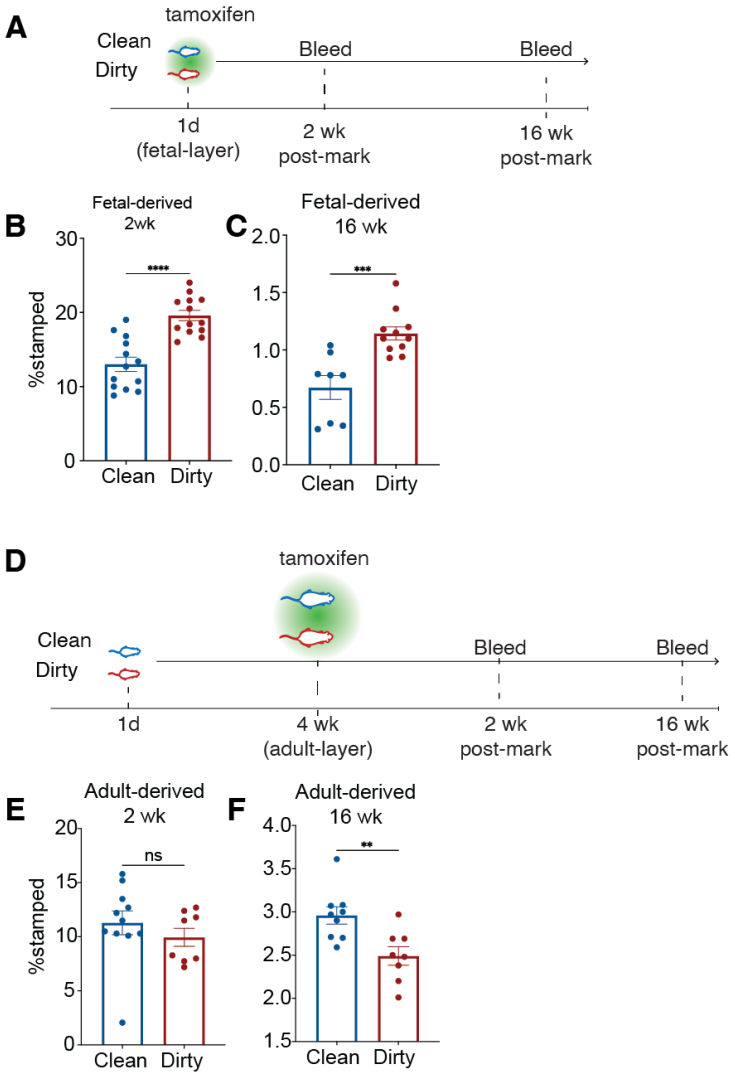
'Clean', 'dirty' mice (before mating), and pet shop mice were screened for microbial load. Blood, feces, oral, and body swabs were collected per the instructions of Charles River Laboratories.

Data availability

The accession number for the RNA-seq and ATAC-seq data reported in this paper is GEO SuperSeries: GSE213831.



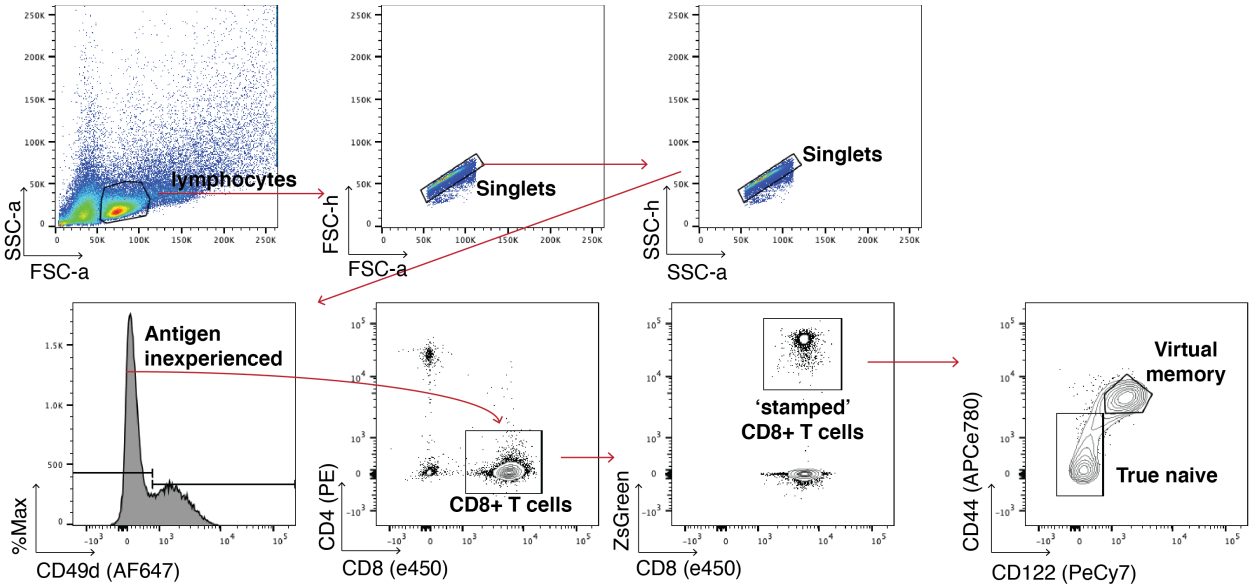
Supplemental Fig 1. Evidence that the CD8+ T cell compartment in dams is altered by exposure to pet-shop material. (A) Representative flow plots showing the $CD44^{hi}CD62L^{lo}$ population in clean and dirty dams prior to mating. **(B)** Proportion of $CD44^{hi}CD62L^{lo}$ CD8+ T cells. **(C)** Representative flow plots showing the $CD44^{hi}KLRG1^{hi}$ population in clean and dirty dams prior to mating **(D)** Proportion of $CD44^{hi}KLRG1^{hi}$ CD8+ T cells forward scatter **(D)** KLRG1 and **(E)** CD62L. Data is pooled from 2 independent experiments (n=4/group). Data for B-D are ±SEM. Statistical significance was determined by student's t-test (** P<0.005, **** P<0.00005)



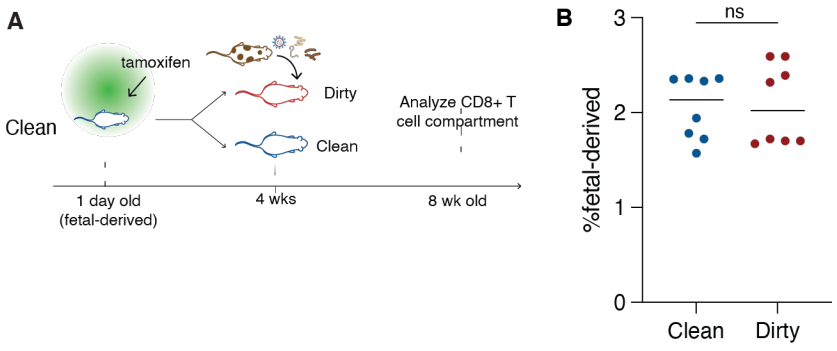
205

Supplemental Fig 2. Early microbial exposure permanently alters the developmental layering of the CD8+ T cell compartment. (A) Approach to mark and track fetal-derived cells. Proportion of stamped fetal-derived cells at **(B)** 2 wk and **(C)** 16 wk post-tamoxifen administration. **(D)** Approach to mark and track adult-derived cells. **(E)** Proportion of stamped fetal-derived cells at

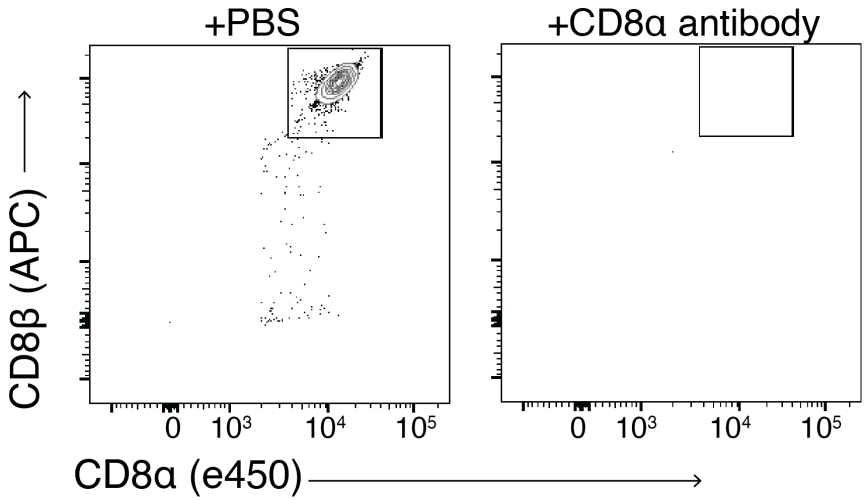
210 2 wk and **(F)** 16 wk post-tamoxifen administration. Data are pooled from two independent experiments (n=4-5 mice/group). Statistical significance was determined by a student's *t*-test (ns not significant, ** p<0.005, ***p<0.0005, **** p<0.00005)



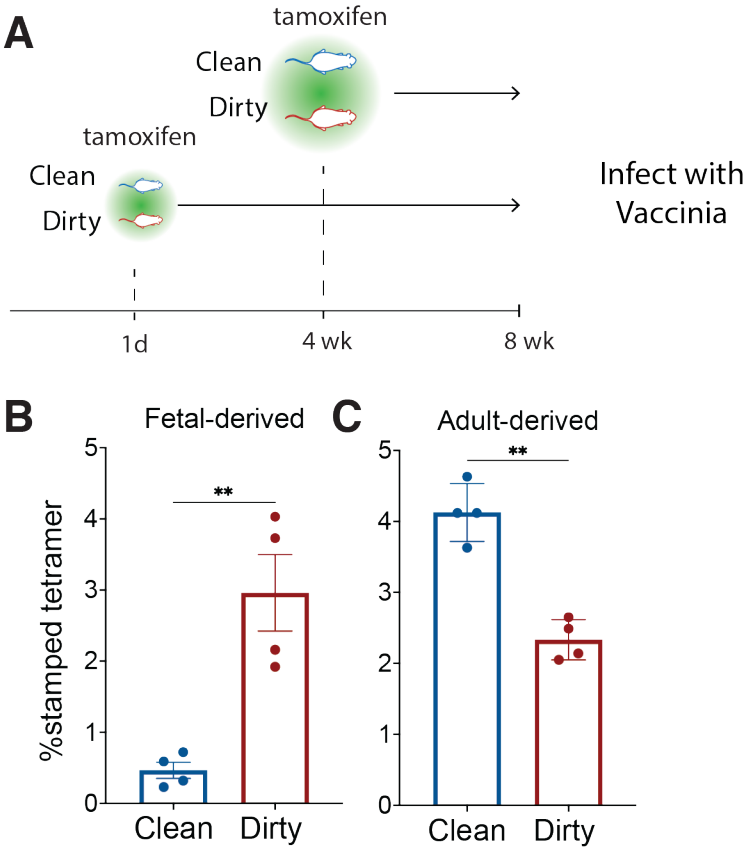
Supplemental Fig 3. Gating strategy for identifying antigen-inexperienced 'stamped' fetal-derived CD8+ T cells for flow analysis. Representative flow plots to identify naïve fetal-derived CD8+ T cells with a virtual memory or true naïve phenotype.



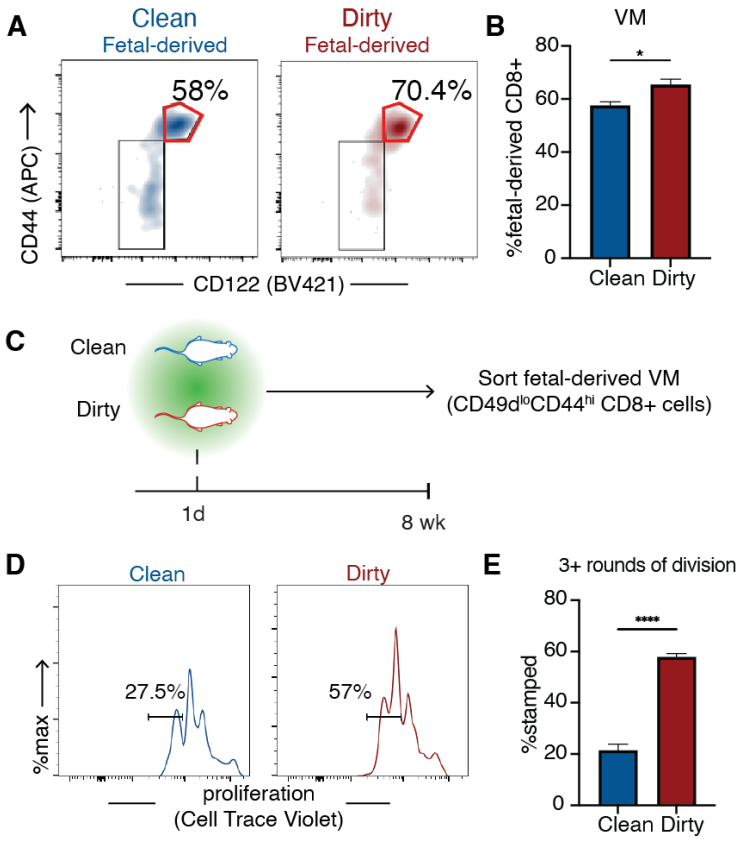
Supplemental Fig 4. Expansion of fetal-layer is dependent on the timing of microbial exposure. **(A)** Experimental schematic showing timestamped mice exposed to a dirty environment during adulthood. **(B)** Proportion of fetal-derived CD8+ T cells present in adulthood. Data is pooled from two independent experiments (n=4 mice/group). Statistical significance was determined by a student's *t-test* (ns not significant).



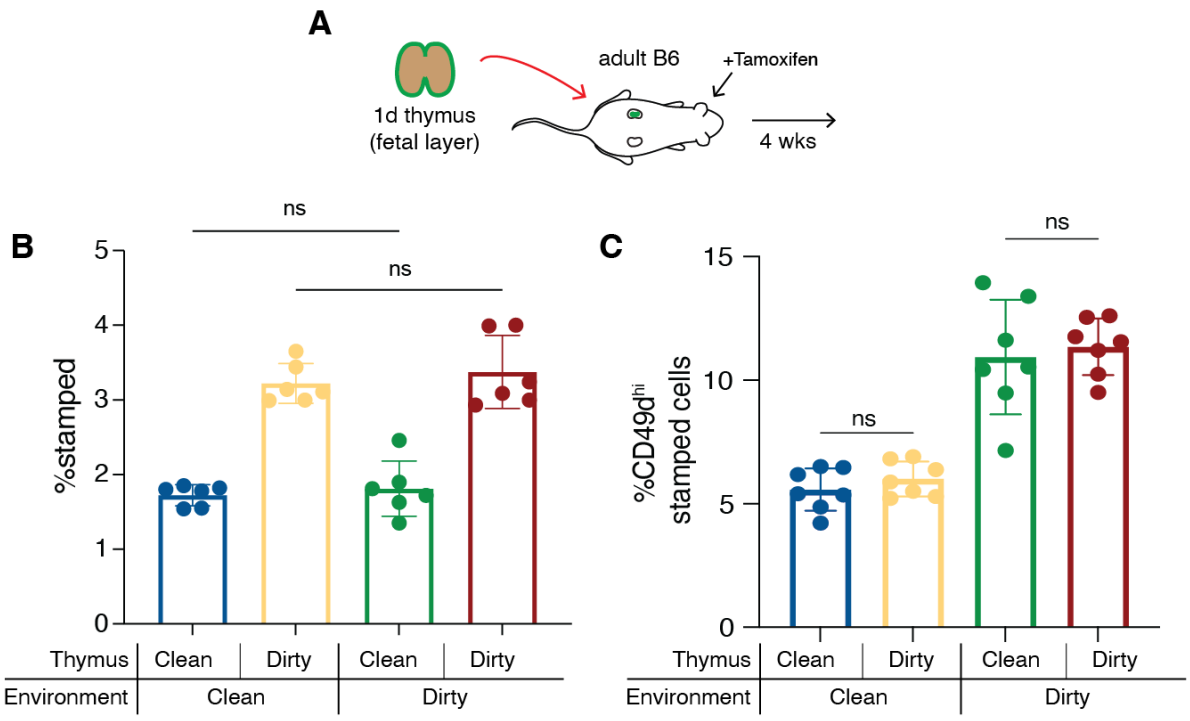
Supplemental Fig 5. CD8α antibody administration removes all CD8+ T cells. Representative contour plots showing the CD8+ T cell population following injection with CD8α antibody or PBS vehicle control. The CD8αβ+ population was gated off of CD3+TCRβ+.



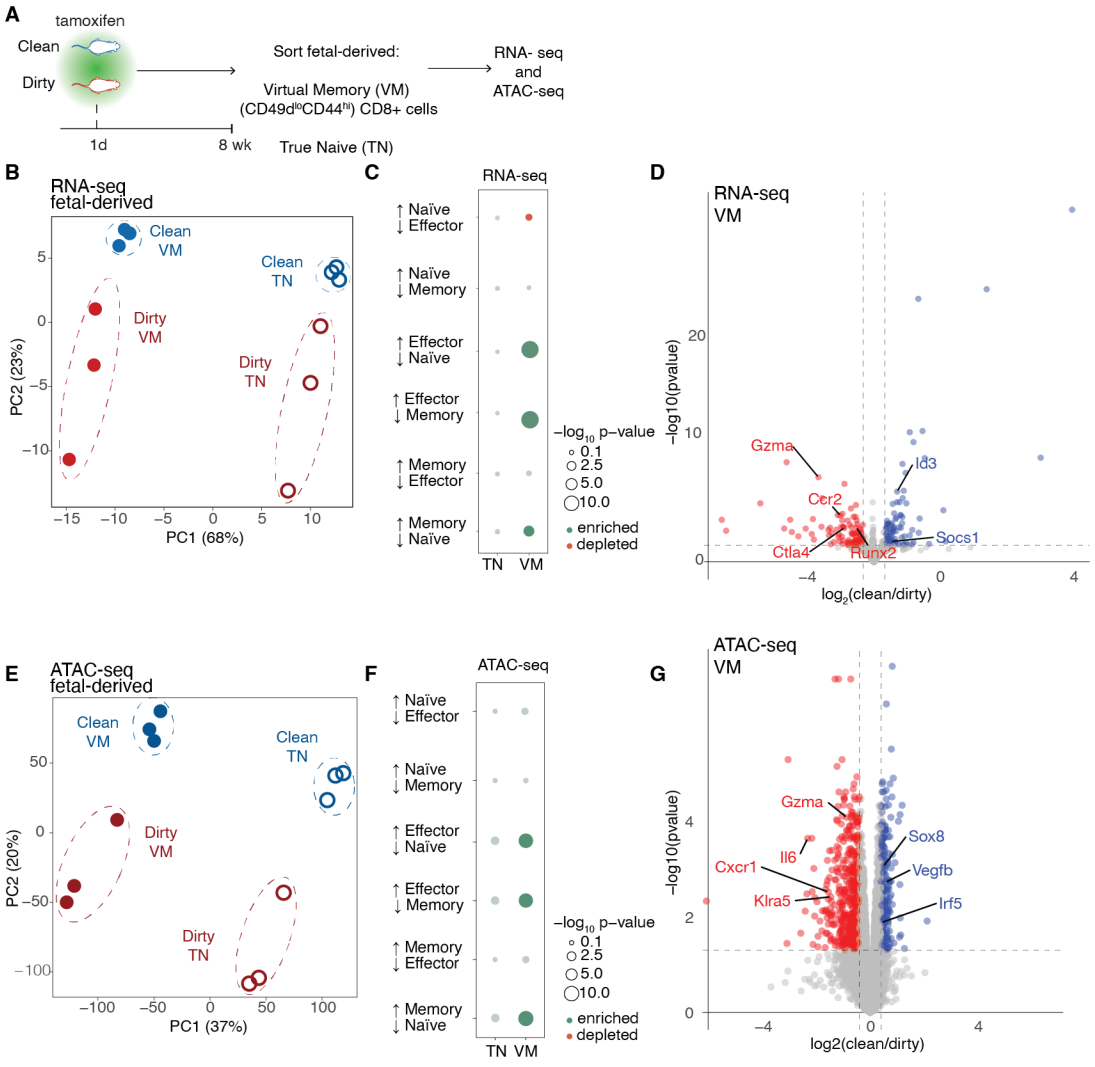
Supplemental Fig 6. Alterations in antigen-specific responses of different layers are not pathogen dependent. (A) Approach to infect timestamp mice with Vaccinia virus. **(B)** The proportion of antigen-specific fetal- and **(C)** adult-derived cells at 5 dpi. Data for B, C are ± SEM. Statistical significance was determined by a student's *t-test* (** $p<0.005$).



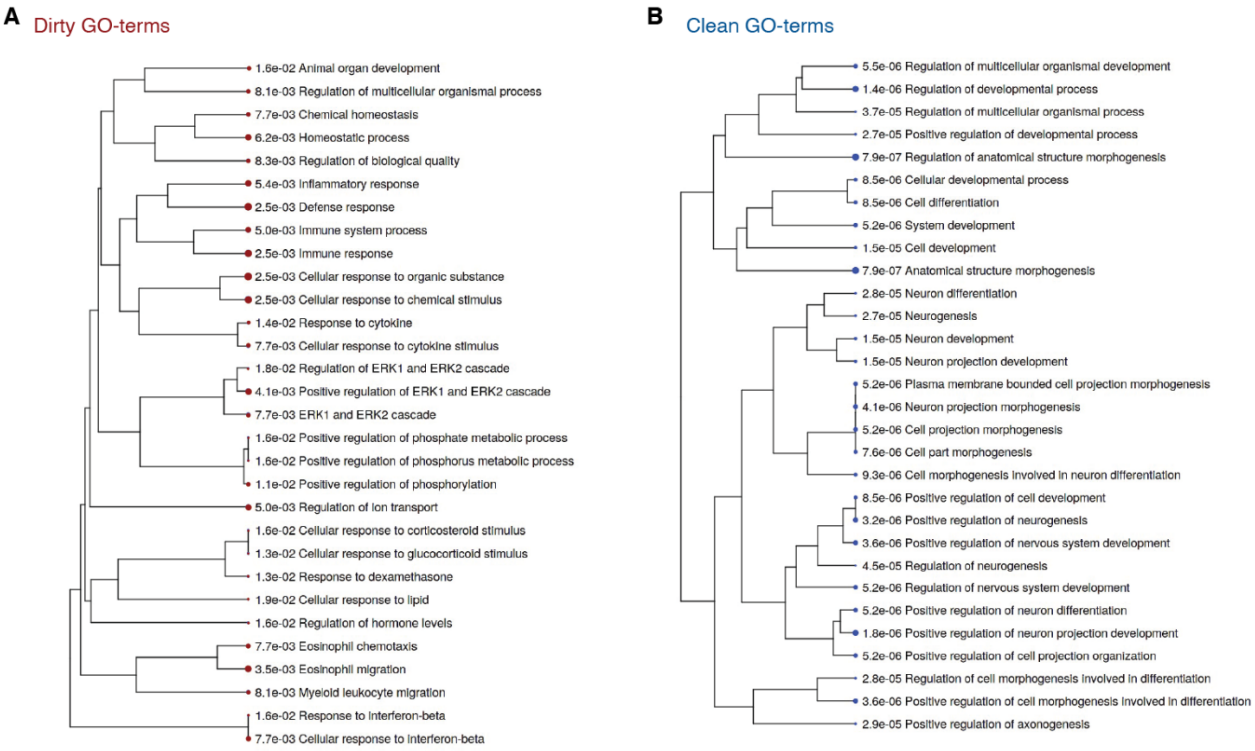
Supplemental Fig 7. Dirty fetal-derived cells are inherently more reactive (A) Representative flow plots of VM population in fetal-derived clean or dirty CD8+ T cells **(B)** Proportion of clean or dirty fetal-derived cells with a virtual memory phenotype **(C)** Experimental approach to sort out antigen-inexperienced VM population **(D)** Representative graphs showing the proportion of cells in the last two rounds of division **(E)** Proportion of cells that underwent 3+ rounds of division. Data for B, E are mean ± SEM and are pooled from 2 independent experiments (n=4-6 mice/group). Statistical significance was determined by a student's t-test (* P<0.05, ** P<0.005)



Supplemental Fig 8. Fetal-derived CD8s from dirty thymus preferentially expand in the periphery. (A) Experimental approach to transplant a fetal thymus under the kidney capsule of an adult recipient. **(B)** Proportion of fetal-derived CD8+ T cells 4 weeks post-transplantation. **(C)** Proportion of antigen-experienced fetal-derived cells 4 weeks post-transplantation. Data is representative of two independent experiments. (n=6 mice/group. are mean ± SEM and are pooled from 2 independent experiments. Statistical significance was determined by a one-way ANOVA with a Tukey multiple comparison test (ns not significant).



Supplemental Fig 9. Cell-intrinsic factors drive enhanced reactivity in fetal-layer. (A) Schematic of sorting strategy for clean or dirty fetal-derived VM or TN cells. **(B)** Principal component analysis, **(C)** Enrichment analysis, and **(D)** Volcano plots showing differentially expressed genes using RNA-seq in clean or dirty fetal-derived with a VM phenotype. **(E)** Principal component analysis, **(F)** Enrichment analysis, and **(G)** Volcano plots showing differentially accessible genes using ATAC-seq from clean or dirty fetal-derived CD8+ T cells with a VM phenotype.



280

Supplemental Fig 10. Dirty fetal-derived cells are poised towards effector-like behavior (A) Dendrograms showing enriched GO-terms for poised genes in dirty and **(B)** clean fetal-derived cells.

		Clean	Dirty	Pet shop
Bacteria	C. Jejuni	-	-	+
	Campylobacter	-	+	+
	Cryptosporidium	-	+	+
	H. bilis	-		+
	H. ganmani	-	+	+
	H. mastomyrinus	-	+	+
	H. typhlonius	-	+	+
	Helicobacter genus	-	+	+
	M. pulmonis	-	+	+
	R. heylii	-	+	+
	R. pneumotropicus	-	+	+
	S. aureus	-	+	+
	Ectromelia	-	-	-
	Campylobacter Genus	-	+	+
	C. bovis	-	-	-
	C. kutscheri	-	-	-
	C. rodentium	-	-	-
	C. piliforme	-	-	-
	K. oxytoca	-	-	-
	K. pneumoniae	-	-	-
	Helicobacter	-	+	+
	M. pulmonis	-	-	+
	R. heylii	-	+	+
	R. pneumotropicus	-	+	+
	Ps. aeruginosa	-	-	-
	Salmonella Genus	-	-	-
	S. aureus	+	+	-
	S. moniliformis	-	-	-
	S. pneumoniae	-	-	-
	Cryptosporidium	-	+	+
	Demodex	-	-	-
	Entamoeba	-	+	+
	P. mirabilis	+	+	+
Virus	MAV 1	-	+	+
	MAV 2	-	+	+
	MHV	-	+	+
	Mouse Norovirus	-	+	+
	Mouse Parvovirus	-	+	+
	Theiler's murine encephalomyelitis virus	-	+	+
	LCMV	-	-	
Pinworms	Aspiculuris tetraptera	-	+	+
	Syphacia obvelata	-	-	+
Parasites	Entamoeba	-	-	-
	Giardia	-	-	-

285

Supplemental Table 1. Microbial contents of clean and dirty dams, and pet shop mice. Bacterial, viral, pinworm, and parasitic load of clean and dirty and pet-shop mice.

References

1 Smith, N. L. *et al.* Developmental Origin Governs CD8(+) T Cell Fate Decisions during Infection. *Cell* **174**, 117-130 e114, doi:10.1016/j.cell.2018.05.029 (2018).

2 Smith, N. L. *et al.* Rapid proliferation and differentiation impairs the development of memory CD8+ T cells in early life. *J Immunol* **193**, 177-184, doi:10.4049/jimmunol.1400553 (2014).

3 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).

4 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).

5 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

6 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).

7 Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**, e47, doi:10.1093/nar/gkz114 (2019).

8 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

9 Luckey, C. J. *et al.* Memory T and memory B cells share a transcriptional program of self-renewal with long-term hematopoietic stem cells. *Proc Natl Acad Sci U S A* **103**, 3304-3309, doi:10.1073/pnas.0511137103 (2006).

10 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

11 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

12 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

13 Li, Q., Brown, J.B., Huang, H., Bickel, P.J. . Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752-1779 (2011).

14 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

15 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

16 Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628-2629, doi:10.1093/bioinformatics/btz931 (2020).