SUPPLEMENTAL DATA for

Rapid Immune Reconstitution of SCID-X1 Canines after In Vivo Gene-Therapy

Short title**: In vivo Gene Therapy for SCID-X1**

Olivier Humbert[1†], Frieda Chan[1†], Yogendra S. Rajawat[1], Troy R. Torgerson[3,4], Christopher R. Burtner[1], Nicholas W. Hubbard[3,4], Daniel Humphrys[1], Zachary K. Norgaard[1], Patricia O'Donnell[2], Jennifer E. Adair[1,4], Grant D. Trobridge[5], Andrew M. Scharenberg[4], Peter J. Felsburg[2], David J. Rawlings[3,4] and Hans-Peter Kiem[1,4 *]

List of Supplemental Items:

1. **Supplemental Materials and Methods**
2. **Supplemental Table S1, S2.**
3. **Supplemental Figures S1-S11.**

SUPPLEMENTAL MATERIALS AND METHODS

Animal breeding

The SCID-X1 dogs used in this study contain a 4 base-pair null mutation on exon 1 of the *IL2RG* gene that results in premature termination of the protein. Progenies were derived from the breeding of SCID-X1 affected male with hemizygous affected females. SCID-X1 pups were identified phenotypically by flow cytometry analysis of CD3+ cells in peripheral blood and by sequencing analysis of genomic DNA isolated from buccal swabs or peripheral blood samples. All SCID-X1 affected dogs are housed in HEPA-filtered facility at the Fred Hutchinson Cancer Research Center. R2258 and R2260 were born to SCID-X1 dogs cured with bone marrow transplant or in vivo FV gene therapy. H864 and H867 were sired by mating in vivo FV gene therapy cured SCID-X1 affected R2260 male with heterozygous SCID-X1 carrier female (**Supplemental Table S2**).

Ex vivo transduction of canine and human CD34+ cells

Human CD34+ cells were collected from volunteers under institutional review board-approved protocol. Human and canine CD34+ cells were isolated, cultured, and transduced with FV vectors as described previously.[21] Briefly, CD34+ cells were cultured overnight in IMDM containing 10% FBS and 100 ng/μL of the respective cytokines (FLT3, SCF, TPO for human; FLT3, TPO, cSCF, cG-CSF for canine), and transduced on CH296 fibronectin (Takara, New York, NY) at 2 μg/mL using MOIs determined by HT1080 titers. Transduction efficiency was evaluated based on fluorophore expression measured by flow cytometry analysis.

Foamy virus vector construction

The FV vector constructs used in this study have been previously described.[30] Briefly, the transfer vector includes a U3-deleted long terminal repeat and a 2.3 Kb FV cis acting region

containing the 3' region of pol and 5' region of env required for efficient gene transfer, with introduced stop codons in the foamy partial gag sequence. The vector was made self-inactivating by deletion of the Tas (Bel-1) transactivator, which is required for transcription from the LTR. The transgene is expressed from the intron-less human elongation factor 1α (EF1α) or from the human phosphoglycerate kinase (PGK) promoters. EGFP and a codon- optimized human common gamma chain receptor (γC) are separated by the Thosea asigna T2A peptide (2A). A safety-modified woodchuck post-transcriptional regulatory element (WPRE) contains the X protein promoter, with four mutated ATG sequences as previously described.[62] To generate EF1α.mCherry.2A.γC.FV and PGK.mCherry.2A.γC.FV, the EF1α.GFP.2A.γC and PGK.GFP.2A.γC, DNA inserts were first extracted by digestion with enzymes BglII/NotI, and ligation into BamHI/NotI of plasmid bluescript SK+ to generate pSK.EF1α.GFP.2A.γC and pSK.PGK.GFP.2A.γC. To construct pSK.PGK.mCherry.2A.γC, the mCherry sequence was PCR amplified using High Fidelity Platinum Taq (Life Technologies) using forward primers 5'-GATCCACCGGTCGCCACCATG-3' and reversed primer 5'-GTCGACGCGGCCGCTTTACTTG-3', digested with AgeI/BsrGI, and ligated into pSK.PGK.GFP.2A.γC cut with the same enzymes. pSK.EF1α.mCherry.2A.γC was constructed by amplification of the EF1α.mCherry sequence from a reference plasmid with forward primer 5'-ACTGCATGCCGATGGCTCCGGTGCCCGTC-3' and reversed primer 5'-GTCGACGCGGCCGCTTTACTTG-3', digestion with SphI/BsrGI and ligation into pSK.EF1α.GFP.2A.γC cut with the same enzymes. PGK.mCherry.2A.γC.FV was constructed by ligation of the AgeI/BamHI fragment from pSK.PGK.mCherry.2A.γC into PGK.GFP.2A.γC.FV. EF1α.mCherry.2A.γC.FV was constructed by ligation of the SphI/NotI fragment from pSK.EF1α.mCherry.2A.γC.FV into PGK.GFP.2A.γC.FV. The PGK- and EF1α-FV constructs overall produced comparable vector titers.

Data availability

All sequence data obtained in this study is available for download at www.ncbi.nlm.nih.gov/bioproject/ (BioProject ID: Pending).

Code availability

All custom R and Python codes are available on request. All other codes are publicly available and cited in the appropriate methods description.

Transduction filter methods. Collisions, IS appearing in distinct samples originating from unique transduction events, were present in the data. Theoretically, this should never be observed and is likely the result of contamination, barcode swapping or other errors in processing. In some cases, it is possible to determine which sample the IS originates from by comparing the number of genomically aligned sequence reads representing the IS in each sample. When examining collisions, the genomically aligned sequence counts were used instead of normalized frequencies to avoid biases introduced by low capture frequency in samples with few genomically aligned reads, because the log base ten-fold difference between the most (129,408) and fewest (108) genomically aligned reads across samples was large (3.08).

Using a custom python script, a list of all collisions was generated. Each transduction event was parsed for observations of ISs in the collision list. For each transduction event in which a collision IS was detected, the mean count of the IS for samples in which it was detected is recorded. For example, an IS at chr10:630,220 was observed in two transduction events. In the first transduction event, it was observed in two samples where it was represented by 100 and 197 genomically aligned sequence reads. It was observed in one sample from another

transduction event where it was represented by a 23 genomically aligned reads. The mean count of the IS in the first transduction event was 148.5 and 23 for the other transduction event.

The ratio of mean counts from each transduction event was compared to the maximum mean count from a single transduction event. If a transduction event had a mean count greater than or equal to one half of the maximum mean count for the IS, the IS was discarded from the dataset; otherwise, the IS was kept for the transduction event in which it had the highest count and removed from the other samples. In other words if the ratio of the maximum mean count to the next highest mean count was greater than 1:2 (½ or 0.5), the IS was discarded. If the ratio was less than 1:2, the IS was retained in the transduction event where it had the highest count and removed from all others. Returning to the previous example, 23 : 148.5 is 0.154. In this case, the non-maximum genomically aligned read count falls below 0.5 and the IS is retained in the first transduction event dataset and removed from the other transduction event dataset. Overall, from an initial set of 12,624 unique IS, 965 collisions (7.6%) were detected and 60 (0.5%) were unresolveable (removed from all datasets).

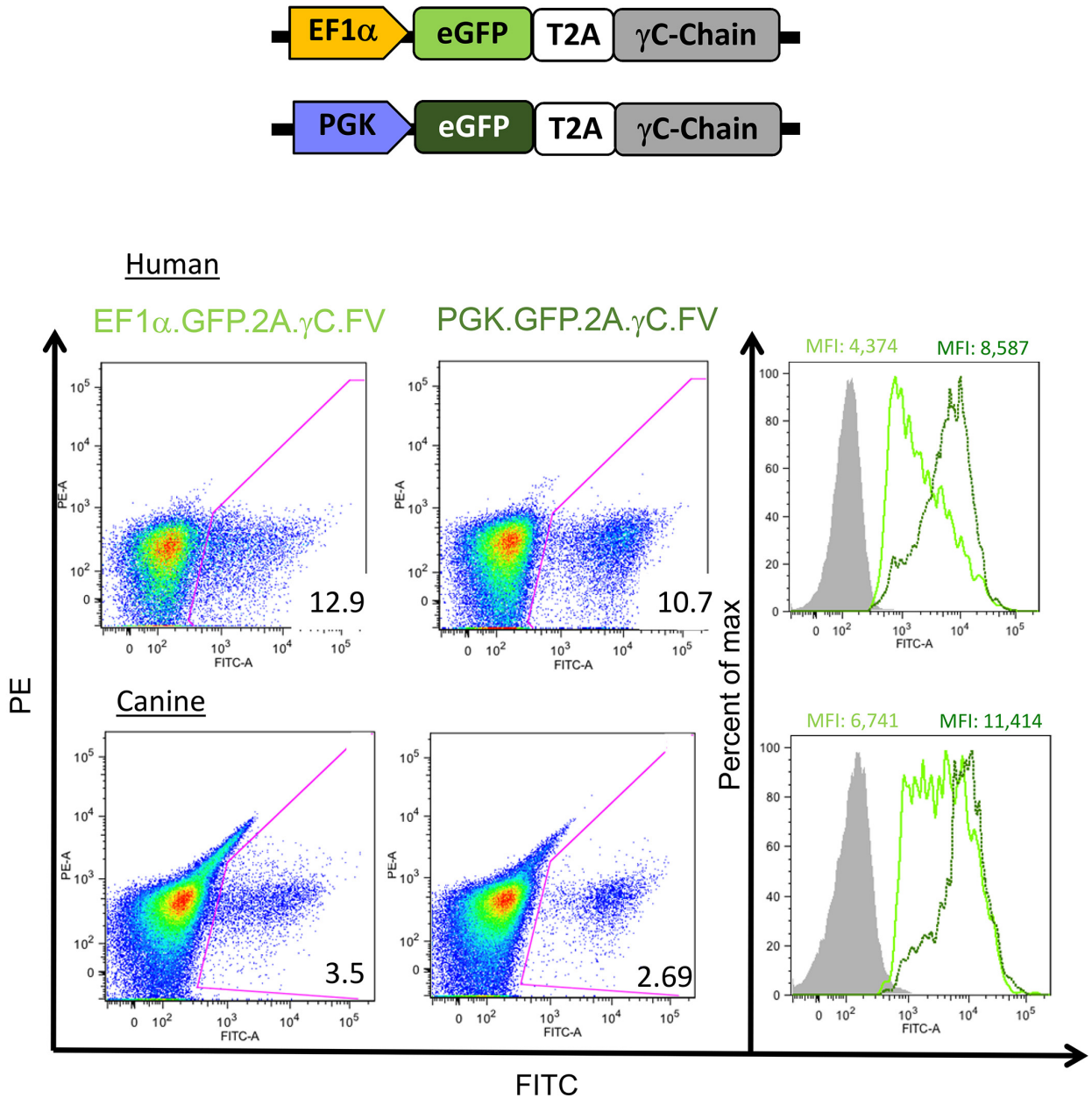
RIS normalization and figure generation. Contributions of each clone were normalized by dividing the number of integration site-associated sequence reads corresponding to that clone by the total number of integration site-associated sequence reads from the same sample. A custom R script was used to generate contribution graphs.

Supplemental Table S1.

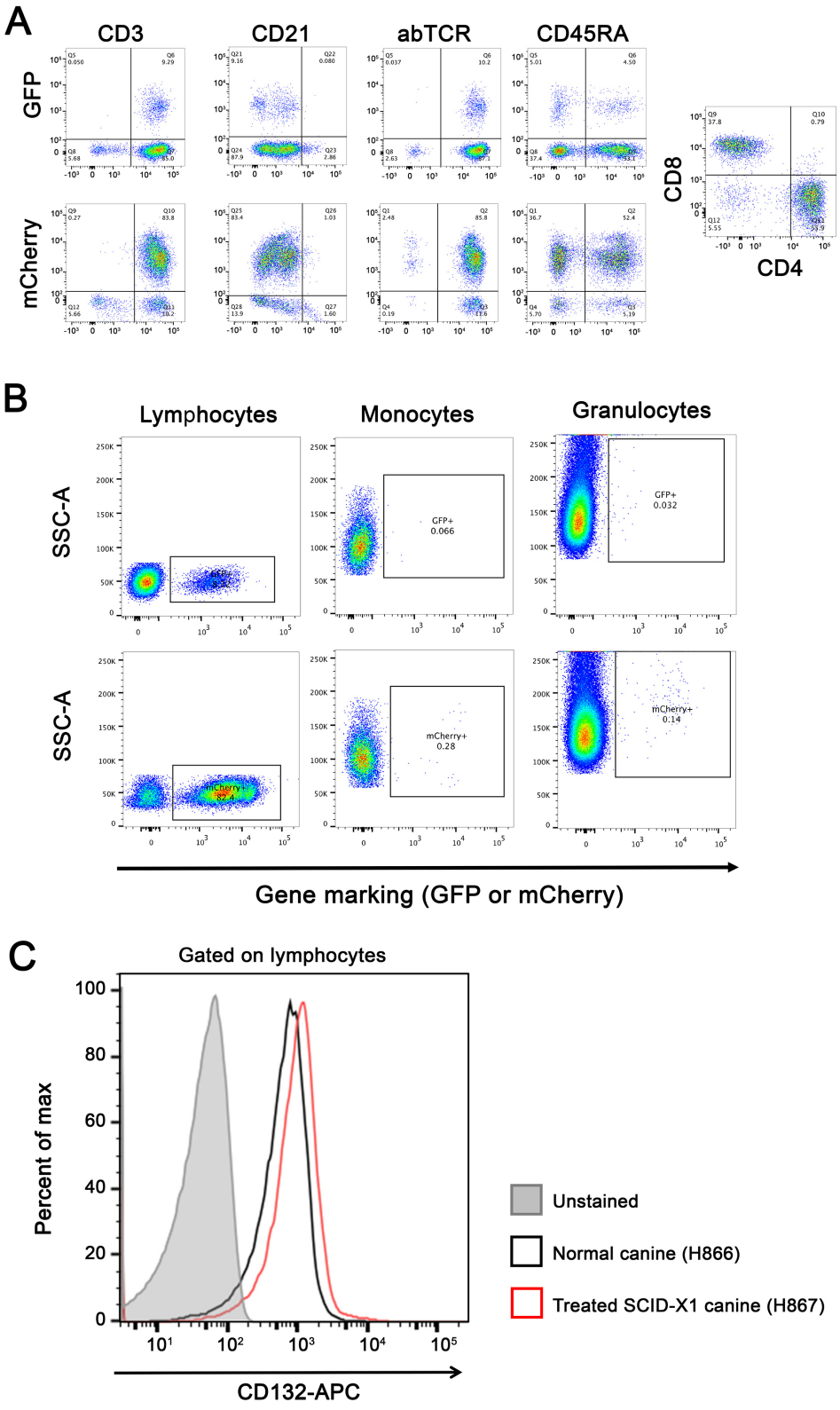
	Sample	Chr	Cstart	hit	% Frequency
a)					
R2258_Ovary	1	chr38	34522	23	4.28
	2	chr1	96161557	6	1.12
	3	chr3	4970058	3	0.56
	4	chr29	5211033	1	0.19
	5	chr10	20208336	1	0.19
	6	chr10	32047839	1	0.19
	7	chr30	31211	1	0.19
	8	chr15	2999557	1	0.19
	9	chr38	429229	1	0.19
	10	chr4	55077884	1	0.19
	11	chr11	24461252	1	0.19
	12	chr8	38928465	1	0.19
R2260_Testis	1	chr12	10898847	1	0.93
	2	chr5	15696040	1	0.93
	3	chr4	4027953	1	0.93
	4	chr38	16348389	1	0.93
	5	chr24	34261332	1	0.93
	6	chr38	16348388	1	0.93
	7	chr38	2094902	1	0.93
	8	chr7	70038184	1	0.93
	9	chr3	2499885	1	0.93
	10	chr4	69749947	1	0.93
	11	chr1	85010894	1	0.93
b)					
H867_semen (unique RIS)	1	chr7	69692227	1	0.19
	2	chr6	52806404	1	0.19
	3	chrUn_JH373264	138754	1	0.19
	4	chr11	57518357	1	0.19
	5	chr13	22499122	1	0.19
	6	chr8	17611696	1	0.19
	7	chr18	34135401	1	0.19
	8	chr2	54318446	1	0.19
	9	chr34	36370455	1	0.19
	10	chr11	44895108	1	0.19
	11	chr10	53723455	1	0.19
	12	chr22	5186427	1	0.19

Supplemental Table S2.

Breeding scheme	Female SCID (X^-X^-)	Female carrier (X^-X^+)	Male SCID (X^-Y)	Male normal (X^+Y)
R2260 (X^-Y) x H725 (X^-X^+)	0	1	0	4
R2260(X^-Y) x H727 (X^-X^+)	1	3	1	0
Observed (expected) frequency	20% (50%)	80% (50%)	20% (50%)	80% (50%)

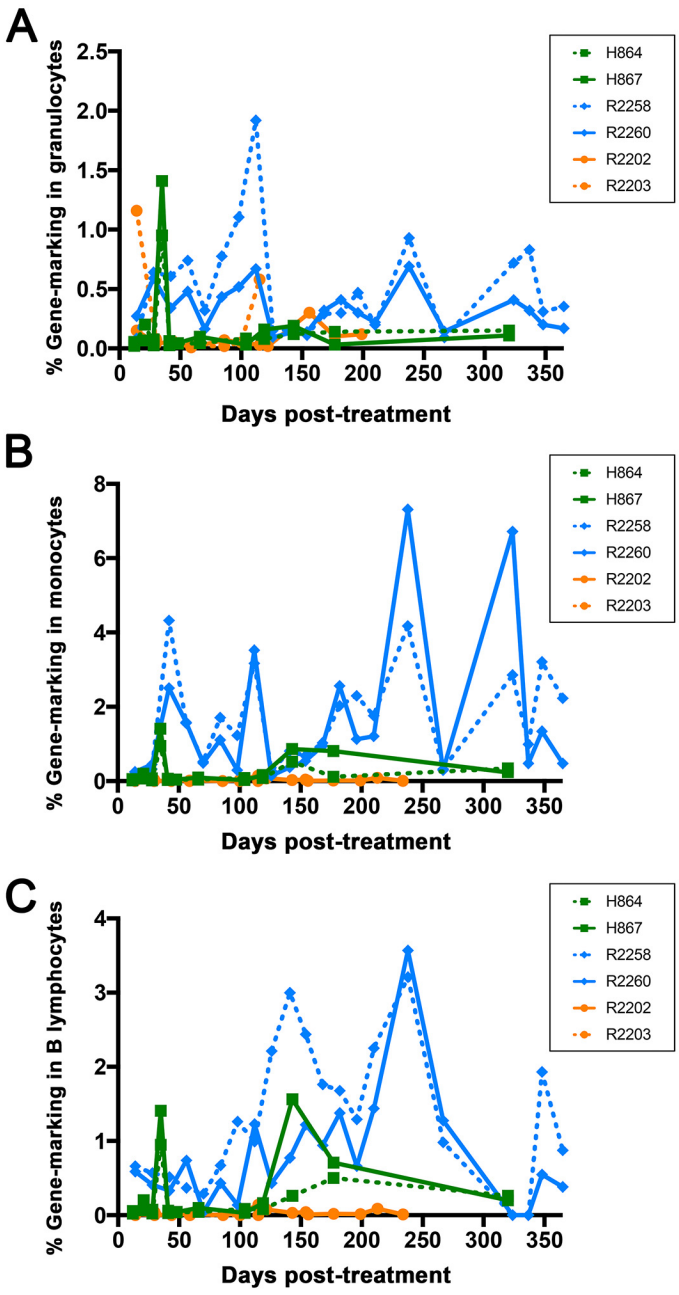


Supplemental Figure S1. Increased transgene expression in CD34+ cells transduced with PGK.GFP.2A.γC.FV. Human and canine CD34+ cells enriched from mobilized peripheral blood and steady state bone marrow, respectively, were transduced with PGK.GFP.2A.γC.FV or EF1α.GFP.2A.γC.FV at an MOI of 10. GFP expression was measured by flow cytometry at 6 days post-transduction and the fraction of GFP+ cells as well as mean fluorescence intensity (MFI) of the GFP+ fraction are shown for each vector. Grey histogram is from untransduced cells.

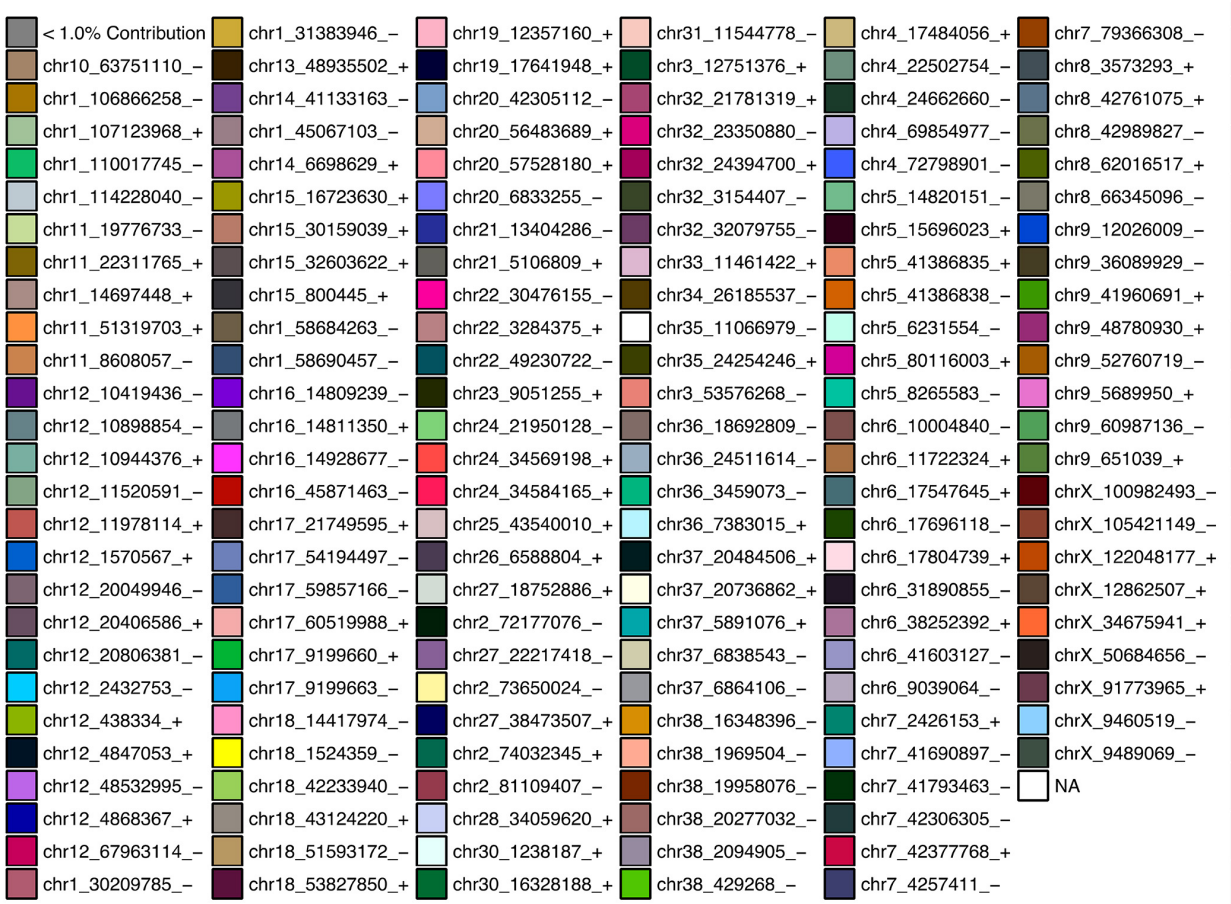


Supplemental Figure S2. Representative phenotypic panel for different blood cell lineages from treated canines. **A)** Peripheral blood gene marking as determined by mCherry or GFP in different lymphocyte subsets from animal R2258 at 620 days post-treatment. Fraction

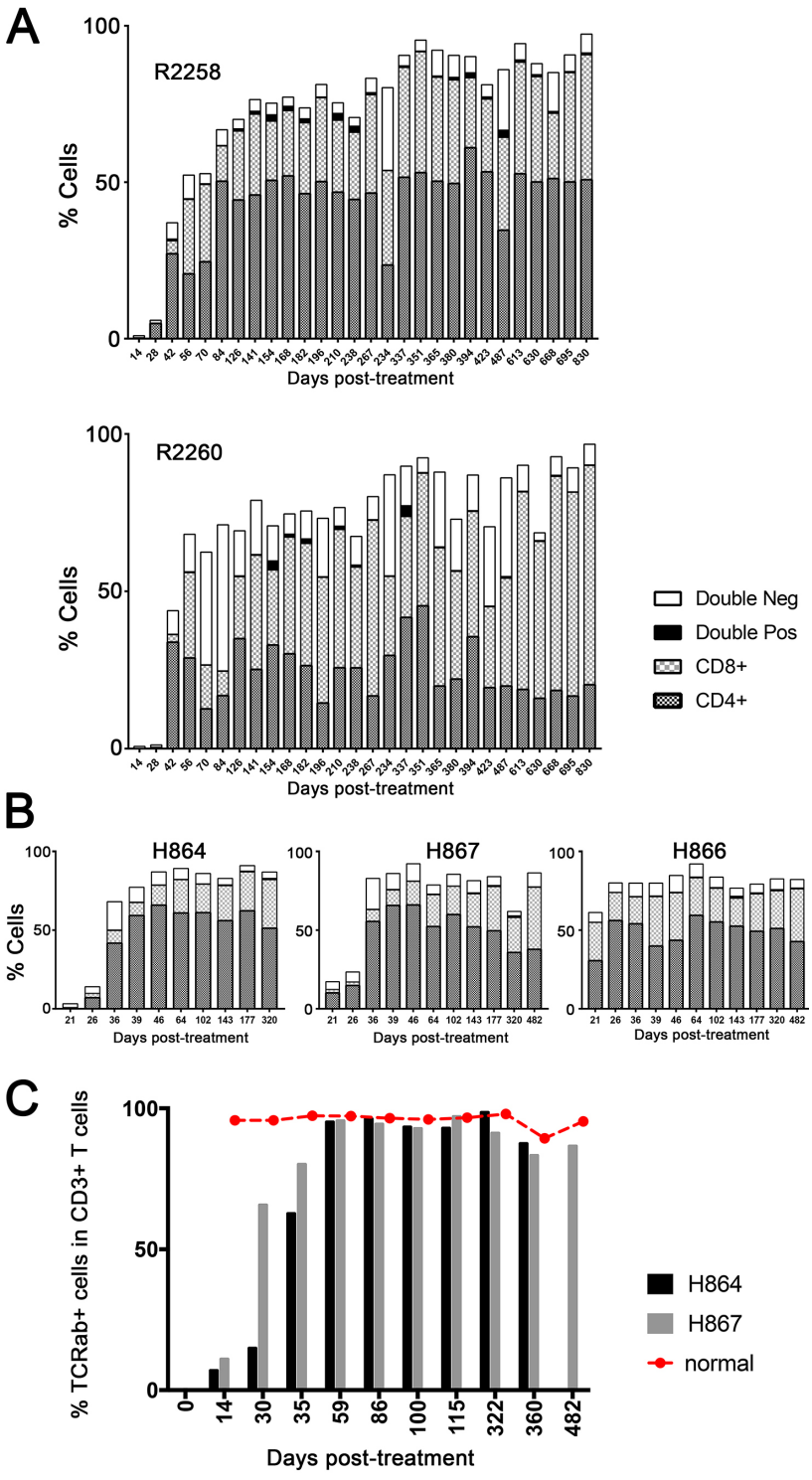
of CD4 and CD8 subsets within CD3+ cells is shown on the right. **B)** Gene marking in different cell lineages from the same animal defined by forward and side scatter. For monocytes and granulocytes, CD3+ cells were first gated out to exclude contaminating lymphocytes. **C)** Surface IL2Rg expression measured in peripheral blood lymphocytes from treated animal SCID-X1 H867 and the littermate control H866 at 550 days post-treatment.



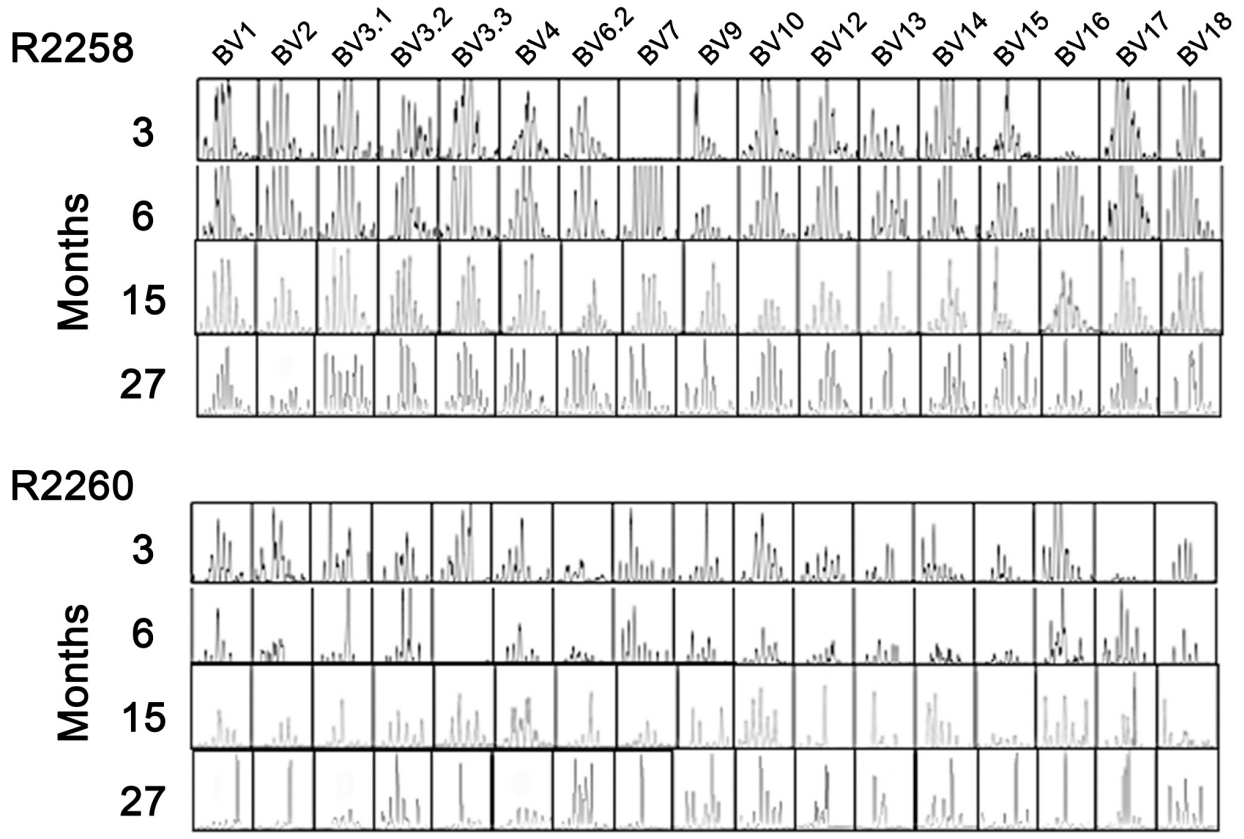
Supplemental Figure S3. Levels of gene marking detected in myeloid and B cells obtained from treated canines. Gene marking was determined based on phenotypic panel described in Supplemental Figure S2 in granulocytes (**A**), monocytes (**B**) and B lymphocytes **(C)** from peripheral blood of treated animals.



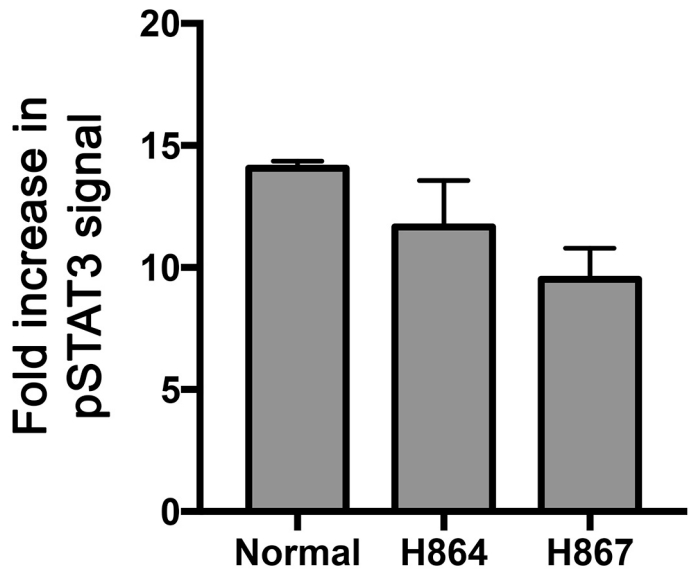
Supplemental Fiγure S4. Location of RIS events analyzed in peripheral blood of FV-vector treated canines. Each color shows the location of a unique integration event that corresponds to the color of the bar graphs depicted in **Figure 3**. Locations are defined as: chromosome (chr) number_bp location_sense (+) or antisense (-) orientation of insertion event.



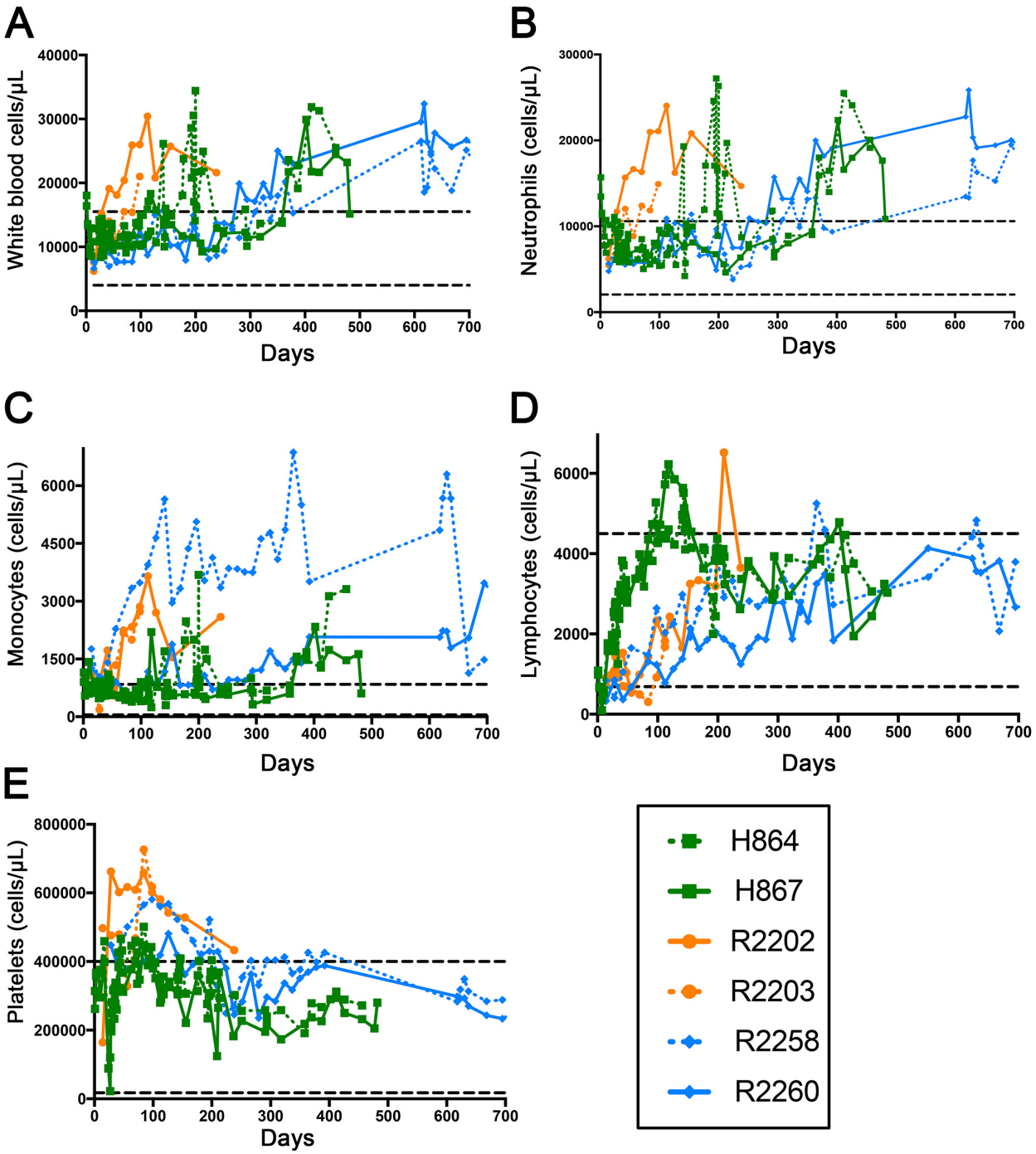
Supplemental Figure S5. CD4/CD8 and TCRab lymphocyte subtypes in FV vector-treated canines. A) Bar graph depicting CD4+, CD8+, CD4/CD8 double positive and CD4/CD8 double negative cells within the CD3+ population as determined by flow cytometry staining in non-mobilized/FV vector treated animals R2258 and R2260. **B)** Same analysis as in (A) in mobilized/ FV vector-treated animals H864 and H867. **C)** Percentage of TCR alpha/beta (ab) lymphocytes within the CD3+ population in animals H864 and H867, and in normal dog.



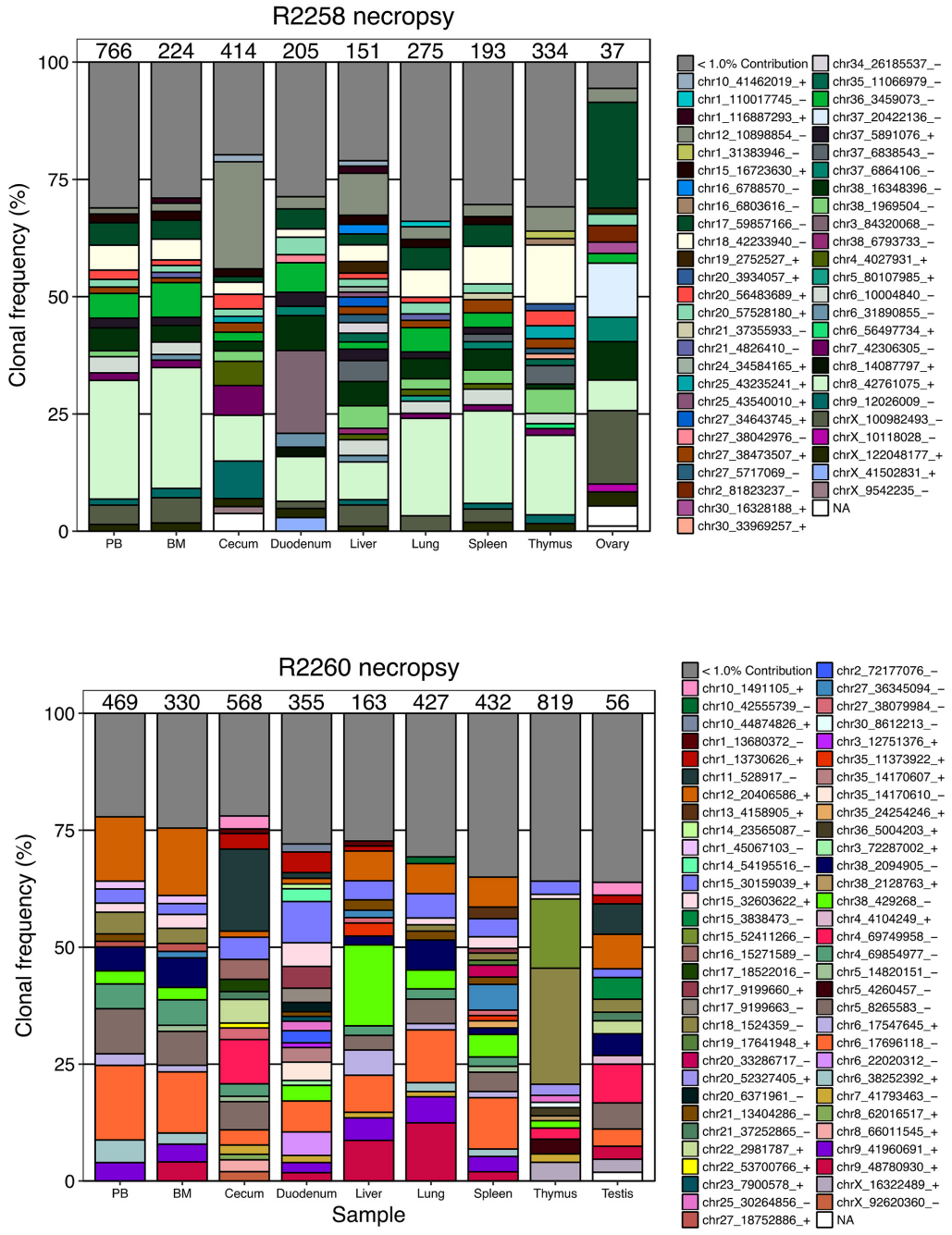
Supplemental Figure S6. TCR Vbeta spectratyping analysis in non-mobilized, FV vector-treated dogs. Rearrangement of the TCR beta chain was assessed by PCR amplification of complementary DNA using 17 different primer pairs (annotated on top) at different time points post-treatment (Month, Mo) in non-mobilized, FV-treated dogs R2258 and R2260.



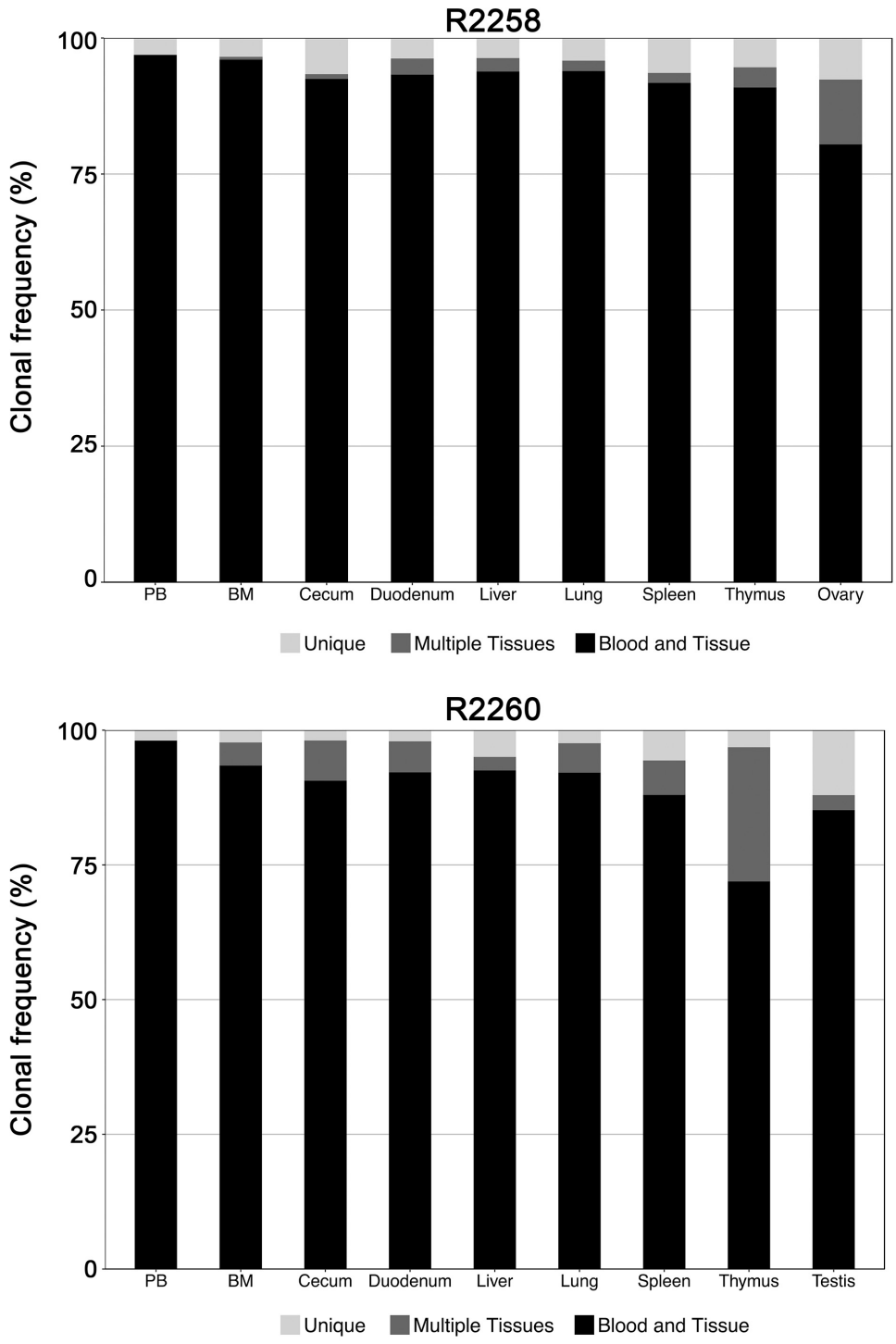
Supplemental Figure S7. pSTAT3 phosphorylation was assessed in mobilized animals H864 and H867 at 1-year post-treatment and compared to a littermate control (normal) as described in **Figure 6A**. Results are expressed as fold increase in cells expressing pSTAT3 when exposed to IL-21 as compared to untreated cells. Error bars show standard deviation for duplicate reactions and no statistical difference was noted between normal and H864 or H867 datasets using a Welch's t test ($p=0.32$ and $p=0.11$, respectively).



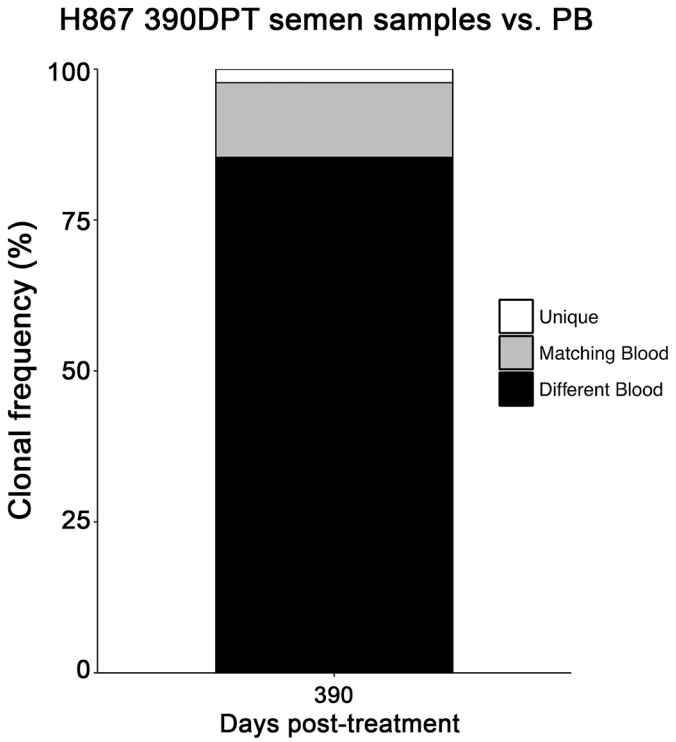
Supplemental Figure S8. Complete blood cell count analysis in all treated dogs. Graph depicts absolute numbers (counts per µL blood) of **A)** total white blood cells, **B)** neutrophils, **C)** monocytes, **D)** lymphocytes, and **E)** platelets in peripheral blood. Dotted line denotes average range of blood cell counts of normal dogs.



Supplemental Figure S9. Retroviral integration site (RIS) analysis of tissues obtained from R2258 and R2260 at time of necropsy. Clonal diversity in different tissues annotated on x-axis was compared to peripheral blood (PB), with total number of RIS events identified in the sample listed on top of the corresponding bar. In all graphs, unique RISs are plotted based on the number of times the RIS was sequenced and normalized to the percentage of total RISs captured in each sample for each animal. Captured RISs appearing at a frequency greater than 1% in each sample are represented by boxes in each bar. Boxes are colored white if they were identified in a single sample, or in matching colors if they were identified in more than one sample at a frequency higher than 1%. Colored numbers in gray portion depict RISs identified in more than one sample with a frequency lower than 1%. BM=bone marrow.



Supplemental Figure S10. Frequency of shared and unique RIS events in tissues obtained from animals R2258 and R2260. RISs from each tissue described in **Supplemental Figure S9** were compared with peripheral blood (PB) and with each other to determine overlap in multiple tissues (black), but absent from PB (dark grey), or that were unique to a specific tissue and absent from PB (light grey).



Supplemental Figure S11. Frequency of shared and unique RIS events observed in semen collected from mobilized canine H867. Semen was collected and gDNA was isolated and subjected to RIS analysis. A total of 16 RISs were identified in this sample and compared with all known peripheral blood (PB) RISs (total of 4,105 RISs; **Figure 3B**) to determine overlap (black). 12 integration sites were unique to the semen sample (white) but were only detected as single hit (**Supplemental Table 1B**), representing 2.28% of all hits. Since the animal remains alive and well, full biodistribution analysis is not yet available.