American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Vector integration and fate in the hemophilia dog liver multi-years following AAV-FVIII gene transfer

Tracking no: BLD-2023-022589R1

Paul Batty (Queen's University, Canada) Sylvia Fong (Queen's University, Canada) Matteo Franco (ProtaGene US Inc, United States) Choong-Ryoul Sihn (BioMarin Pharmaceutical Inc., United States) Laura Swystun (Queen's University, Canada) Saira Afzal (ProtaGene CGT GmbH, Germany) Lori Harpell (Queen's Universtiy, Canada) David Hurlbut (Queen's University, Canada) Abbey Pender (Queen's University, Canada) Cheng Su (BioMarin Pharmaceutical Inc., United States) Hauke Thomsen (GeneWerk (Protagene CGT) GmBH, Germany) Christopher Wilson (BioMarin Pharmaceutical Inc., United States) Loubna Youssar (ProtaGene CGT GmbH, Germany) Andrew Winterborn (Queen's University, Canada) Irene Gil-Farina (GeneWerk (Protagene CGT) GmBH, Germany) David Lillicrap (Queen's University, Canada)

Abstract:
Gene therapy using adeno-associated viral (AAV) vectors is a promising approach for the treatment of monogenic disorders. Long-term multi-year transgene expression has been demonstrated in animal models and clinical studies. Nevertheless, uncertainties remain concerning the nature of AAV vector persistence and whether there is a potential for genotoxicity. Here, we describe the mechanisms of AAV vector persistence in the liver of a severe hemophilia A dog model (male = 4, hemizygous, and female = 4, homozygous), more than a decade after portal vein delivery. The predominant vector form was non-integrated episomal structures with levels correlating with long-term transgene expression. Random integration was seen in all samples (median frequency= 9.3e-4 sites/cell), with small numbers of non-random common integration sites associated with open chromatin. No full-length integrated vectors were found, supporting predominant episomal vector-mediated long-term transgene expression. Despite integration, this was not associated with oncogene upregulation or histopathological evidence of tumorigenesis. These findings support the long-term safety of this therapeutic modality.

Conflict of interest: COI declared - see note

COI notes: S.F., C-R.S., C.S., and C.W. are employees and stockholder of BioMarin Pharmaceutical Inc. P.B. has received research support from BioMarin, and consulting fees or honoraria from BioMarin, Octapharma, Novo Nordisk, CSL Behring, Pfizer, and Institute for Nursing and Medication Education (IMNE). D.L. has received research support from BioMarin, CSL-Behring, and Sanofi and has received consulting fees or honoraria from BioMarin, CSL-Behring, Novo Nordisk, Pfizer, and Sanofi. L.L.S. has received consulting fees from BioMarin. H.T., and I.G-F. were employees of ProtaGene CGT GmbH. S.A. and L.Y. are employees of ProtaGene CGT GmbH and M.F. is an employee of ProtaGene US Inc. The remaining authors declare no competing financial interests.

Preprint server: No;

Author contributions and disclosures: Authorship contribution: D.L., P.B., and S.F. designed the research study; L.H., A.P., and A.W. cared for the animals in the study. P.B., D.H., S.F., and D.L. analyzed the data; P.B., S.F., and D.L. wrote the first draft of the paper and L.L.S. revised the manuscript. M.F., C-R.S., and I.G.-F. conducted and interpreted experiments. S.A., H.T., C.S., L.Y., and C.W. performed the bioinformatic analyses. All authors reviewed and critically edited the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: The TES and LAM-PCR data is available at the NIH Sequence Read Archive. (BioProject ID: PRJNA1074750, https://www.ncbi.nlm.nih.gov/bioproject/1074750). Data sharing is also available through email with the corresponding author

Clinical trial registration information (if any):

Vector integration and fate in the hemophilia dog liver multi-years following AAV-FVIII gene transfer

Authors: Paul Batty[1,2], Sylvia Fong[1,3], Matteo Franco[4], Choong-Ryoul Sihn[3], Laura L. Swystun[1], Saira Afzal[5], Lorianne Harpell[1], David Hurlbut[1], Abbey Pender[1], Cheng Su[6], Hauke Thomsen[5,7], Christopher Wilson[6], Loubna Youssar[5], Andrew Winterborn[1], Irene Gil-Farina[5], David Lillicrap[1]

Institution Addresses:

[1]Department of Pathology and Molecular Medicine, Queen's University, Kingston, Ontario, Canada

[2]Cancer Institute, University College London, London, UK

[3]Research, BioMarin Pharmaceutical, Novato, CA, USA

[4]ProtaGene US Inc, Burlington, MA, USA

[5]ProtaGene CGT GmbH, Heidelberg, Germany

[6]Data Science, BioMarin Pharmaceutical, Novato, CA, USA

[7]MSB Medical School Berlin, Berlin, Germany

Corresponding Author: Dr David Lillicrap, Department of Pathology and Molecular Medicine, Queen's University, Kingston, Ontario, Canada. Telephone: +1-613-533-6342 Fax: +1-613-533-2907 Email: david.lillicrap@queensu.ca

The TES and LAM-PCR data is available at the NIH Sequence Read Archive. (BioProject ID: PRJNA1074750, https://www.ncbi.nlm.nih.gov/bioproject/1074750). Data sharing is also available through email with the corresponding author

Word Count (Abstract): 148 words

Word Count (Body of Text): 4093 words

Figures: 5, **Tables:** 4

Running Short Title: AAV-FVIII Persistence in Canine Hemophilia

Scientific Category: Gene therapy

Keywords: Gene Therapy, AAV, Hemophilia A, Integration, Episomal

Key Points:

1. In hemophilia A dogs, AAV-cFVIII vectors predominantly persist in the liver long-term in non-integrated episomal forms.
2. AAV vector integration was seen at low frequencies, and occurred commonly in in areas of open chromatin, with no effect on gene expression.

Abstract

Gene therapy using adeno-associated viral (AAV) vectors is a promising approach for the treatment of monogenic disorders. Long-term multi-year transgene expression has been demonstrated in animal models and clinical studies. Nevertheless, uncertainties remain concerning the nature of AAV vector persistence and whether there is a potential for genotoxicity. Here, we describe the mechanisms of AAV vector persistence in the liver of a severe hemophilia A dog model (male = 4, hemizygous, and female = 4, homozygous), more than a decade after portal vein delivery. The predominant vector form was non-integrated episomal structures with levels correlating with long-term transgene expression. Random integration was seen in all samples (median frequency= 9.3e-4 sites/cell), with small numbers of non-random common integration sites associated with open chromatin. No full-length integrated vectors were found, supporting predominant episomal vector-mediated long-term transgene expression. Despite integration, this was not associated with oncogene upregulation or histopathological evidence of tumorigenesis. These findings support the long-term safety of this therapeutic modality.

Introduction:

Current treatment to prevent bleeding in patients with severe hemophilia A involves regular intravenous infusion of FVIII concentrates or subcutaneous injection of the bispecific antibody emicizumab (Roche, USA). The monogenic recessive nature of hemophilia has made it an ideal candidate for a gene therapy approach. The most advanced approach for *in vivo* gene delivery for hemophilia utilizes adeno-associated viral (AAV) vectors.[1] Pre-clinical studies in large animals, including hemophilia dogs, have described therapeutic transgene expression for over a decade following a single AAV vector infusion.[2] Although median follow-up is shorter in clinical studies (hemophilia A: 6 years[3,4] and hemophilia B: 6.7 years[5]), multiple studies are ongoing.[1] AAV5-FVIII gene therapy (valoctocogene roxaparvovec, BioMarin, USA) is approved by the U.S. Food and Drug Administration and conditionally approved by the European Medicines Agency for the treatment of adult patients with severe hemophilia A.

The origin of long-term transgene expression from recombinant AAV (rAAV) vectors, and potential for genotoxicity remain unclear. In young and selected mouse models, recurrent rAAV integration has been reported at the *Rian* locus, with development of hepatocellular carcinoma.[6,7] In hemophilia dogs, clonal hepatocyte expansion was reported following rAAV treatment, albeit without tumorigenesis.[8] Additionally, hepatocellular carcinoma has been reported in an individual with hemophilia B treated with an rAAV vector.[9,10] Although vector sequences did not appear to play a pathogenic role in this case, the nature of long-term rAAV vector persistence remains poorly understood. To address this issue, we studied long-term genomic implications of rAAV gene therapy a decade after treatment in the hemophilia dog model. This represents the longest follow-up reported to date post-AAV gene therapy, at the natural lifespan of this animal model. These dogs

possess a spontaneous *F8* mutation and bleeding phenotype similar to humans with severe hemophilia A.[11,12] Here, we describe the frequency, sites of integration, and genomic implications of long-term rAAV vector persistence.

Methods:

AAV vector

Canine FVIII AAV vectors were constructed and administered as described previously.[13] Briefly, the AAV-cFVIII expression cassette consisted of a liver-specific transthyretin (TTR) promoter, synthetic intron, non-codon optimized canine B-domain deleted SQ-FVIII cDNA, and synthetic polyadenylation sequence (Supplementary Figure 1; vector map located in Supplementary Information).[2,13] Liver samples (multiple samples from different liver regions) were flash-frozen in liquid nitrogen and stored at -80^{o}C.

Quantification of AAV-canine FVIII vector genome forms

AAV-canine FVIII (cFVIII) vector genome extraction and quantification was performed using AllPrep DNA/RNA micro kits (Qiagen, Germantown, MD) and drop-phase droplet digital polymerase chain reaction (ddPCR) assays[14] using different primers/probe sets. The cR1-cR11 linked assay measures full-length AAV-cFVIII genomes capable of giving rise to stable cFVIII transcript and the cSQ assay measures overall vector genomes (full-length and fragments) (Supplementary Table 1). DNA was digested with KpnI to separate vector genome units from concatemers. Plasmid Safe™ ATP-Dependent DNase (PS-DNAse) (Lucigen, Middleton, WI) was used to hydrolyze linear DNA and isolate circular DNA.[14] ddPCR of the endogenous gene canine transferrin receptor protein 1 (cTfrc) was performed to provide a normalization reference.

Integration site analysis by TES and bioinformatic analysis of TES amplicons

TES (target enrichment sequencing)[15] was performed by double-capture using two different RNA 120 bp-long bait sets, both designed based on 8x tiling (Supplementary Figure 2). The first bait set was homologous to the whole vector sequence and the second covered only vector regions diverging

between the vector and the dog genome: ITRs, promoter-transgene junction, transgene exon-exon junctions and the polyA. TES was performed first on control samples consisting of vector plasmid spiked into untransduced canine genomic DNA simulating a vector load of 1 vector genome/cell and non-spiked control. Samples were analyzed in duplicate using 1000 ng DNA per replicate. DNA was sheared to ~500bp length using ultrasonicator (Covaris) and fragment size was verified by TapeStation (Agilent). Libraries were prepared using the Agilent SureSelectXT2 kit in line with manufacturer's instructions. Additional hybridization and PCR enrichment steps were performed on the final library. After library concentration and size distribution evaluation, libraries were sequenced by 2x250bp symmetric paired-end on the Illumina MiSeq platform.

GENE-IS was used for bioinformatic analyses of linear amplification-mediated (LAM)-PCR- and TES-derived data.[16] Raw sequence data were filtered according to 100% sample barcode identity and sequence quality (Phred 20 for TES and Phred 30 for LAM-PCR). Sequencing reads were aligned to the canine reference genome (canFam3 or canFam4) and vector for IS analysis for both methods.

Vector coverages for each replicate were analyzed and illustrated in Integrative Genome Viewer (IGV). The average vector coverage, normalized by the average coverage on the subgenomic regions, was also used to estimate the VCN for the analyzed samples. Sequence coverage on the vector and on five integration sites bearing the highest frequencies (≥15%) was analyzed and illustrated in IGV. Where applicable, the liftover of IS locations from dog to the human genome (hg38) was carried out using the UCSC liftover tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver), and their proximity to the cancer genes was assessed using an in-house bioinformatic tool using the cancer gene data available from the Cancer Gene Census database (http://cancer.sanger.ac.uk/census; updated to the 5th of September 2019, v90).

Statistical analyses

Descriptive data are summarized as mean, median, range, and frequencies. Correlation was performed using Pearson (parametric) or Spearman (non-parametric). Analyses of gene expression were performed using mixed effects models (Dunnett's adjusted) to account for variance within and

between dogs. All tests were two-tailed with a p-value of <0.05 used for statistical significance. Analyses were performed using GraphPad Prism v. 9.0.0 for Windows (GraphPad Software, CA), and SAS software v 9.4 for modeling the gene expression.

For additional Methods please see Supplementary Information.

Results:

Eight severe hemophilia A dogs were treated with a single infusion of AAV-cFVIII (6e12 – 2.7e13 vg/kg) at a median of 9.5 months age (Table 1).[2] After a median follow-up of 10.7 years (range: 8.2-12.0 years), transgene-derived FVIII:C (chromogenic) activity ranged from 1.8 – 8.6% in six responding dogs. Improvement in the bleeding phenotype was seen in all eight dogs. Analysis of post-mortem samples by qPCR/RT-PCR demonstrated the liver as the primary source of vector-derived FVIII expression. There was no evidence of chronic liver disease or liver tumors at post-mortem. A full report of the phenotypic outcomes from this study has been previously described.[2]

As an extension to these studies, the regional distribution of AAV-cFVIII vector genome copies and cFVIII mRNA expression was evaluated using DNA and mRNA co-isolated from multiple liver samples by ddPCR. In the majority of AAV-cFVIII treated dogs, little intra-animal regional variation in hepatic vector copy number (VCN) or cFVIII mRNA expression was seen, whereas in some, more marked regional variation was present (e.g., JUN) (Figure 1).

Preclinical and clinical studies of AAV5-hFVIII-SQ transduced livers have demonstrated that circularized monomeric and concatemeric episomes are the major AAV DNA forms associated with long-term transgene expression.[14,17] Here, we evaluated the vector genome forms mediating long-term expression (i.e., episomal or integrated) using three orthogonal methodologies. Firstly, liver circular episomal AAV-cFVIII was enriched by digestion with PS-DNAse and followed by KpnI to allow enumeration of transgene cassettes in the circular genomes.[18] Absolute quantification of full-length vector units was performed using drop-phase ddPCR with primers/probes to the proximal end of the 5' and 3' ITRs (D-segment).[17] Full-length circular episomes were detected using ddPCR in the liver of all AAV-cFVIII treated dogs (Figure 2A). Importantly, this assay underestimates the number of

circular vector genomes due to shearing/linearization of large concatemeric episomes during DNA extraction, and degradation of 25-35% of episomes by PS-DNAse.[17,18] A significant positive correlation was observed between the circular full-length episomal vector levels and terminal FVIII:C ($r=0.812$, $p=0.0078$, Figure 2B) and cFVIII RNA ($r=0.83$; $p<0.0001$, Figure 2C), suggesting FVIII expression was primarily derived from episomal vectors. These findings were confirmed using Southern blotting (Figure 2D), performed on a single sample from each dog, demonstrating primarily head-to-tail configured full-length episomes with band intensities in each sample correlating with circular genome copies measured by ddPCR (Supplementary Figure 3).

Complementary analysis was performed using targeted enrichment next-generation sequencing (TES, Figure 2E) (two samples/dog). Sequencing reads containing vector-vector (VV) or vector-genome (VG) junctions were analyzed providing estimates of the proportion of episomal and integrated forms. The majority (95.4%, range: 83.6-98.9%) of sequencing reads resulted from vector-vector (approximated as episomal) reads, with only 4.6% (range: 1.2-16.4%) resulting from vector-genome (integrated) reads (Table 1, Figure 2E). Previous in vitro studies of wild-type AAV have suggested that integrated vector concatemers are infrequent.[19] Collectively, results obtained from orthogonal methods support that AAV predominantly persists within the liver in an episomal form a decade after vector delivery. However, we cannot rule out the presence of integrated concatemers, as was recently reported using long-read sequencing in non-human primates treated with AAV8 vectors encoding rhLDLR, hLDLR, or GFP.[20]

Low Frequencies of Integrations Were Seen Predominantly in Intergenic Regions

Although AAV-cFVIII vectors predominantly persist episomally, integration was seen in all liver samples analyzed. Integration frequencies were calculated using TES and confirmed using an orthogonal LAM-PCR approach (Table 1 & Supplementary Figure 2). Using TES, there were 43,246,183 sorted sequencing reads, with 11,533 IS reads. This resulted in 5,746 uniquely mappable IS and a mean integration frequency of 9.3e-4 IS/cell (range 1.21e-4 – 2.72e-3), equating to 0.93 integration events per 1000 cells (Table 1). Integration frequency correlated ($r=0.94$, $p=0.0005$) with

hepatic vector genome copies. Characterization of integration site (IS) locations relative to annotated genes, demonstrated the majority (93.8%) occurred in intergenic regions of the canine genome (Figure 3A). An overview of unique in-gene insertions (total=355, of which *F8* = 195) is provided in Table 3 and Supplementary Table 3.

Evaluation of Integration Site Abundance

Individual sample IS frequencies were determined and the ten most abundant (top-10 IS) are shown in Figure 3B-3D and Supplementary Figure 4. Differences in IS profiles were seen between dogs and samples. Clonal abundance was evaluated from frequencies of unique IS in individual samples. This provides approximation of abundance, with thresholds based on data from retroviral integration studies.[21] Most samples (10/16) exhibited integration profiles where each of the top-10 IS had a frequency of <10% of the total IS sequence count (average=1.59%) (Figure 3B). For the 6 samples with IS frequencies >10%, the average frequency for the top-10 IS was 6.03% (Figure 3C). IS with frequencies >30% were only seen in single samples in the two non-responding dogs (MG and ANG) (Figure 3D) with plasma FVIII:C below the limit of detection but improved whole blood clot times post-AAV-cFVIII as previously described.[2] MG-1 had two IS ~162kb upstream of *CCND1* (33.0% and 53.0%) and ANG-2 had one IS ~424kb upstream of *MIR320* (31.7%). In these two samples, the top-10 IS accounted for >90% of the total IS, although this is skewed in MG-1 by the low total number of IS. Both MG and ANG had low levels of total AAV-cFVIII vgDNA (Figure 1A), and circular full-length vgDNA (Figure 2A) suggesting that low rates of vector integration may account for the increased IS frequency observed. Data from the orthogonal LAM-PCR strategy is provided in Supplementary Figure 5. In summary, these datasets demonstrate the presence of specific IS clones at higher abundances in two samples when compared to the overall polyclonal integration profile observed in the majority of samples.

No Deletions Seen in the Host Genome at Integration Sites

We then evaluated the implications of AAV-cFVIII integration on adjacent host genome regions for the higher frequency (>15%) IS: *MIR320* (ANG-2, n=2), *MIR1296* (JUN-1, n=2), *KCNIP2* (ALX-2), *MET*

(MZ-1) and *CCND1* (MG-1) (Supplementary Table 4). Using TES short-read sequencing, possible deletions were seen for *MIR320* (ANG-2: 2nd IS 3bp deletion), *MIR1296* (JUN-1: 2nd IS 176bp deletion), *MET* (MZ-1: 185bp deletion) and *CCND1* (MG-1: 70bp deletion) (Supplementary Figure 6). Long-read nanopore sequencing was used to further investigate these putative deletions. After sequence read alignment, no deletions were seen, which likely represent short-read sequencing artifacts using TES. In this small sample of IS, these data suggest that integration occurred without significant impact on the surrounding host genome.

Common Integration Sites Were Seen with No Effect on Gene Expression

Analysis of the distribution of IS demonstrated these occurred mostly randomly throughout the canine genome (Figure 4A). A non-biased systems biology approach was then used to evaluate whether IS formed into clusters, or common integration sites (CIS). Using the TES and LAM-PCR methodologies, 37 and 4 common integration sites with a CIS order ≥5 were seen, respectively (Table 2 & Supplementary Table 5). The most frequent CIS are shown in Figure 4B, on chromosome 14 (*ABCB1*), 28 (*KCNIP2*), and X (*F8* and *CLIC2*). The CIS predominantly occurred in intergenic regions, with the only common in-gene CIS, occurring within the *F8* and *ALB* loci. No differences were observed in the hepatic gene expression of *ABCB1*, *CLIC2*, *KCNIP2,* and *ALB* in AAV-cFVIII treated, compared to normal or untreated hemophilia dogs (Figure 4C-4E, Supplementary Figure 7 and Supplementary Table 6). Of note, samples used were not co-isolated from those used for TES and LAM-PCR, and multiple samples were analyzed per dog to ensure accurate representation. Importantly, none of these CIS are located within genomic regions linked with wild-type AAV integration or regions associated with adverse events (clonal outgrowth or malignant transformation) in clinical gene therapy trials.[21-28] No evidence of tumorigenesis was observed on histopathological assessment (Supplementary Table 7).[2]

No Enrichment of Integration is Seen Relative to Known Human Cancer Genes

We then evaluated whether IS were enriched in proximity to known oncogenes. Since there is no canine cancer gene database, a liftover of IS referenced to canFam4 was performed to the human

genome (hg38). Successfully lifted over IS (3,538/5,755) coordinates were referenced against 7 human cancer gene databases (Supplementary Table 8) for locations proximal (≤100kb) to the transcriptional start site (TSS) of known oncogenes. Using the curated Sanger database (https://cancer.sanger.ac.uk/cosmic), 8.4% (n=297) IS occurred <100kb from a known cancer gene. This included 13 of the top-10 IS, proximal to *MET*=5, *MTCP1*=4, *NCOA4*=1, *LCP1*=1, *MALAT1*=1 and *NF2*=1 (Supplementary Table 9); all with relative frequencies <20%. The highest IS frequency (18.2%) was seen in MZ-1, proximal to the Hepatocyte Growth Factor Receptor (*MET*) locus, however, no change in hepatic *MET* expression was seen in liver samples from any of the AAV-cFVIII treated dogs compared to normal or untreated hemophilia dogs (Figure 4F). IS frequencies relative to cancer genes were then compared to 10,000 simulated data sets. This demonstrated IS frequencies fell within the predicted normal range (0.065 - 0.084), indicating no enrichment in IS proximal to human cancer genes (Figure 4G).

Common Integration Sites Occurred at Chromatin Accessible Regions of the Canine Genome

We then went on to evaluate genomic mechanisms underlying these CIS. Firstly, top-10 TES CIS (canFam3) were analyzed to determine whether there was sequence homology between IS regions and vector sequences. Excluding CIS within the *F8* gene, no homology was seen comparing CIS sequences to vector sequences. In view of homology between the cFVIII vector and the native canine *F8* gene, we analysed IS mapped to the *F8* locus. *F8* intronic IS were regarded as true IS due to lack of these elements within the vector and *F8* exonic IS were considered as either true integration events or an internal vector-vector (VV) junction sequence. Although some *F8* IS could not be excluded from being false positives due to exonic location, integration was seen within introns of the native canine *F8* gene (Table 3). We then evaluated whether integration occurred more frequently in regions of increased chromatin accessibility in the canine genome (canFam4) using ATAC-Seq (Transposase-Accessible-Chromatin sequencing). Analyses of regions surrounding IS sites were performed on dog liver samples (n=7, Figure 4H) and associations of ATAC-Seq scores were calculated using receiver operating characteristic (ROC) scores compared to a random sample. ROC

scores varied from 0.4 (significant negative association) to 0.6 (significant positive association) [29]. Although, for regions >10 kbp, we did not observe many deviations in ROC score from 0.5, increased ROC scores were seen for IS regions of 1 kbp down to 50 bp. This supports the hypothesis that focused regional enhanced chromatin accessibility has a facilitatory influence on vector integration. To determine whether DNA methylation plays a role in AAV integration, IS distributions were compared to whole genome bisulfite sequencing (Methyl-seq) results (3 canine datasets, 47,000,000 sites) and the average measurement of the overall enrichment for corresponding genomic regions determined by ATAC-Seq. No association with ROC values for methylation were seen. These results suggest an association between integration and regions of higher chromatin accessibility, which could enable passive vector integration at locations where DNA is loosened from the nucleosome complex.

No Full-Length Integrated Recombinant AAV Vectors Were Detected

Finally, we evaluated whether full-length integrated AAV-cFVIII could be detected which could result in transgene expression. TES sequencing reads were analyzed to determine whether these were homogeneously represented and/or if recurrent deletions had occurred. Most samples (15/16) demonstrated homogeneous coverage (Supplementary Figures 8 and 9). One sample (ANG-2) displayed high coverage in the promoter and poly(A), with no transgene reads, indicating partially deleted vector genomes. Irregular coverage profiles were seen for two other samples (ANG-2 and MG-1), with peaks downstream of the 5' ITR, promoter, and at exon-exon junctions, but without interruptions that could imply vector breakage. Although this approach provides an impression of overall vector sequence coverage, it cannot provide information on genome contiguity.

To address this, long-read sequencing was performed in 8 higher (>10%) frequency IS (Supplementary Table 10). Sorted reads were aligned to their reference integration sites and demonstrated that no full-length integrated vector copies were found (Figure 5, Supplementary Figure 10). For the IS investigated, six had structures comprising the 5' ITR and most of the promoter and three showed parts of the poly-A and 3'ITR (Figure 5A). For most (7 out of 8), the 5' ITR and

parts of the promoter were retrieved (Figure 5B), and in 3 samples, parts of the poly(A) and 3' ITR were also retrieved (Figure 5C). Examples of sequencing reads from different size classes (0-500bp, 501-1000bp, and >1000bp) were investigated to confirm these identified structures. For most samples, extracted longer reads confirmed the integrated structure shown by the coverage plots. In particular, several vector fragments were retrieved suggesting vector recombination within integrated forms. However, this approach is not able to distinguish between truly rearranged forms and chimeric reads arising from the library preparation. Finally, to exclude sequencing bias from the ITRs and vector sequences that might limit full-length sequence detection, a cF8x104 plasmid control (AAV_TTR_FVIII) was amplified and sequenced using the same protocol (Figure 5D, Supplementary Figure 11). This demonstrated uniform coverage of the full-length plasmid control, excluding the possibility that full-length integrated vectors could not be retrieved. These findings further support the observation that full vector sequences predominantly (and possibly exclusively) persist in episomal rather than as integrated forms.

Discussion:

This study has provided detailed analyses of the fate and forms of vectors over a decade post-rAAV-cFVIII treatment. Using orthogonal strategies, we demonstrated that the majority of vectors exist long-term in episomal structures that correlate with transgene RNA expression. These findings are in keeping with descriptions of episomal forms seen 2-4 years following AAV5-hFVIII-SQ administration in humans.[14] Integration frequencies are similar to those reported after shorter follow-up for an AAV2/5-cohPBGD vector used for treatment of acute intermittent porphyria in non-human primates (4 weeks, 2.0e-4 IS/cell) and humans (52 weeks, 1.17e-3 IS/cell) suggesting integration events occur early after gene transfer and remain stable over time.[30] These integration frequencies are lower than seen with lentiviral vectors and the spontaneous somatic mutation rate.[31] Analyzing integrated vector sequences, we only found fragmented and rearranged copies which is similar to the findings from other investigators.[8] Collectively, this provides evidence that long-term FVIII expression was

episomally derived, although we acknowledge further study is required on the source of transgene expression (episomal or integrated) based on a recent study reporting the presence of integrated AAV concatemers in non-human primates.[20]

Although integration occurred throughout the canine genome, we identified common integration sites. Using two different sequencing strategies, these were identified in proximity to five genes: *KCNIP2, CLIC2, ABCB1, F8, and ALB*, all of which are liver-expressed. Recurrent rAAV IS have been described in murine and canine studies. In mice, the predominant site of recurrent integration is within the *Rian* locus, with the majority occurring in the *mir341* locus, which is not encoded within the canine or human genomes.[32] In a similar hemophilia A dog study, recurrent IS were seen in proximity to *EGR2*, *EGR3*, *CCND1*, *ALB,* and *DUSP1.*[8] An overlap is seen in the genomic loci of common integration sites in both dog studies, with the exception of *DUSP1* (Table 4).[8] Considering that IS are assigned to the closest annotated gene, the different canine genome annotation tracks used in this and the previous study may account for these differences. Nonetheless, we did not observe evidence of dysregulated expression of adjacent genes for the common IS seen in AAV-cFVIII treated dogs, compared to normal or untreated hemophilia dogs (Supplementary Table 11). With albumin being highly expressed in hepatocytes, this locus has an accessible chromatin landscape facilitating vector integration. This is supported by our findings of high ATAC-Seq scores in CIS regions, suggesting that integration occurs through "passive invasion" of the genome at regions of chromatin accessibility.

Integrations into the native *F8* and *CLIC2* (located ~240 kb 5' of *F8*) loci seen in these dogs may also be driven by homologous recombination and appear distinct from vector-into-vector events reported in a similar study.[8] Integrations at *ABCB1* and *KCNIP2* are more difficult to explain. GTEx analysis of human liver shows low expression of *ABCB1* and *KCNIP2* and a multi-BLAST alignment showed 77 short homologies (50-190 bp) were observed between *KCNIP2* and *ABCB1*. A single short (21 bp) area of homology was found between *ABCB1* and the vector sequence. Further research is needed to better understand factors contributing to increased integration at these loci.

Overall, the influence of insertions located within genes that are oncogenic or tumorigenic in humans (*MET, CCND1, MIR-1296, MIR-320*) on long-term safety is difficult to predict as there is no cancer reference for the canine genome and limited evidence that these genes are associated with tumorigenesis in dogs. We then examined IS locations relative to cancer genes, lifting over coordinates from the canine to the human genome, and referencing to seven human cancer gene datasets. We found that as the number of cancer genes in a dataset increased, there was an increase in the number of integrations adjacent to these genes. However, when comparing our data to a simulated random dataset, the proportion of IS proximal to the TSS of cancer genes fell within the normal distribution of the random control.

The final question we addressed was whether clonal expansion occurred following AAV vector delivery. A report by Nguyen and colleagues noted FVIII elevations in two dogs four years after rAAV-FVIII treatment.[8] Molecular analyses demonstrated expansion of cells containing IS, close to cancer-associated genes in the animals with rising FVIII levels and in four other dogs. We did not see similar elevations in FVIII activity in this study.[2] Using orthogonal strategies to quantify IS abundance, we demonstrated that in 10 of 16 samples the top-10 most frequent IS represented on average 1.59% of the total IS. In the remaining 6 samples, the top-10 IS represented on average, 6.03% of the total integrations. These results indicate that while IS abundance is variable across different dogs and samples, no obvious clonal dominance was seen in the liver for most samples. Two biopsies exhibited integration frequencies >30% of total IS sequences, a threshold set for clonality in retroviral integration studies. However, this threshold for clonality has not been investigated in solid tissues or non-retroviral studies and its applicability to AAV gene therapy has not been validated.[21] Within these studies, we cannot rule out whether clonal growth occurred due to natural liver growth or changes with age (e.g., nodular hyperplasia). Whether higher clonal abundance occurred in areas of nodular hyperplasia is an active area of investigation and outside the scope of this manuscript. It should be noted that clonal dominance does not directly correlate with adverse events, as shown in the RT-PCR data and lack of tumorigenesis seen in both studies.[2,8]

In summary, our study demonstrates the nature of AAV vector structures a decade after administration is complex. Most intact vectors were episomal, and appear to be responsible for long-term transgene expression. Variable levels of integration site abundance were seen, but with no clear evidence for dominant clonal evolution. These results must be considered in the context of a small study population and limited biopsy material. These results provide further justification for ongoing investigation of AAV vector fate and function, to enhance informed decisions on the long-term safety and efficacy of AAV gene therapy.

Acknowledgements: In Memoriam of Dr. Manfred Schmidt, a brilliant pioneer in the cell and gene therapy field and a truly exceptional colleague. This study was supported in part by a Canadian Institutes for Health Research Foundation Grant FDN 154285, and by a grant from BioMarin Pharmaceutical. The authors thank Dr. Evan Witt for overseeing the upload of genomic data to GEO.

Authorship contribution: D.L., P.B., and S.F. designed the research study; L.H., A.P., and A.W. cared for the animals in the study. P.B., D.H., S.F., and D.L. analyzed the data; P.B., S.F., and D.L. wrote the first draft of the paper and L.L.S. revised the manuscript. M.F., C-R.S., and I.G.-F. conducted and interpreted experiments. S.A., H.T., C.S., L.Y., and C.W. performed the bioinformatic analyses. All authors reviewed and critically edited the manuscript.

Conflict-of-interest disclosure: S.F., C-R.S., C.S., and C.W. are employees and stockholder of BioMarin Pharmaceutical Inc. P.B. has received research support from BioMarin, and consulting fees or honoraria from BioMarin, Octapharma, Novo Nordisk, CSL Behring, Pfizer, and Institute for Nursing and Medication Education (IMNE). D.L. has received research support from BioMarin, CSL-Behring, and Sanofi and has received consulting fees or honoraria from BioMarin, CSL-Behring, Novo Nordisk, Pfizer, and Sanofi. L.L.S. has received consulting fees from BioMarin. H.T., and I.G-F. were employees of ProtaGene CGT GmbH. S.A. and L.Y. are employees of ProtaGene CGT GmbH and M.F. is an employee of ProtaGene US Inc. The remaining authors declare no competing financial interests.

References:

1. Batty P, Lillicrap D. Hemophilia Gene Therapy: Approaching the First Licensed Product. *Hemasphere*. 2021;5(3):e540.

2. Batty P, Mo AM, Hurlbut D, et al. Long-term follow-up of liver-directed, adeno-associated vector-mediated gene therapy in the canine model of hemophilia A. *Blood*. 2022;140(25):2672-2683.

3. Pasi KJ, Laffan M, Rangarajan S, et al. Persistence of haemostatic response following gene therapy with valoctocogene roxaparvovec in severe haemophilia A. *Haemophilia*. 2021;27(6):947-956.

4. Laffan M, Rangarajan S, Lester W, et al. Hemostatic results for up to 6 years following treatment with valoctocogene roxaparvovec, an AAV5-hFVIII-SQ gene therapy for severe hemophilia A. *Res Pract Thromb Haemost*. 2022;6(S1):e12787.

5. Nathwani AC, Reiss UM, Tuddenham EG, et al. Adeno-Associated Mediated Gene Transfer for Hemophilia B: 8 Year Follow up and Impact of Removing "Empty Viral Particles" on Safety and Efficacy of Gene Transfer. *Blood*. 2018;132(S1):92-5856.

6. Donsante A, Miller DG, Li Y, et al. AAV vector integration sites in mouse hepatocellular carcinoma. *Science*. 2007;317(5837):477.

7. Dalwadi DA, Torrens L, Abril-Fornaguera J, et al. Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice. *Mol Ther*. 2021;29(2):680-690.

8. Nguyen GN, Everett JK, Kafle S, et al. A long-term study of AAV gene therapy in dogs with hemophilia A identifies clonal expansions of transduced liver cells. *Nat Biotechnol*. 2021;39(1):47-55.

9. Schmidt M, Foster GR, Coppens M, et al. Liver Safety Case Report from the Phase 3 HOPE-B Gene Therapy Trial in Adults with Hemophilia B. *Res Pract Thromb Haemost*. 2021;5.

10. Schmidt M, Foster GR, Coppens M, et al. Molecular evaluation and vector integration analysis of HCC complicating AAV gene therapy for hemophilia B. *Blood Adv*. 2023;7(17):4966-4969.

11. Hough C, Kamisue S, Cameron C, et al. Aberrant splicing and premature termination of transcription of the FVIII gene as a cause of severe canine hemophilia A: similarities with the intron 22 inversion mutation in human hemophilia. *Thromb Haemost*. 2002;87(4):659-665.

12. Giles AR, Tinlin S, Greenwood R. A canine model of hemophilic (factor VIII:C deficiency) bleeding. *Blood*. 1982;60(3):727-730.

13. Scallan CD, Lillicrap D, Jiang H, et al. Sustained phenotypic correction of canine hemophilia A using an adeno-associated viral vector. *Blood*. 2003;102(6):2031-2037.

14. Fong S, Yates B, Sihn CR, et al. Interindividual variability in transgene mRNA and protein production following adeno-associated virus gene therapy for hemophilia A. *Nat Med*. 2022;28(4):789-797.

15. Oziolor EM, Kumpf SW, Qian J, et al. Comparing molecular and computational approaches for detecting viral integration of AAV gene therapy constructs. *Mol Ther Methods Clin Dev*. 2023;29:395-405.

16. Afzal S, Wilkening S, von Kalle C, Schmidt M, Fronza R. GENE-IS: Time-Efficient and Accurate Analysis of Viral Integration Events in Large-Scale Gene Therapy Data. *Mol Ther Nucleic Acids*. 2017;6:133-139.

17. Sihn CR, Handyside B, Liu S, et al. Molecular analysis of AAV5-hFVIII-SQ vector-genome-processing kinetics in transduced mouse and nonhuman primate livers. *Mol Ther Methods Clin Dev*. 2022;24:142-153.

18. Schnepp BC, Chulay JD, Ye GJ, Flotte TR, Trapnell BC, Johnson PR. Recombinant Adeno-Associated Virus Vector Genomes Take the Form of Long-Lived, Transcriptionally Competent Episomes in Human Muscle. *Hum Gene Ther*. 2016;27(1):32-42.

19. Janovitz T, Sadelain M, Falck-Pedersen E. Adeno-associated virus type 2 preferentially integrates single genome copies with defined breakpoints. *Virol J*. 2014;11:15.

20. Greig JA, Martins KM, Breton C, et al. Integrated vector genomes may contribute to long-term expression in primate liver after AAV administration. *Nat Biotechnol*. 2023.

21. Ott MG, Schmidt M, Schwarzwaelder K, et al. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med*. 2006;12(4):401-409.

22. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*. 2003;302(5644):415-419.

23. Deichmann A, Hacein-Bey-Abina S, Schmidt M, et al. Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J Clin Invest*. 2007;117(8):2225-2232.

24. Hacein-Bey-Abina S, Garrigue A, Wang GP, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest*. 2008;118(9):3132-3142.

25. Cavazzana-Calvo M, Payen E, Negre O, et al. Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature*. 2010;467(7313):318-322.

26. Braun CJ, Boztug K, Paruzynski A, et al. Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity. *Sci Transl Med*. 2014;6(227):227ra233.

27. Nault JC, Mami I, La Bella T, et al. Wild-type AAV Insertions in Hepatocellular Carcinoma Do Not Inform Debate Over Genotoxicity Risk of Vectorized AAV. *Mol Ther*. 2016;24(4):660-661.

28. Chandler RJ, Sands MS, Venditti CP. Recombinant Adeno-Associated Viral Integration and Genotoxicity: Insights from Animal Models. *Hum Gene Ther*. 2017;28(4):314-322.

29. Berry CC, Nobles C, Six E, et al. INSPIIRED: Quantification and Visualization Tools for Analyzing Integration Site Distributions. *Mol Ther Methods Clin Dev*. 2017;4:17-26.

30. Gil-Farina I, Fronza R, Kaeppel C, et al. Recombinant AAV Integration Is Not Associated With Hepatic Genotoxicity in Nonhuman Primates and Patients. *Mol Ther*. 2016;24(6):1100-1105.

31. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between germline and somatic mutation rates in humans and mice. *Nat Commun*. 2017;8:15183.

32. Chandler RJ, LaFave MC, Varshney GK, Burgess SM, Venditti CP. Genotoxicity in Mice Following AAV Gene Delivery: A Safety Concern for Human Gene Therapy? *Mol Ther*. 2016;24(2):198-201.

Figure 1: Regional variation of AAV-cFVIII seen in the liver after long-term follow-up in the hemophilia A dog model, quantified using drop-phase digital droplet polymerase chain reaction (ddPCR). NR = non-responder dogs with plasma FVIII:C (OSA and CSA) levels that remained below the limit of detection but showed evidence of improved whole blood clot times compared with untreated hemophilia A dogs.[2] **A:** AAV-cFVIII vector genome (VG) copies quantified in copies per diploid genome. **B:** cFVIII mRNA expression normalized to Beta-2-Microglobulin (B2M).

Figure 2: Episomal AAV-cFVIII vector forms detected in the liver after long-term follow up, with correlation between vector copies and FVIII expression. A: Circular full-length AAV-cFVIII vector genomes, detected by drop-phase droplet digital PCR (ddPCR) following treatment with Plasmid Safe™ ATP-Dependent DNase (PsDNase) and KpnI to enrich circular genomes and quantify full-length monomers. **B.** Strong correlation (r=0.832, p=0.0201) between circular full-length vg (vector genome) copies and terminal FVIII:C levels measured by chromogenic substrate assay. For ELI, where terminal FVIII:C was unavailable, a mean of earlier FVIII:C on-study was used. **C:** Strong correlation (r=0.83, p<0.001) between full-length episomal vg copies and cFVIII mRNA expression, supporting episomal AAV-cFVIII as the primary source of FVIII expression. **D:** Southern Blotting, following PsDNase and BamHI, demonstrates full-length AAV-cFVIII vector genomes in an H-T configuration (band size 4kb). Sample from JU contains two fragments (~1600 and 850bp) suggesting AAV-cFVIII truncation. **E:** Target enrichment sequencing (TES) comparing frequencies of vector-vector, possibly episomal (blue) to vector-canine genome, integrated (orange) sequencing reads demonstrating the majority resulted from vector-vector (episomal) reads.

Figure 3: Targeted enrichment sequencing (TES) evaluation of integration sites. A: Location of unique integration sites relative to annotated canine genes (canFam3) detected using TES. The majority (93.8%) of unique integration sites (IS) were located in intragenic regions. For in-gene insertions, 206/355 (58%) fell within an intron and 149/355 (42%) in an exon. NR = non-responder. Upstream insertions are defined as intergenic regions 5' of the closest gene, and downstream

insertions are 3' of the closest gene. In-gene insertions are defined as those located within the transcriptional unit. **B-D:** Cumulative sequence counts of the top-10 integration sites were identified. **B:** Most (10/16) samples exhibited integration profiles where the top IS demonstrated frequencies <10% of the total sequences. Representative examples include VC-2 and FLO-1. **C:** 4/16 samples had at least 1 IS with a frequency of >10% and <30% of the total IS sequence count. Representative examples include JU-1 and MZ-1. **D:** 2/16 samples had at least 1 IS with a frequency of >30%. Both of these samples were obtained from non-responding dogs (ANG-2 and MG-1). See Supplementary Figure 4 for top-10 IS data for all 16 samples analyzed.

Figure 4: Features and cellular implication sites of hepatic AAV-cFVIII integration after long-term follow up in the hemophilia A dog model. A: Heat map of integration sites in 10MB window sizes, demonstrating integration across the canine genome detected by target enrichment sequencing; frequencies displayed on a density scale ranging from 0 IS/10Mb to >697 IS/10Mb, with highest frequency IS, shown in yellow & red. **B:** Representation of common integration sites (CIS) seen within 10MB window sizes on chromosomes 14 (*ABCB1*), 28 (*KCNIP2*) & X (*F8* and *CLIC2*). **C-E:** No significant dysregulation in gene expression of genes in proximity to common integration sites in the liver of AAV-cFVIII treated hemophilia A dogs, compared to untreated hemophilia A or normal dogs; quantification by drop-phase digital droplet (ddPCR) and normalized to beta-2-microglobulin expression. N-dog = normal dog. NTC= no template control. **F:** No dysregulation in MET expression in the liver of AAV-cFVIII treated hemophilia A dogs, compared to untreated hemophilia A or normal dogs; quantification by ddPCR and normalized to beta-2-microglobulin expression. **G:** Comparison of IS frequencies in proximity (≤100kb) to cancer genes to simulated data sets (n=10,000), demonstrated this frequency is within the predicted normal range (0.065 - 0.084). **H:** Evaluation of association ($p < 0.05$) of integration sites with areas of chromatin accessibility (ATAC-Seq) and methylation status (Bisulfite sequencing).

Figure 5: Evaluation of Vector Integrity by Long Read Sequencing. Long range sequencing was performed for 8 IS with a higher (>10%) observed frequency. **A:** Suggested structures of integrated AAV-cFVIII vector for selected IS. Structure 1, included the 5' ITR and part of the promoter (identified in ANG-2 *MIR320*, MZ-1 *MET*, ANG-1 *MIR1296* and ANG-2 *MET*). Structure 2, included the 5' ITR, part of the promoter, parts of the poly-A and 3' ITR (identified in MG-1 *CCND1*, MZ-1 *CCND1* and JU-1 *MIR1296*). **B:** Vector coverage for samples MZ-1 *MET* Sequencing data aligned to putative integration site reference, including the AAV_TTR_FVIII vector at the IS location retrieved by TES. Scale is set to 0-70664 reads. **C:** Vector coverage for sample MG-1 CCND1. Sequencing data aligned to putative integration site reference, including the AAV-cFVIII vector at the IS location retrieved by TES. Scale is set to 0-120555 reads. **D:** Coverage for all sorted reads obtained from the AAV-cFVIII amplicon. Sequencing data aligned to the AAV-cFVIII vector sequence. Scale is set to 0-3131072 reads. For other examples see Supplementary Figure 10.

						TES				LAMPCR		
ID	AAV Dose vg/kg	Treatment Age (years)	Follow Up (years)	CSA FVIII:C (%)	N	TES Seq	Total IS Seq	TES Unique IS	Int VCN vg/cell	LAMPCR Seq	LAMPCR Unique IS	LAMPCR IS/cell
ALX	AAV6 1.0e13	0.7	10.5	8.6%	1 2	2324084 3525725	293 688	239 434	6.18e-4 1.12e-3	87797 n/a	360 n/a	6.20e-4 n/a
ANG*	AAV2 1.5e13	1.0	11.5	0.5%	1 2	2264496 3016959	1363 1377	310 58	8.01e-4 1.50e-4	284365 n/a	167 n/a	2.88e-4 n/a
ELI	AAV2 6.0e12	0.6	8.2	#	1 2	2915252 2406745	385 656	269 496	6.95e-4 1.28e-3	257439 n/a	213 n/a	3.67e-4 n/a
FLO	AAV8 1.0e13	1.0	10.5	1.8%	1 2	3057875 3237835	838 887	362 413	9.35e-4 1.07e-3	n/a n/a	n/a n/a	n/a n/a
JUN	AAV2 2.7e13	1.0	12.0	7.2%	1 2	2286381 2494473	948 700	550 534	1.42e-3 1.38e-3	160937 n/a	380 n/a	6.55e-4 n/a
MG*	AAV6 1.7e13	0.5	11.5	0.3%	1 2	2597851 2525320	436 152	49 105	1.27e-4 2.71e-4	146726 n/a	67 n/a	1.15e-4 n/a
MZ	AAV6 1.0e13	1.3	10.1	7.9%	1 2	2369218 2355983	650 458	243 353	6.28e-4 9.12e-4	n/a n/a	n/a n/a	n/a n/a
VC	AAV2 1.5e13	0.7	11.0	2.9%	1 2	3855884 2012102	1349 353	1055 276	2.73e-3 7.13e-4	n/a 228238	n/a 196	n/a 3.38e-4
TB	Control	n/a	n/a	n/a	1	2805661	4	4	1.03e-5	n/a	n/a	n/a

Table 1: Quantification of AAV-cFVIII vectors in the liver following long-term follow up in the hemophilia A dog model. TES = target enrichment sequencing. LAM-PCR = linear-amplification mediated PCR. vg = vector genomes. *non-responder. Control = untreated hemophilia A control dog. CSA=Chromogenic substrate FVIII activity assay, using a pooled normal canine standard (12-20 normal dogs). # = terminal sample not available: earlier CSA FVIII:C on-study: 3.4 – 5.7%. N=Liver sample number (i.e. ALX sample 1 = ALX-1). Seq = sequencing reads. IS = integration sites. Int VCN=integrated vector copy number. Epi VCN = episomal vector copy number. n/a = not available.

A	Rank	Events	Chr	Mean Integration Locus	Dimension (bp)	Gene	Entropy
TES	1*	740	28	14066254	30324	*KCNIP2*	0.90
	2*	521	X	123771584	25052	*CLIC2*	0.94
	3*	373	14	16288512	37578	*ABCB1*	0.94
	4*	361	X	122970340	146353	*F8*	0.90
	5	69	5	75600480	17415	*LOC479649*	0.88
	6	30	13	62161330	16400	*ALB*	0.90
	7	26	4	14746173	72601	*MIR1296*	0.93
	8	25	1	121363006	948	*MIR578*	0.97
	9	18	13	32788006	754	*MIR30D*	0.99
	10	15	6	43486296	3975	*MIR8880*	0.67
B	**Rank**	**Events**	**Chr**	**Mean Integration Locus**	**Dimension (bp)**	**Gene**	**Entropy**
LAMPCR	1*	45	28	14062005	11015	*KCNIP2*	0.87
	2*	29	X	123771226	16893	*CLIC2*	0.93
	3*	11	14	16281704	7201	*ABCB1*	0.99
	4*	10	X	122948039	80519	*F8*	0.88

Table 2: Location of the top 10 common integration sites (CIS) detected by target enrichment sequencing (TES) (A) and linear-amplification mediated PCR (LAMPCR) (B). *CIS seen in both sequencing methodologies. Chr = chromosome. bp = base pairs. Entropy = Contribution of CIS within individual or different samples, this ranges from 0 to 1, reflecting homogeneous (i.e., low values - found in a single sample) to heterogeneous (i.e., high values - found in several samples) CIS respectively. High entropy values within this study signify the heterogeneous nature of the CIS.

ID	Sample	Total IS	F8 Intron IS	F8 Exon IS	True F8 IS (%)	Undetermined F8 IS (%)
ANG	1	4	2	2	50	50.0
ANG	2	113	0	113	0	100.0
ELI	1	11	9	2	81.8	18.2
ELI	2	13	1	12	7.7	92.3
VEC	1	63	22	41	34.9	65.1
VEC	2	12	5	7	41.7	58.3
ALX	1	5	1	4	20.0	80.0
ALX	2	10	2	8	20.0	80.0
JU	1	18	6	12	33.3	66.6
JU	2	42	9	33	21.4	78.6
MZ	1	5	2	3	40.0	60.0
MZ	2	12	6	6	50.0	50.0
FLO	1	13	9	4	69.2	30.8
FLO	2	26	21	5	80.8	19.2
MG	1	5	2	3	40.0	60.0
MG	2	0	0	0	0.0	0.0

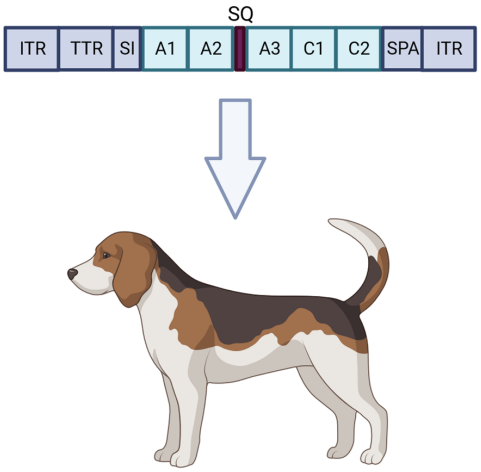
Table 3: In-gene integration sites (IS) mapped to the canine *F8* gene using targeted enrichment sequencing (TES). Integration sites mapped within a *F8* intron were deemed as "true" IS as these regions are not represented in the AAV-cFVIII vector. IS mapped to a *F8* exon were considered "undetermined", as these could represent either true exonic insertions or vector-vector insertion events. Overall, 97/352 (27.6%) insertion sites were located in *F8* introns.

	UNC Integration Clusters					Queen's TES Common Integration Site			
Chr	Hits	Start	End	Width	Gene*	CIS# (Events)	Mean Locus	Width	Gene*
4	7	14530145	14924592	39448	EGR2	7 (26)	14746173	72601	MIR1296
4	26	39589684	39589877	194	DUSP1	nr	nr	nr	nr
13	5	62155847	62632404	476558	ALB	6 (30)	62161330	16400	ALB
18	14	48484014	48821243	337230	CCND1	26 (7)	48641914	60162	CCND1
25	15	34458017	34601125	143109	EGR3	12 (14)	34589168	7314	MIR320

Table 4: Overlapping integration clusters in two canine AAV-cFVIII studies after long-term follow up. Comparison of common integration sites (CIS) occurring in this study to those reported by Nguyen et al.[8] nr = not represented in Target Enrichment Sequencing (TES) CIS dataset. Hits = number of calls against 50 random sites. UNC = University of North Carolina. Chr = Chromosome. * Differences in gene annotations are noted in these two analyses.

Vector Integration and Fate in the Hemophilia Dog Liver Multi-Years Following AAV-FVIII Gene Transfer

AAV-canine FVIII portal vein infusion



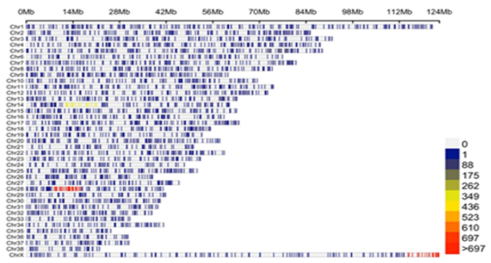
N = 8 severe hemophilia A dogs

Dose: 6.0e12 – 2.7e13 vg/kg

Follow-up: 8.2 – 12.0 years

1. The predominant vector form was non-integrated episomal structures with levels correlating with long-term transgene expression

2. Vector integration was seen across the canine genome by targeted enrichment next-generation sequencing (TES) and linear amplification-mediated (LAM)-PCR



Conclusions: 1. AAV-cFVIII vectors predominantly persist in the liver long-term in non-integrated episomal forms. 2. AAV vector integration was seen at low frequencies and occurred commonly in areas of open chromatin with no effect on gene expression.

Batty et al. DOI

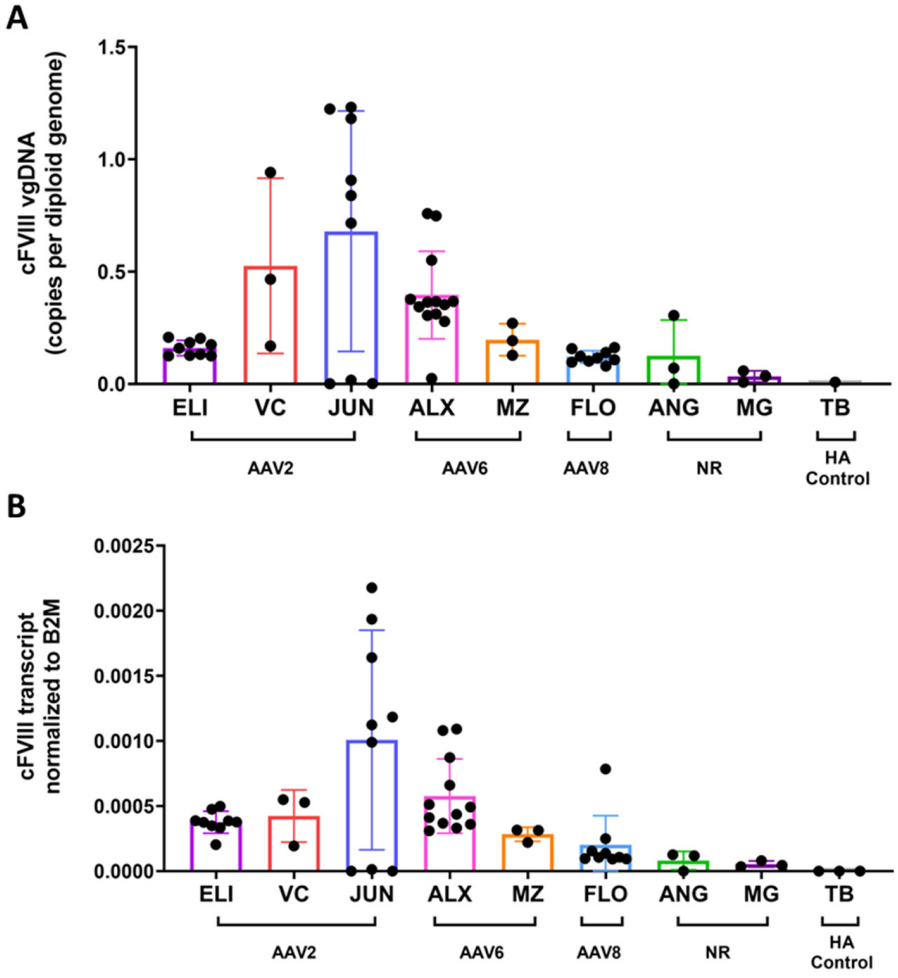

A
cFVIII vgDNA
(copies per diploid genome)
1.5
1.0
0.5
0.0
ELI
VC
JUN
ALX
MZ
FLO
ANG
MG
TB
AAV2
AAV6
AAV8
NR
HA
Control
B
cFVIII transcript
normalized to B2M
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
ELI
VC
JUN
ALX
MZ
FLO
ANG
MG
TB
AAV2
AAV6
AAV8
NR
HA
Control

Figure 1

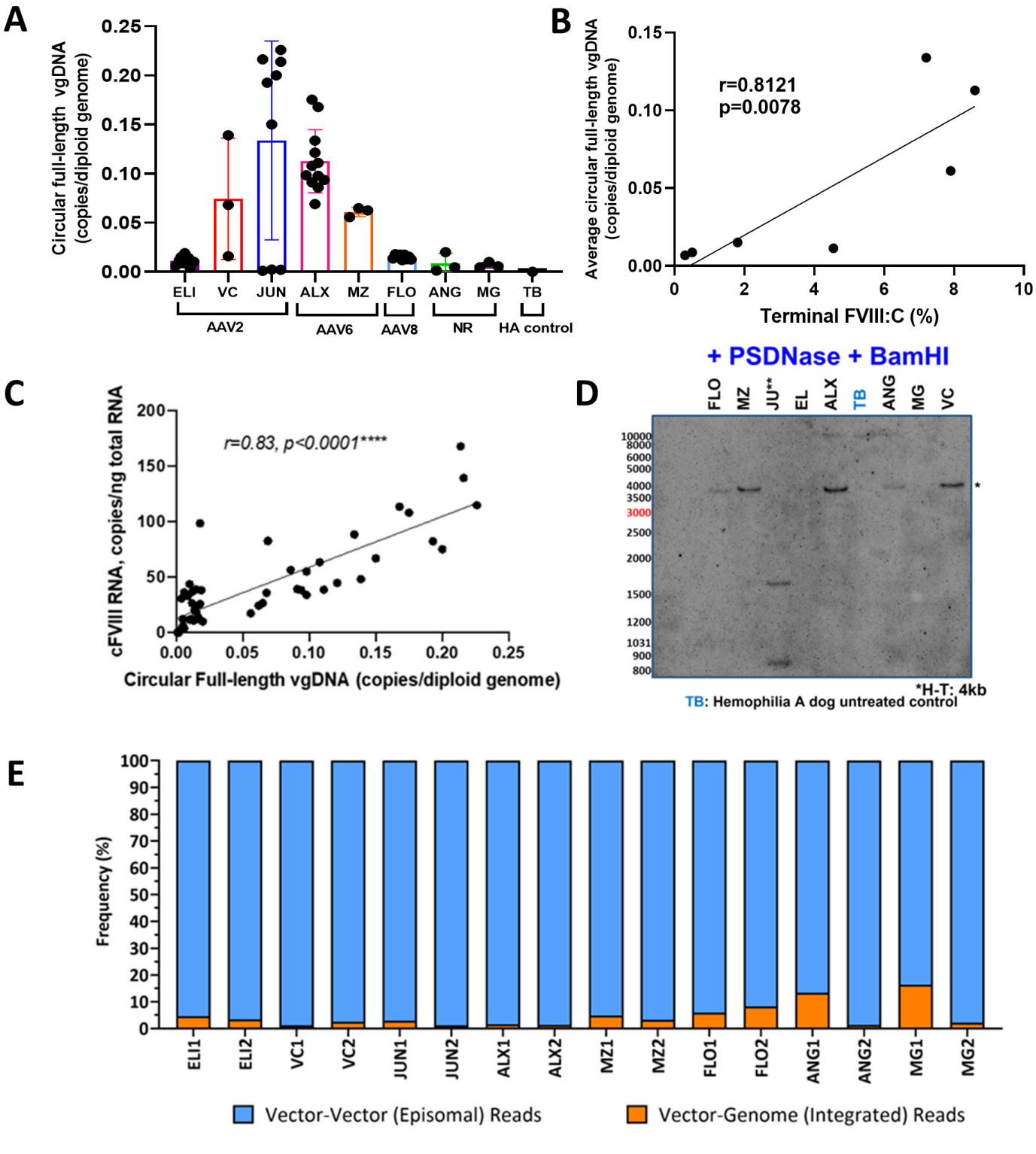
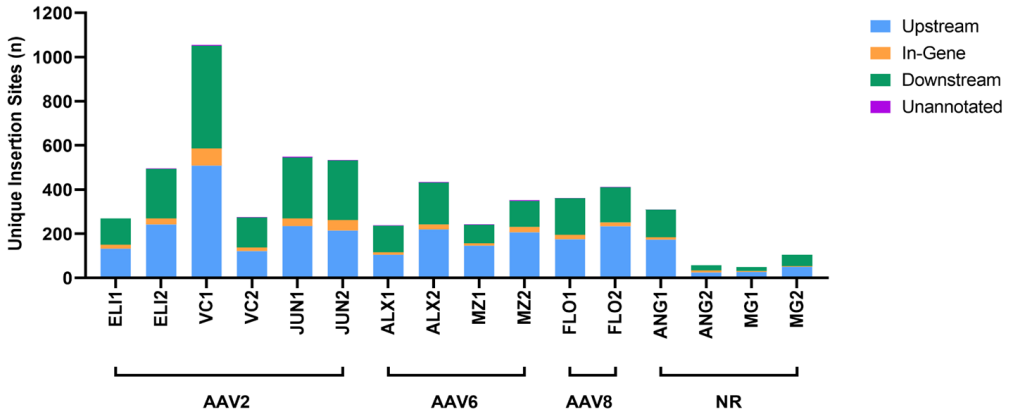

A
Circular full-length vgDNA (copies/diploid genome)
0.25
0.20
0.15
0.10
0.05
0.00
ELI
VC
JUN
ALX
MZ
FLO
ANG
MG
TB
AAV2
AAV6
AAV8
NR
HA control
B
Average circular full-length vgDNA (copies/diploid genome)
r=0.8121
p=0.0078
Terminal FVIII:C (%)
C
cFVIII RNA, copies/ng total RNA
r=0.83, p<0.0001****
Circular Full-length vgDNA (copies/diploid genome)
D
+ PSDNase + BamHI
FLO
MZ
JU**
EL
ALX
TB
ANG
MG
VC
10000
8000
6000
5000
4000
3500
3000
2500
2000
1500
1200
1031
900
800
*H-T: 4kb
TB: Hemophilia A dog untreated control
E
Frequency (%)
ELI1
ELI2
VC1
VC2
JUN1
JUN2
ALX1
ALX2
MZ1
MZ2
FLO1
FLO2
ANG1
ANG2
MG1
MG2
Vector-Vector (Episomal) Reads
Vector-Genome (Integrated) Reads

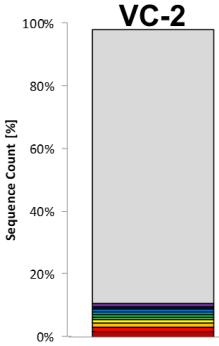
Figure 2

A



B

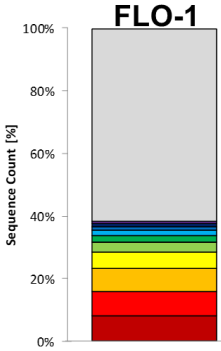
VC-2



Rank	Gene Name	Seq. Count	Location
Top1	CXCL12	6	28+3022679
Top2	KCNIP2	5	28+14061905
Top3	NCSTN	5	38+21935336
Top4	ABCB1	4	14+16285602
Top5	ABCB1	3	14+16281042
Top6	PLEKHA8	3	14+42478522
Top7	CFTR	3	14+60743670
Top8	cOR13M4	3	16+5504630
Top9	KCNIP2	3	28+14067815
Top10	MIR16-2	3	34+26789181
#All Other mapp. IS	0	315	

Seq Count 10 Strongest	38
Seq Count all other mapp. IS	315
Total Seq Counts Used	353

FLO-1

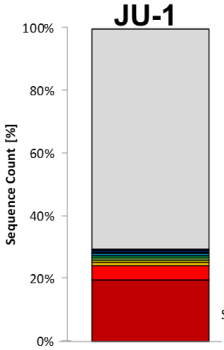


Rank	Gene Name	Seq. Count	Location
Top1	MIR1296	68	4+14710285
Top2	LOC478001	66	2+23981744
Top3	LOC479649	61	5+75605276
Top4	LOC478001	43	2+23981746
Top5	KCNIP2	27	28+14061114
Top6	ABCB1	19	14+16284253
Top7	KCNIP2	14	28+14058969
Top8	ABCB1	9	14+16284237
Top9	CTSC	8	21+11968362
Top10	KCNIP2	6	28+14061044
#All Other mapp. IS	352	517	

Seq Count 10 Strongest	321
Seq Count all other mapp. IS	517
Total Seq Counts Used	838

C

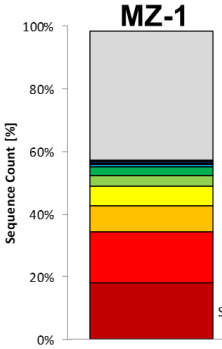
JU-1



Rank	Gene Name	Seq. Count	Location
Top1	MIR1296	188	4+14742806
Top2	MIR1296	44	4+14476720
Top3	MIR1296	9	4+14476737
Top4	SLC23A1	8	2+32892924
Top5	GJB2	6	25+17985399
Top6	MIR1296	6	4+14720327
Top7	MIR8858	5	18+51762112
Top8	SLC12A4	5	5+81616667
Top9	CYB5R1	5	7+581283
Top10	COR13C13	4	11+61199922
#All Other mapp. IS	540	668	

Seq Count 10 Strongest	280
Seq Count all other mapp. IS	668
Total Seq Counts Used	948

MZ-1

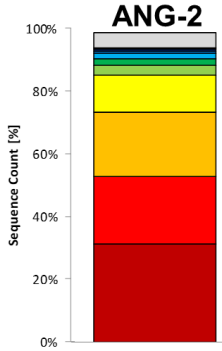


Rank	Gene Name	Seq. Count	Location
Top1	MET	119	14+55615149
Top2	MET	109	14+55614962
Top3	CCND1	54	18+48611833
Top4	CCND1	42	18+48611849
Top5	RPL15	21	23+16465615
Top6	RPL15	20	23+16465613
Top7	FGFR2	4	28+32082932
Top8	CLIC2	3	X+123773810
Top9	ABCB1	3	14+16274913
Top10	ABCB1	3	14+16276726
#All Other mapp. IS	233	272	

Seq Count 10 Strongest	378
Seq Count all other mapp. IS	272
Total Seq Counts Used	650

D

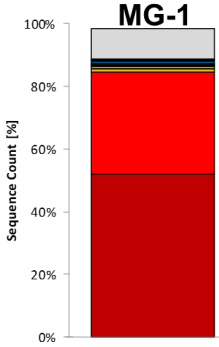
ANG-2



Rank	Gene Name	Seq. Count	Location
Top1	MIR320	437	25+34592658
Top2	MIR320	300	25+34589409
Top3	MIR320	289	25+34589406
Top4	MET	163	14+55620574
Top5	F8	45	X+122990517
Top6	F8	27	X+123043093
Top7	F8	24	X+123000564
Top8	F8	12	X+122986004
Top9	MIR1296	9	4+14713493
Top10	MET	6	14+55620515
#All Other mapp. IS	18	65	

Seq Count 10 Strongest	1,312
Seq Count all other mapp. IS	65
Total Seq Counts Used	1,377

MG-1



Rank	Gene Name	Seq. Count	Location
Top1	CCND1	231	18+48671995
Top2	CCND1	144	18+48671926
Top3	CLGN	4	19+2586882
Top4	CLIC2	3	X+123774454
Top5	IL15	2	19+421885
Top6	BRCA2	2	25+3928864
Top7	NEFH	2	26+22784680
Top8	KCNIP2	2	28+14056188
Top9	CRYAA	2	31+37875327
Top10	CCL28	2	4+66356565
#All Other mapp. IS	39	42	

Seq Count 10 Strongest	394
Seq Count all other mapp. IS	42
Total Seq Counts Used	436

Figure 3

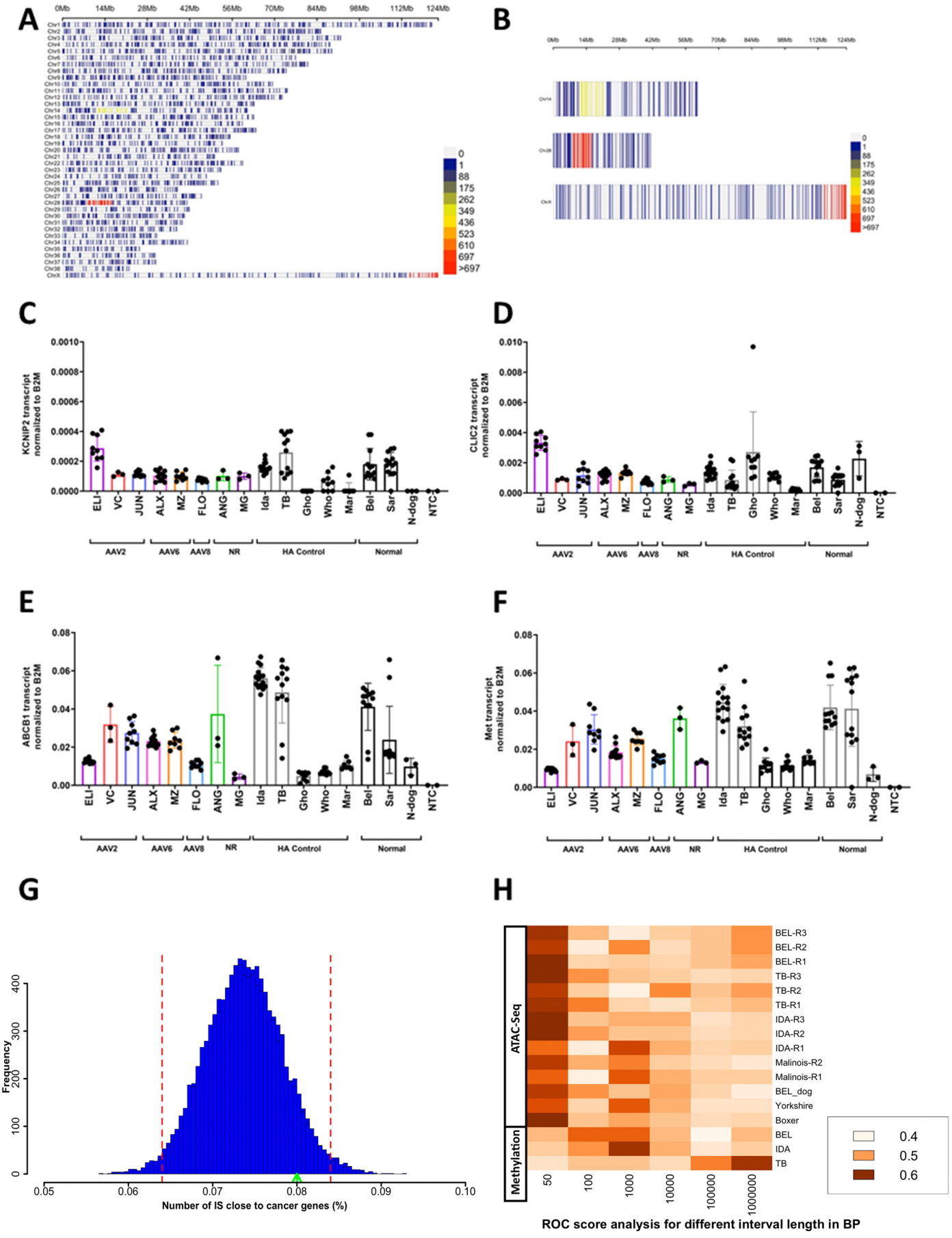

A
B
C
D
E
F
G
H
KCNIP2 transcript normalized to B2M
CLIC2 transcript normalized to B2M
ABCB1 transcript normalized to B2M
Met transcript normalized to B2M
ELI VC JUN ALX MZ FLO ANG MG Ida TB Gho Who Mar Bel Sar N-dog NTC
AAV2 AAV6 AAV8 NR HA Control Normal
Frequency
Number of IS close to cancer genes (%)
ATAC-Seq
Methylation
BEL-R3
BEL-R2
BEL-R1
TB-R3
TB-R2
TB-R1
IDA-R3
IDA-R2
IDA-R1
Malinois-R2
Malinois-R1
BEL_dog
Yorkshire
Boxer
BEL
IDA
TB
ROC score analysis for different interval length in BP

Figure 4

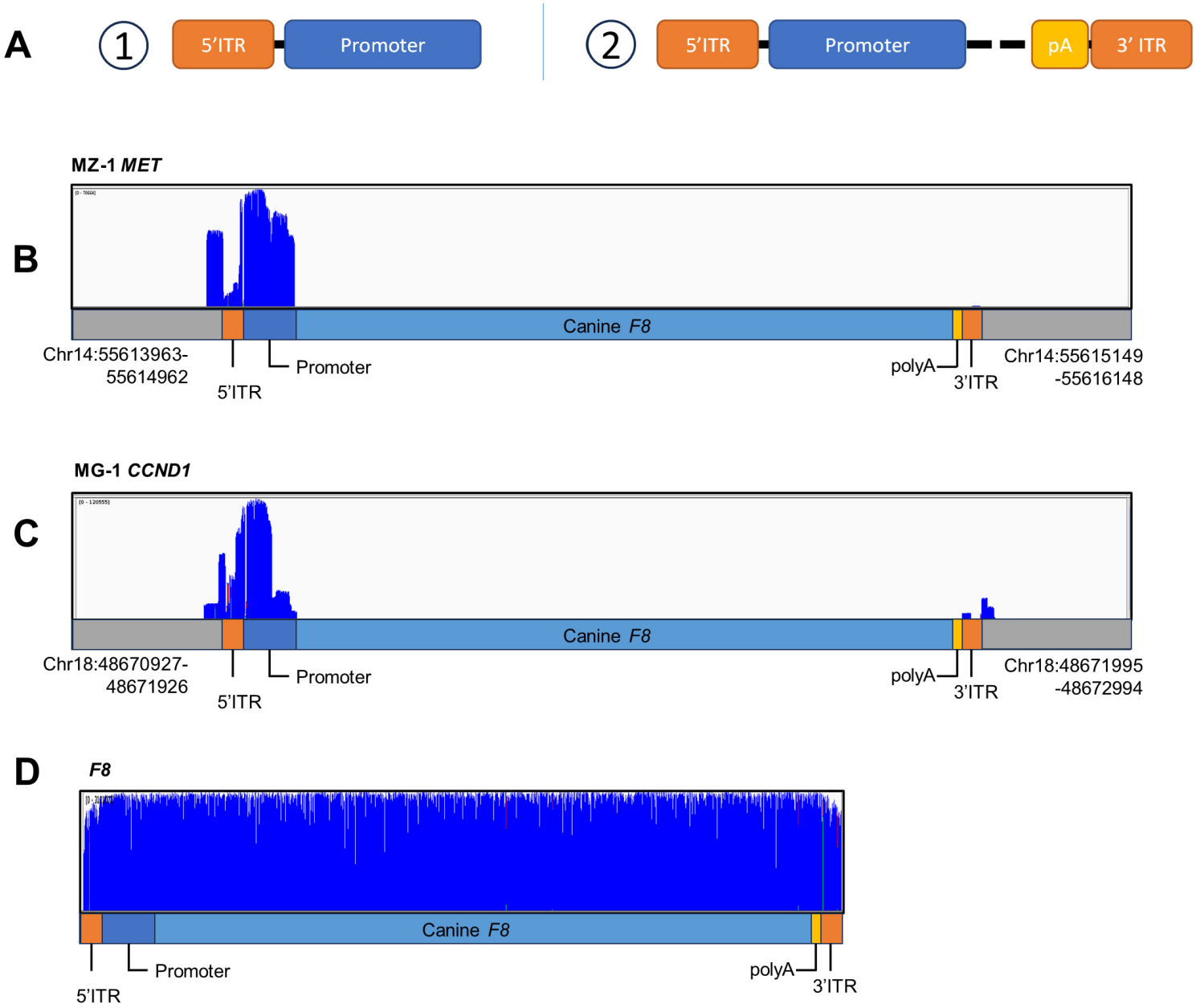

A
1
5'ITR
Promoter
2
5'ITR
Promoter
pA
3' ITR
B
MZ-1 MET
Canine F8
Chr14:55613963-
55614962
5'ITR
Promoter
polyA
3'ITR
Chr14:55615149
-55616148
C
MG-1 CCND1
Canine F8
Chr18:48670927-
48671926
5'ITR
Promoter
polyA
3'ITR
Chr18:48671995
-48672994
D
F8
Canine F8
5'ITR
Promoter
polyA
3'ITR

Figure 5