

A

C	A	S	S	L	A	G	G	N	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCT	GGG	GGG	AAC	CAA	GAC	ACC	CAG	TAC	TTT
---	---	---	AGC	TTA	GCA	GGG	GGG	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCG	GGG	GGC	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCG	GGG	GGG	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCT	GGG	GGG	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCT	GGG	GGT	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCC	GGG	GGT	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCC	GGG	GGG	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCG	GGG	GGT	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCC	GGG	GGT	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTG	GCT	GGG	GGG	AAC	CAA	---	---	---	---	---
---	---	---	AGC	TTG	GCT	GGG	GGT	AAC	CAA	---	---	---	---	---

Multiple nucleotide sequences encoding same amino acid sequence

C	A	S	S	L	A	G	G	D	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCT	GGG	GGG	GAC	CAA	GAC	ACC	CAG	TAC	TTT
---	---	---	AGC	TTA	GCA	GGG	GGC	GAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCA	GGG	GGG	GAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCA	GGG	GGG	GAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCG	GGG	GGC	GAC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCG	GGG	GGG	GAC	CAA	---	---	---	---	---
---	---	---	AGC	TTG	GCT	GGG	GGG	GAC	CAA	---	---	---	---	---

C	A	S	S	L	A	G	G	A	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCT	GGG	GGG	GCC	CAA	GAC	ACC	CAG	TAC	TTT
---	---	---	AGC	TTA	GCT	GGG	GGG	GCC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCA	GGG	GGG	GCC	CAA	---	---	---	---	---
---	---	---	AGC	TTG	GCT	GGG	GGG	GCC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCC	GGG	GGG	GCC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCG	GGG	GGC	GCC	CAA	---	---	---	---	---
---	---	---	AGC	TTA	GCG	GGG	GGG	GCC	CAA	---	---	---	---	---

C	A	S	S	L	A	G	G	G	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCT	GGG	GGG	GGC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	A	G	G	H	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCG	GGG	GGG	CAC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	A	G	G	P	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCG	GGG	GGG	CCC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	A	G	G	S	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCT	GGG	GGG	TCC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	A	G	G	T	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GCT	GGG	GGG	ACC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	T	G	G	A	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	ACT	GGG	GGG	GCC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	T	G	G	G	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTG	ACT	GGG	GGG	GGC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	T	G	G	I	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTG	ACT	GGG	GGG	ATC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	T	G	G	S	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	ACT	GGG	GGG	AGC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	G	G	G	A	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GGT	GGG	GGG	GCC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	G	G	G	D	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GGG	GGG	GGC	GAC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	G	G	G	H	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GGG	GGG	GGG	CAC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	G	G	G	N	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	GGG	GGG	GGG	AAC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	S	G	G	N	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	TCT	GGG	GGG	AAC	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	S	G	G	A	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	TCT	GGG	GGG	GCG	CAA	GAC	ACC	CAG	TAC	TTT

C	A	S	S	L	S	G	G	H	Q	D	T	Q	Y	F
TGT	GCA	AGC	AGC	TTA	TCT	GGG	GGG	CAC	CAA	GAC	ACC	CAG	TAC	TTT

B

S	L	A	G	G	N	Q	No. of n.t. additions
AGC	TTA	GCT	GGG	GGG	AAC	CAA	0
---	TTA	GCT	GGG	GGG	AAC	---	1
---	TTA	GCT	GGG	GGG	AAC	---	1
---	TTA	GCT	GGG	GGG	AAC	---	1
---	TTA	GCT	GGG	GGG	AAC	---	1
---	TTA	GCT	GGG	GGG	AAC	---	1
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2
---	TTA	GCT	GGG	GGG	AAC	---	2

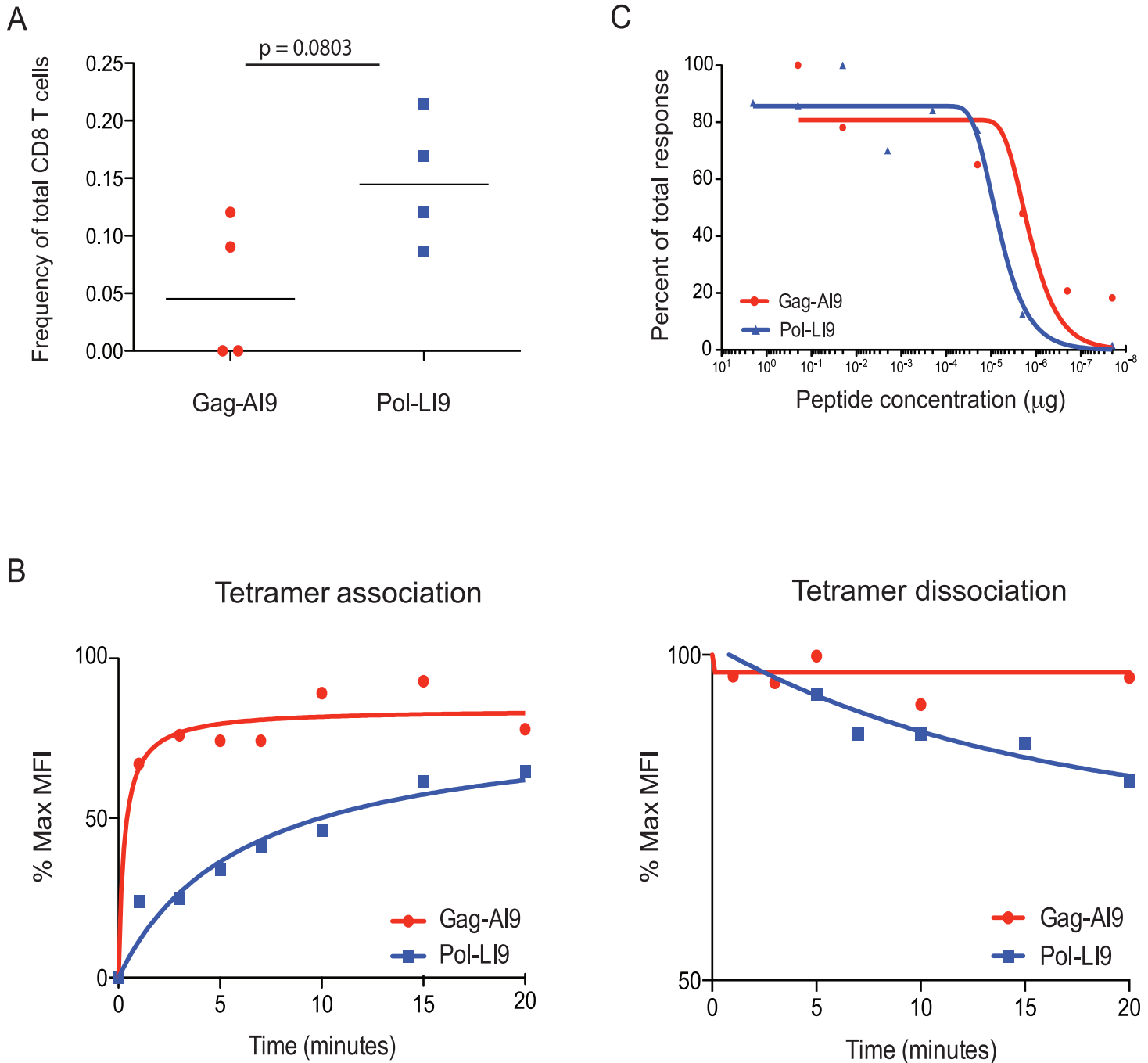
Multiple recombination events producing same nucleotide sequence

C

Germline genes														
TRBV16:	TGT	GCA	AGC	AGC	TTA	GA								
TRBD1:	GGG	CAG	GGGGGC											
TRBD2:	GGG	ACT	GGGGGGGC											
TRBJ2-5:	AAC	CAA	GAC	ACC	CAG	TAC	TTT							

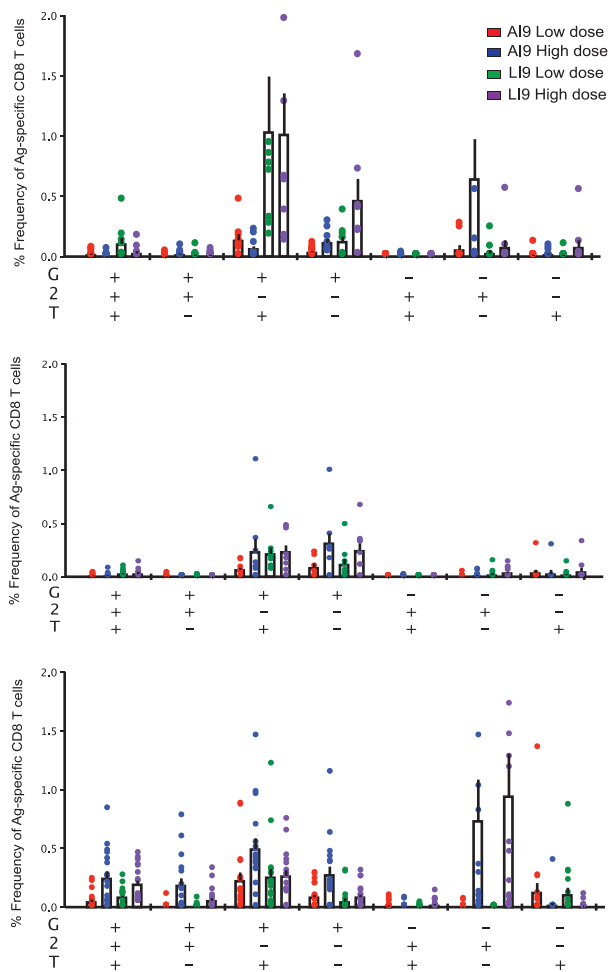
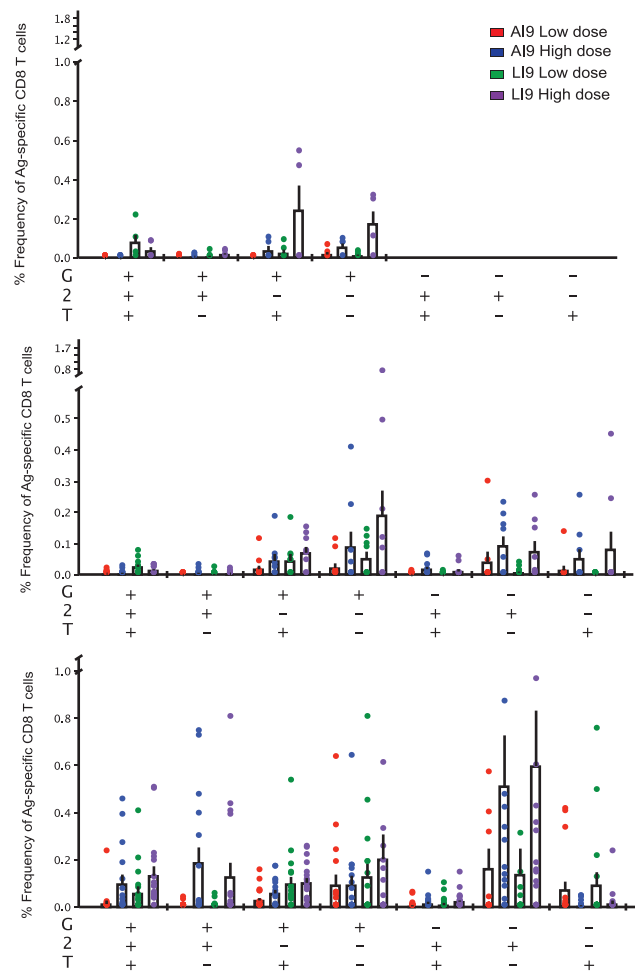
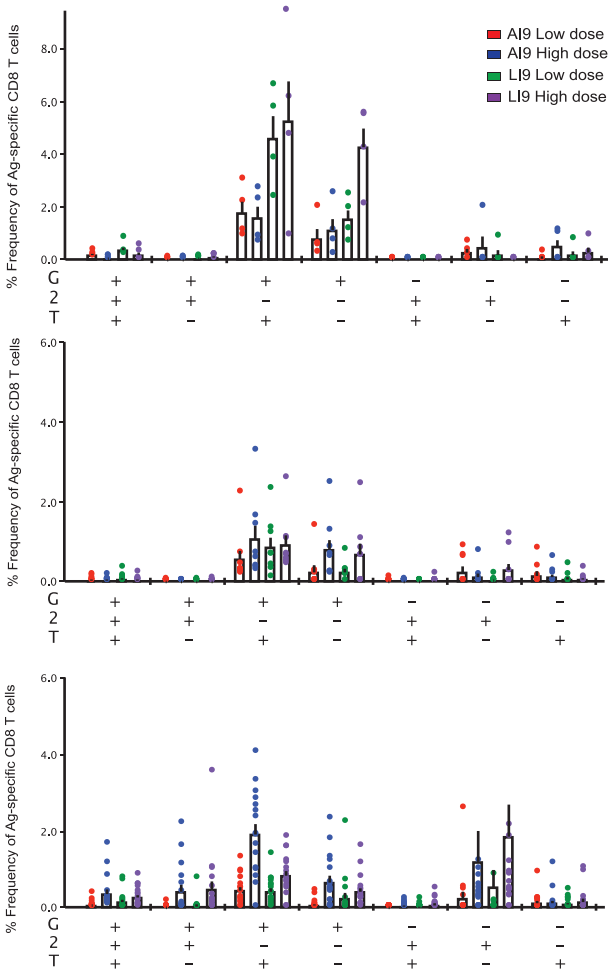
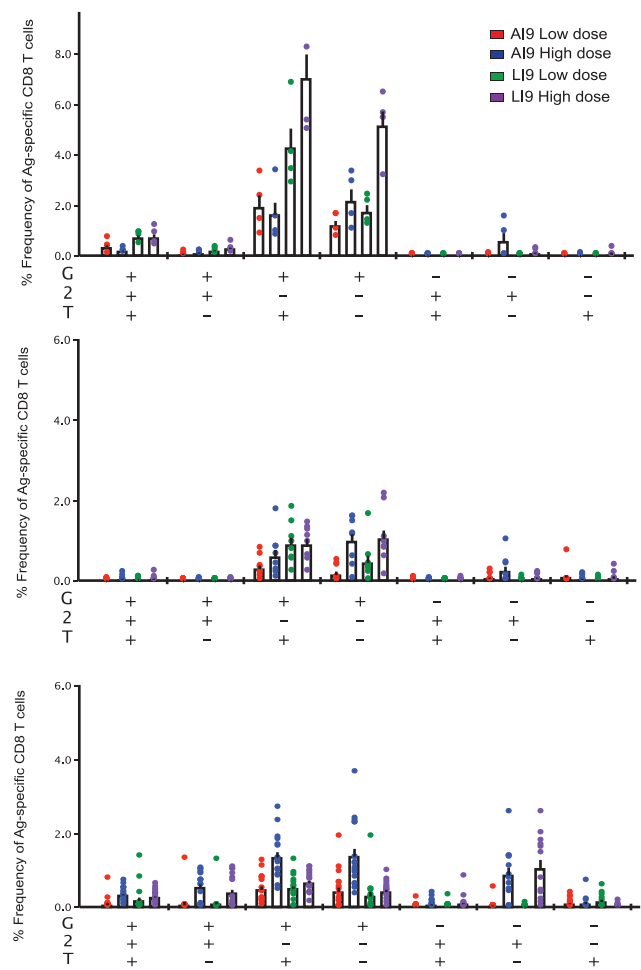
Supplemental Figure 1. Convergent recombination in the Pol-LI9-specific CD8 TCRβ repertoire.

(A) The Pol-LI9-specific CD8 TCRβ repertoire using the *TRBV16* and *TRBJ2-5* genes comprises the majority of sequences that conform to the CASSLXGGXQDTQYF amino acid motif. A subset of each of the consensus sequences is shown. One possible alignment of the nucleotide sequences with the *TRBV16* (blue), *TRBD1* or *TRBD2* (red), and *TRBJ2-5* (green) genes involving a minimal number of nucleotide additions (black) is shown for each nucleotide sequence. Some amino acid sequences were found to be encoded by a variety of nucleotide sequences. This is demonstrated for the amino acid sequences CASSLAGGNQDTQYF, CASSLAGGDQDTQYF and CASSLAGGAQDTQYF. (B) An example of the multiple recombination events at the V(D)J junction, involving no more than two nucleotide additions, that can produce one of the nucleotide sequences (yellow box) encoding CASSLAGGNQDTQYF is shown. Some of these gene recombination events involve few nucleotide additions and are thus expected to recur frequently. (C) The germline genes are shown for reference.



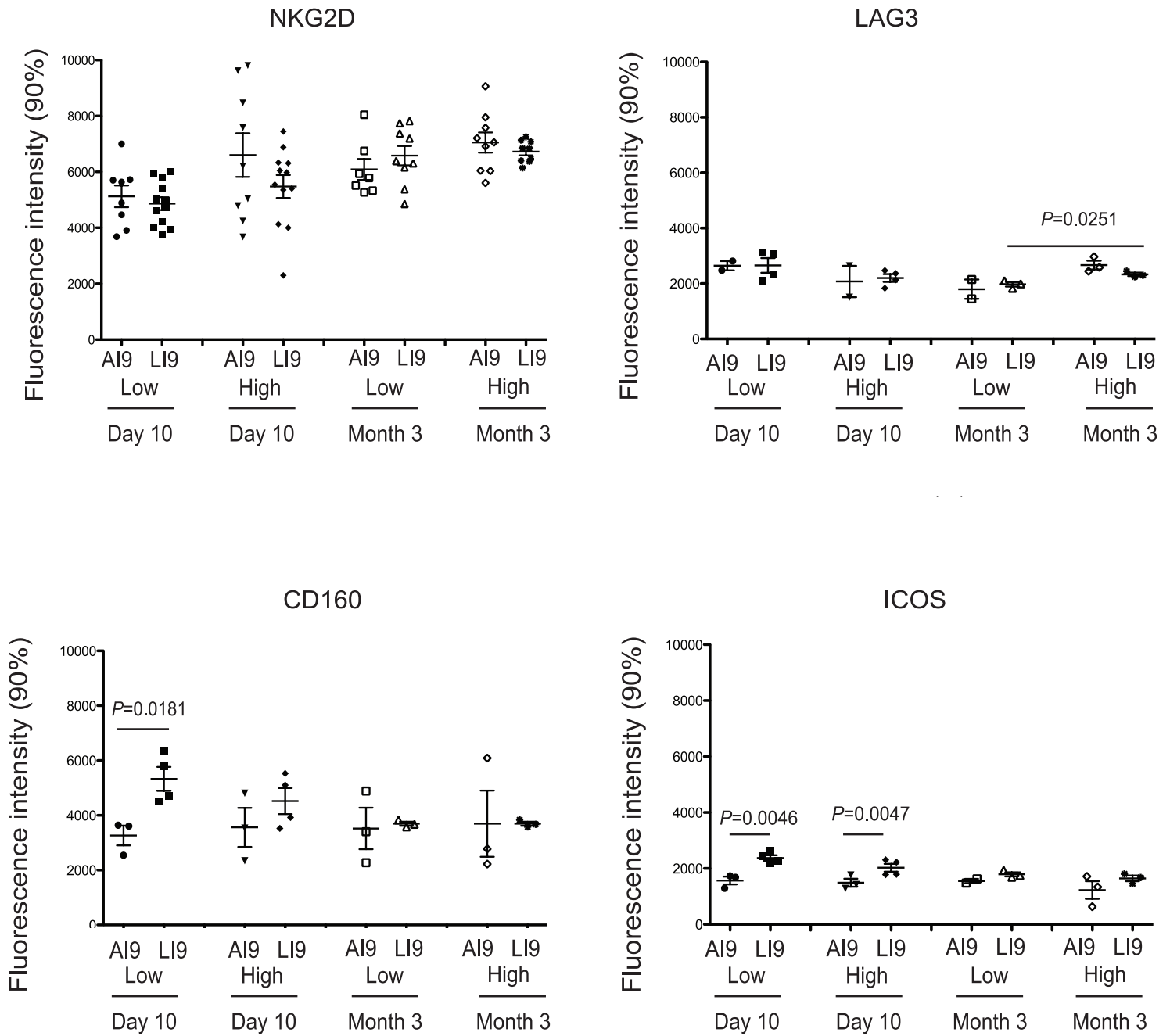
Supplemental Figure 2. Analysis of precursor frequency, antigen avidity and functional sensitivity.

(A) Precursor frequency measurements are shown as proportions of the total CD8 T cell pool. Four naive animals were assayed for the presence of Gag-AI9-specific and Pol-LI9-specific CD8 T cells by tetramer staining. Statistical analysis was performed using the Wilcoxon matched-pairs signed rank test. (B) Representative tetramer association and dissociation analyses are shown for Gag-AI9-specific and Pol-LI9-specific CD8 T cell populations. Non-linear least squares fitting of the described model yielded the following association (left panel) values: for Gag-AI9, $Y = 11.1\%$, $I_0 = 75.7$, $I_\infty = 0.104$, $m = 13.0$; for Pol-LI9, $Y = 4.49\%$, $I_0 = 99.1$, $I_\infty = 0.0223$, $m = 14.9$. All rates are expressed in min^{-1} . The important parameter is Y , which is proportional to the TCR/pMHC1 on-rate. Parameter estimates for tetramer dissociation (right panel) were as follows: for Gag-AI9, $\mu = 0$, $\nu = 0.17$, $y_0 = 93\%$; for Pol-LI9, $\mu = 0.17$, $\nu = 0.29$, $y_0 = 71\%$. All rates are expressed in min^{-1} . (C) Functional sensitivity analysis was performed using intracellular cytokine flow cytometry across a range of peptide concentrations. Peptides were assayed at ten-fold dilutions in splenocyte cell suspensions from a vaccinated animal. Values as percentages of the total response were calculated by nonlinear regression analysis. Red circles, Gag-AI9; blue symbols, Pol-LI9.

A**B****C****D**

Supplemental Figure 3. Functional analysis of antigen-specific CD8 T cells after vaccination.

(A-D) SPICE analysis of functional CD8 T cell responses specific for Gag-AI9 and Pol-LI9 in spleen (A), lung (B), lymph node (C) and blood (D) at day 10 (upper panel), month 1 (middle panel) and month 3 (lower panel) after vaccination with either low dose or high dose Ad5.Gag.Pol as indicated. Intracellular cytokine staining was performed for IFN γ (G), TNF (T) and IL-2 (2) after peptide stimulation ex vivo. Boolean gate arrays were created using the FlowJo platform to determine the frequency of each of the eight possible response patterns per cell. Non-specific background events were subtracted from responses measured in the stimulated samples for each response pattern individually. Frequencies for each functional permutation are shown in bar charts with error bars representing the standard error of the mean. Responses from eight mice for each tissue at each time point are displayed.



Supplemental Figure 4. Activation and exhaustion status of antigen-specific CD8 T cells after vaccination. Polychromatic flow cytometry was used to determine the surface expression of activation (ICOS) and exhaustion (CD160, LAG3 and NKG2D) markers at day 10 and month 3 post-vaccination. The Gag-AI9-specific and Pol-LI9-specific CD8 T cell populations were compared across vaccine doses and time points as indicated. Statistical analyses were performed using the Mann-Whitney and Wilcoxon matched-pairs signed rank tests.