Published in final edited form as:

Virology. 2017 November ; 511: 40–48. doi:10.1016/j.virol.2017.08.011.

Parvovirus B19 Integration into Human CD36+ Erythroid Progenitor Cells

Tyler Janovitz[a,b,#], **Susan Wong**[c], **Neal S. Young**[c], **Thiago Oliveira**[d], and **Erik Falck-Pedersen**[b,*]

[a]Tri-Institutional MD-PhD Program, Weill Medical College of Cornell University, New York, NY 10065 USA

[b]Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10065 USA

[c]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD

[d]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA

Abstract

The pathogenic autonomous human parvovirus B19 (B19V) productively infects erythroid progenitor cells (EPCs). Functional similarities between B19V nonstructural protein (NS1), a DNA binding endonuclease, and the Rep proteins of Adeno-Associated Virus (AAV) led us to hypothesize that NS1 may facilitate targeted nicking of the human genome and B19 vDNA integration. We adapted an integration capture sequencing protocol (IC-Seq) to screen B19V infected human CD36+ EPCs for viral integrants, and discovered 40,000 unique B19V integration events distributed throughout the human genome. Computational analysis of integration patterns revealed strong correlations with gene intronic regions, H3K9me3 sites, and the identification of 41 base pair consensus sequence with an octanucleotide core motif. The octanucleotide core has homology to a single region of B19V, adjacent to the P6 promoter TATA box. We present the first direct evidence that B19V infection of erythroid progenitor cells disrupts the human genome and facilitates viral DNA integration.

Keywords

Parvovirus; B19; latency; integration; high throughput sequencing; human erythroid progenitor cell

*Correspondence: efalckp@med.cornell.edu (E.F.P.).

#Present Address: Department of Pathology, Brigham and Women's Hospital, Boston, Ma.

Introduction

Parvovirus B19 is currently the only well established pathogenic human parvovirus. By adulthood approximately 60% of the population is seropositive for B19V (Nabae et al., 2014). In the majority of individuals, seroconversion occurs after asymptomatic infection. Symptomatic infections may result in several diseases including fifth disease (erythema infectiosum) a common childhood infection; transient aplastic crisis; hydrops fetalis; or polyarthropathy (reviewed in (Young and Brown, 2004)). In symptomatic infections, B19V can reach titers of $>10^{12}$ ge/ml at the height of viremia (Young and Brown, 2004). In one study, patients presenting with acute phase erythema infectiosum were determined to have a median average viral serum titer of 7.6×10^{5} genome equivalents (ge)/ml, whereas patients presenting with aplastic crisis had titers as high as 10^{12} ge/ml (Ishikawa et al., 2014).

Resolution of acute virus infection occurs through self-limiting cytotoxicity (Kurtzman et al., 1987; Ozawa et al., 1987), innate inflammatory/immune mechanisms (Kerr et al., 2001), production of high titer B19V neutralizing antibody(Kurtzman et al., 1989), and B19V specific cytotoxic T cells(Streitz et al., 2008). Complete viral clearance from the circulatory system may require several months to years (Dobec et al., 2007; Lindblom et al., 2005). Several studies have identified the existence of persistent viral genomes in a variety of tissue types including bone marrow (reviewed in (Servant-Delmas et al., 2010)). The viral and host determinants that establish the clinical course of viral infection and the potential for viral persistence are not completely understood (Kerr, 2005).

B19V production appears to occur predominantly in bone marrow erythroid progenitor cells (EPCs). Not all erythropoietic cells are equally permissive for viral replication (Bua et al., 2016; Takahashi et al., 1990). B19V infects EPCs by binding P antigen (glycosphingolipid globoside), the primary B19V receptor (Brown et al., 1993). Studies have suggested α5β1 integrin and KU80 as secondary receptors that facilitate virus internalization (Munakata et al., 2005; Weigel-Kelley et al., 2001, 2003). Ex-vivo differentiated CD36+ erythroid progenitor cells are permissive for B19V gene expression, replication, and production of infectious virus (Wong et al., 2008). However, limited B19V yield has hampered establishment of an in vitro B19V production system. Therefore studies characterizing B19V infection most often depend on use of infectious human serum. Nonerythroid cell types can be infected by B19V, but are nonpermissive or semi-permissive due to defective transcription, splicing, and translational functions which compromise replication and virus production (reviewed in (Adamson-Small et al., 2014)).

Replication of the B19V 5.5 KB single stranded DNA genome utilizes the basic mechanism established for other members of the parvovirus family (Chen et al., 2011; Ganaie et al., 2017; Guan et al., 2009; Luo and Qiu, 2015). The viral inverted terminal-repeat elements present at both ends of the linear genome (Figure 1D) serve as replication primers for cellular DNA polymerase. The nonstructural protein NS1 contributes to terminal end resolution through binding NS1 DNA binding elements (NSBE) in the origin region, and nicking viral DNA at the terminal resolution site (trs) of the minimal origin (Guan et al., 2009; Tewary et al., 2014). NS1 has functional domains similar to those found in the well-characterized AAV2 large Rep protein (Im and Muzyczka, 1990; Maggin et al., 2012;

Zarate-Perez et al., 2012; Zhou et al., 1999)(Figure 1A). The N-terminal domain includes both DNA binding and endonuclease activities(Sanchez et al., 2016), a linker region bridges to a putative SF3 helicase/ATPase domain and the weakest region of homology is found in the C-terminal putative Zinc binding region. NS1 expression influences gene expression patterns, induces cell cycle arrest (Lou et al., 2012; Luo et al., 2013; Wan et al., 2010) and contributes to apoptotic cell death (Moffatt et al., 1998; Sol et al., 1999). Overexpression of NS1 is cytotoxic and contributes to the cytopathology associated with aplastic crisis.

NS1 expression may lead to nicking of host chromosomal DNA, presenting an opportunity for DNA damage and B19V DNA integration into the host genome. To test this possibility, we have utilized B19V infection of permissive CD36+ erythroid progenitor cells as an experimental model. Using a modified high throughput genome wide integration sequencing strategy (IC-Seq) we have identified over 4×10^4 unique B19V integration events following infection of CD36+ erythroid progenitor cells. Computational analysis of integration sites revealed a novel pattern of integration correlates for this autonomous parvovirus, and identified a consensus chromosomal sequence domain that closely mirrors elements present in the B19V P6 promoter.

Materials and Methods

Ex-vivo generated CD36$^+$ erythroid progenitor cells

Primary human CD34$^+$ peripheral blood stem cells (termed A31) were cultured in serum-free expansion medium as previously described (Wong et al., 2008). Briefly, 10^4 cells/ml were placed in expansion medium containing a 1:5 dilution of BIT 9500 (STEMCELL Technologies, Vancouver, British Columbia, Canada) in Alpha minimum essential medium (Mediatech, Herndon, VA) with added supplements. After 4 days of culture in the expansion medium, 1 volume of cell culture was expanded into 4 volumes with fresh medium, and thereafter maintained at less than 2×10^6 cells/ml. Differentiating cells were characterized by the presence of cell surface antigens over a course of two weeks. Cell marker analysis by flow cytometry using fluorescein isothiocyanate (FITC) conjugated antibodies to CD34 and CD36 (BD Biosciences, Franklin Lakes, NJ) established the maturation stage of erythroid progenitor cell populations. Day 9 CD36+ A31 cells were aliquoted and stored as frozen stock cell lines. For B19V integration studies, Day 9 CD36 A31 cells were thawed, allowed to expand for 3 days, and infected with virus as indicated.

Virus and infections

High-titer B19V viremic plasma (4×10^{11} viral particles per milliliter) seronegative for B19V IgG and IgM was purchased from Seracare Life Sciences (Milford, Ma.). Serial dilutions were used to determine the virus dilution (10^{-3}) that provided 5–10% of cells positive for B19V capsid protein expression (by immunofluorescence (IF) analysis), with limited cytopathic effect at 24 hours post inoculation. Using parameters established by analytical screening, experimental replicates were carried out by infecting approximately 1.5×10^7 CD36+ Day 12 EPC with 2×10^3 genome equivalents (ge)/cell.

Immunofluorescence for B19V capsid proteins

As previously described (Wong et al., 2008), cells were harvested and cytocentrifuged at 1,500 rpm, fixed in acetone and methanol (1:1), washed in phosphate-buffered saline, and incubated with monoclonal antibody 521-5D (1:500 dilution, Chemicon, Temecula, CA). For IF staining, FITC-labeled goat anti-mouse immunoglobulin G (1:200, BD Biosciences) was used as a secondary antibody and counterstained with Evans blue (1:200).

Real-time RT-PCR and PCR

As previously described (Wong et al., 2008), real-time reverse transcription-PCR (RT-PCR) for B19V NS and VP1 capsid transcripts and real-time PCR for B19V viral DNA were carried out using the PerfeCTa Multiplex qPCR Supermix (Quanta Biosciences Inc., West Chester, PA). For RT-PCR, RNA was extracted from the cells by using the TurboCapture ®96 mRNA kit (Qiagen, Valencia, CA) and converted to cDNA by using random primers with Moloney murine leukemia virus reverse transcriptase (Invitrogen, Carlsbad, CA). Real-time RT-PCR was performed using 5 ul of the resulting cDNA, which was amplified as a multiplex with β-actin as an internal control. For real-time PCR, DNA was extracted using the QIAamp DNA blood mini kit (Qiagen) and 5 ul of the resulting DNA was used for analysis. The same primer and probe sets were used for both cDNA and DNA for real time PCR analysis. Primers and probes: B19 **Capsid** F-CCTGGGCAAGTTAGCGTAC, R-ATGAATCCTTGCAGCACTGTCA, Probe 56FAM-TATGTTGGGCCTGGCAA-3IABlk_FQ, **NS1** F-GGGCAGCATGTGTTAAAGTGGA, R-TGGCCATTGCCAAGTTTGT, NS1 probe 5TYE665-TTATGGGCCGCCAAGTACAGGAAA-31AbRQsp, **Actin** F-GGCACCCAGCACAATGAAG, R-GCCGATCCACACGGAGTACT, actin probe 5MAX550-TCAAGATCATTGCTCCTCCTGAGCGC-3IABlk_FQ I (IDTDNA, Coralville, Iowa).

IC-Seq Library Preparation

For each replicate, three aliquots of 6 million CD36+EPCs, containing ~250 ug of genomic DNA in total were processed using the TC-Seq and IC-Seq protocols as previously described (Janovitz et al., 2013; Janovitz et al., 2014; Klein et al., 2011; Oliveira et al., 2012) (illustrated in Figure 1E). Briefly, aliquots were lysed in Proteinase K buffer (100 mM Tris (pH 8), 0.2% SDS, 200 mM NaCl, 5 mM EDTA) with 200 ug/mL Proteinase K. Genomic DNA was purified by phenol-chloroform extraction and ethanol precipitation. Sonication was conducted using a Bioruptor (Diagenode, Denville, N.J.) to generate DNA smears of 600–1200bp, with an 850bp core. DNA was polished using the End-It DNA Repair Kit (Epicenter), purified, and dA-tailed utilizing the 3′–5′ exo–Klenow Fragment (NEB). Then, fragments were ligated to 200pmol of annealed linkers.

Viral Junction Amplification

All PCRs were conducted using Herculase II Fusion DNA Polymerase (Agilent Technologies) according to manufacturer specifications. Sequences of the pLinker primer and asymmetric linker oligos were described previously (Janovitz et al., 2013; Janovitz et al., 2014; Klein et al., 2011). The B19V primer sequences utilized were: P1: 5’-[Bio-

TEG]CAGAGCAAAGAATGGCGGCGG-3’ and P2: 5’-GCGGGCTCGTTCACTCGGTC-3’ (approximate location illustrated in Figure 1D). Pooled, linker-ligated DNA was divided into 800ng aliquots and subjected to linear-amplification (single-primer) PCR with biotinylated P1: 1×(98°C-3min) 12×(98°C-40s, 65°C-30s, 72°C-45s) 1×(72°C-1min). Reactions were then spiked with pLinker and subjected to exponential PCR amplification: 1×(98°C-3min) 35×(98°C-40s, 65°C-30s, 72°C-45s) 1×(72°C-5min). Amplification products 600bp-1.2kb were isolated by agarose gel electrophoresis, and virus primer-specific products were enriched by magnetic streptavidin bead pull-down. Semi-nested PCR was performed with P2 and pLinker: 1×(98°C-3min) 35×(98°C-40s, 65°C-30s, 72°C-40s) 1×(72°C-5min). Amplification products 600bp-1.2kb were isolated by agarose gel electrophoresis.

Paired-End Library Production and Sequencing

Linkers were digested with AscI and removed by agarose gel purification. Fragments were then polished, purified, and dA-tailed as after the genomic DNA sonication. Fragments were then ligated to Illumina paired-end adapters and isolated by agarose gel electrophoresis. A final, 30-cycle, library PCR was conducted utilizing Illumina primers PE1.0 and PE2.0 according to manufacturer specifications, and amplification products 350bp-1kb were isolated by agarose gel electrophoresis. The final libraries were submitted to 2×50 paired-end deep sequencing using an Illumina HiSeq 2000. IC-Seq data have been uploaded into the NIH NCBI Bioproject Site (accession number PRJNA208398) SRA sample numbers SAMN07260902 and SAMN07260902 to be released upon publication.

Computational Analysis-Read Validation and Alignment

Each end of paired-end reads was 3’-trimmed to 36 bp and validated to ensure that correct priming and processing had occurred using Bowtie. Viral ends required the 20 bp P2 and the next 16 bp of viral sequence that is contiguous with the primer, allowing two mismatches. On the target side, presence of a perfect match to the remaining 7 bp of linker sequence was required. The 29 bp remainder of the target side was aligned with the human genome (hg18/NCBI Build 36.1) using Bowtie. Up to 2 mismatches were allowed and unique alignments in the best alignment stratum were required. Identical target alignments, same strand and position, were combined into a single putative unique integration event, and any event supported by a single alignment was not considered in further analyses.

Determination of Integration Hotspots

Individual integration hotspots were defined as a region of at least three integration events for which the frequency of events differed in a statistically significant fashion, $p < 1E\text{-}9$ as determined by a negative binomial test, from a random distribution along the genome (Klein et al., 2011; Oliveira et al., 2012). Hotspots with 100% repeat overlap, as defined by RepeatMasker, or present on the Y chromosome were removed from consideration as probable artifact. Circos was used to generate circular whole-genome visualizations (Krzywinski et al., 2009).

Identification of consensus motif

MEME-Chip was utilized to identify a consensus integration motif within 50 bp of integration sites (Machanick and Bailey, 2011). Default program settings were applied, with a single modification allowing a search for a maximum of 50 bp consensus.

Integration correlates—Analyses of genomic correlation were conducted utilizing PyBedTools (Dale et al., 2011; Quinlan and Hall, 2010) and ENCODE data sets (Consortium, 2011).

Results

B19V and the host factor interactions that contribute to its viral life cycle are not fully characterized. The nonstructural NS1 protein is essential for viral gene expression and DNA replication (Guan et al., 2009; Tewary et al., 2014). NS1 possesses structural and enzymatic functions that are similar to the AAV Rep proteins, with a high degree of amino acid homology (Figure 1A). Activity of the replication proteins appears to be the major determinate for integration of AAV (Chiorini et al., 1996; Huser et al., 2014; Huser et al., 2010; Janovitz et al., 2013; Janovitz et al., 2014; Kotin et al., 1990). We hypothesized that B19V infection can result in DNA damage and vDNA integration into human host DNA, where genomic insertion sites are largely driven by the sequence affinity and activity of the NS1 replication protein. We set out to investigate whether B19V infection of its known target cell, erythroid progenitor cells, leads to vDNA integration.

To test this hypothesis, we modified the IC-Seq protocol previously used to characterize integration of AAV2 and 5 into HeLa cells (Janovitz et al., 2013; Janovitz et al., 2014). Using commercially available patient derived B19V serum containing ~4×10^{11} ge/ml, we established infection conditions (~2,000 ge/cell and a final concentration of ~4×10^{8} ge/ml) for ex-vivo-expanded CD36+ EPCs (day 12 post differentiation), in which approximately 5–10% of cells are positive for viral capsid 24 hr post inoculation (Figure 1B). Under these conditions, high levels of B19V NS1 and capsid (VP1) transcripts were expressed (Figure 1C), and low-level viral genome amplification occurred (Figure 1C). We considered high level viral gene expression, a relatively low level of cytopathic effect, and a modest level of genome replication optimal for capturing B19V integration events. Large-scale infections of 1.5×10^{7} cells at 2000 ge/cell were carried out, and total DNA was harvested 24 hours post inoculation. Purified DNA from two samples was processed through the IC-Seq protocol (Figure 1E (Janovitz et al., 2013; Janovitz et al., 2014) described in the materials and methods) using B19V specific primers flanking the P6 promoter (illustrated in Figure 1D). The rigorous fragmentation, purification, and amplification steps of the IC-Seq protocol are designed to eliminate PCR artifact (see materials and methods).

Parvovirus B19 Integration and Paucity of Hotspots

As described, validation of each read required at viral ends the 20 bp P2 primer and the next 16 bp of viral sequence that is contiguous with the primer, allowing two mismatches. On the target side, the presence of a perfect match to the remaining 7 bp of linker sequence was required. The 29 bp remainder of the target side was aligned with the human genome (hg18/

NCBI Build 36.1) using Bowtie. In total from the combined experimental sample data (Figure 2A), we captured and mapped over 564,000 validated B19V integration events to the human genome, which represented over 40,000 unique genomic positions.

B19V integration was pervasive throughout the human genome, targeting every chromosome with many hundreds to thousands of events (Figure 2B). In this circular depiction of the human genome, the outer ring depicts the size of each chromosome as well as chromosomal banding patterns. Each blue line of the inner ring indicates the location and number of reads corresponding to each integration event. Most chromosomes maintained a roughly comparable frequency of viral integration approximately 14 integrant sites/Mb (Figure 2C), with the X chromosome being the single exception. We found only chromosomes 13, 17, 19, and 20 possessed integration frequencies more than 25% divergent from the mean. We next determined whether there were genomic hotspots of B19V integration. Genomic hotspots are specific regions attracting particularly high densities of integration activity. Hotspot analysis utilizes an algorithm to scan the human genome comparing the observed integration frequency with that expected by chance, using a statistical significance cutoff of $p < 1E\text{-}9$ (Janovitz et al., 2013; Janovitz et al., 2014; Oliveira et al., 2012). This unbiased analysis revealed a single definitive hotspot (Figure 2B). This hotspot occurred in chromosome 1, within the calmodulin binding transcription activator 1 (CAMTA1) tumor suppressor gene, and contained integration events at ten distinct nucleotide positions. Thus, despite widespread integration, a lone B19V hotspot was identified, representing approximately 0.02% of all genomic integration events.

Integrant Association with Genes and Transcription

We next investigated the relationship between B19V integrations and specific gene regions. This analysis demonstrated significant enrichment of B19V insertions within genes as compared to that expected by chance (Figure 3A). Additionally, we found the distribution of B19V integration activity is not uniform across specific gene regions. Integrations were underrepresented around the beginning and end of genes (transcription start site (TSS) and transcription termination site (TTS) respectively), but were highly enriched within introns ($p < 0.001$). Thus, B19V integration occurs preferentially within genes and is most enriched within introns. While there was a paucity of statistically significant hotspots, we noted that many genes in addition to CAMTA1 were targeted with multiple integration events. We better visualized this phenomenon by generating a word cloud diagram of the 1,000 genes with the greatest number of B19V insertion events (Figure 3B). This format highlights that many genes received some degree of integration, while only a relatively small subset of genes including CAMTA1 (calmodulin binding transcription activator 1), PTPRN2 (Protein Tyrosine Phosphatase, Receptor Type N2), and CDH4 (Cadherin 4), were targeted with greater frequencies. Having established that B19V integrated preferentially within genes, we next investigated whether transcriptional activity influenced viral insertion sites in CD36+ erythroid progenitors. We found that the expression level of genes targeted by B19V did not differ substantially from the expression frequencies of the gene array itself (Figure 3C). Thus, the transcriptional level of genes did not correlate with B19V integration frequency.

We next sought to understand the relationship between B19V integration and chromatin structure. To accomplish this, we analyzed the enrichment of B19V integrations associated with specific histone markers. Employing data from CD36+ erythroid progenitor CHIP-Seq experiments (Cui et al., 2009), we discovered that B19V integration was highly enriched within chromatin marked by H3K9me3 ($p < 0.001$) (Figure 3D). H3K9me3 is an established marker for heterochromatin and associates with repressed or inactive gene loci (reviewed in (Kimura, 2013)). Genomic regions associated with other histone markers attracted less integration than expected by chance ($p < 0.001$). Thus, statistical analysis reveals a unique nonrandom integration profile for B19V following infection of CD36+ erythroid progenitor cells, positive correlation with H3K9me3 sites, but inverse correlation with all other histone markers that we investigated.

Proto-oncogenes and Tumor Suppressors

We noted that the lone B19V integration hotspot and several of the genes targeted with multiple integration events were either proto-oncogenes or tumor suppressors (Figure 3B). To investigate this observation further, we used the Sanger Institute Cancer Gene Census (SICGI) to determine the B19V integration activity within known causal oncogenes (Figure 4A). In this Circos plot, each B19V integration site is represented by a blue line, each of the SICGI genes is shown as a green line on the interior of the Circos plot, and B19V integrations into SCGI genes is depicted by a red line forming the inner circle of the plot. We discovered that B19V integrated within 287 individual genes in the Cancer Gene Census. In fact, 2.1% of all B19V integration activity targeted these genes, a frequency two-fold higher than expected by chance ($p < 0.001$).

To better visualize the frequency of integration within individual proto-oncogenes and tumor suppressors, we created a word cloud diagram (Figure 4B). This analysis demonstrated that a small subset of these genes, including CAMTA1 (Calmodulin Binding Transcription Activator 1), RUNX1 (Runt Related Transcription Factor 1), and FHIT (Fragile Histidine Triad), were targeted at greater frequencies. Thus, B19V integration was enriched in tumor-associated genes and a select subset of these was predominantly targeted.

Discovery of Consensus Integration Motif

We hypothesized that B19V integration site selection may be influenced by sequence homology between the viral and host genomes that provide NS1 binding opportunities. Similar biology has been demonstrated to be the major driver of integration for the related adeno-associated viruses (Huser et al., 2014; Huser et al., 2010; Janovitz et al., 2013; Janovitz et al., 2014). To assess this possibility in an unbiased manner, we used an algorithm to scan for repeated sequence motifs in the human DNA within fifty base pairs of B19V integration sites. This unbiased analysis revealed that the largest and most statistically significant hit was a forty-one base pair consensus motif with a p-value of 1.7E-100 (Fig. 5A). The most conserved regions of this sequence bear a striking similarity with B19V NS1 binding sites and a portion of the viral P6 promoter region, additionally, the six most statistically significant sequences supporting this consensus possess a remarkable degree of homology (Figure 5B).

One region of the B19V consensus integration motif was particularly highly conserved. This core sequence is 6–10 bases long and has viral P6 promoter homology. We noted that the sequence includes one poorly conserved nucleotide, with either an A or G being most common. These sequences we termed the “G variant” (CTGTAGTC) and the “A variant” (CTGTAATC), and sought to investigate the relationship between B19V integration and the genomic distribution of these sequences. We found that the chromosomal distribution of B19V integrations was highly correlated with the presence of both variants of the core sequence motif (Figure 5C). This correlation explained up to 87% of the variability in chromosomal B19V integrant distribution. We hypothesized that the distribution of these core consensus motifs influenced the unusual pattern of B19V integration events within specific gene regions. Therefore, we compared the observed frequency of the two variant core motifs occurring within specific gene regions to that expected by chance (Figure 5D). This analysis revealed that both A and G variants of the promoter-like motif were highly enriched in genes, especially introns.

Discussion

The functional homology between parvoviral nonstructural replication proteins led us to hypothesize that parvoviruses have the capacity to nick host cellular DNA in a manner that facilitates viral DNA integration. Using infection of $CD36^{+}$ EPCs by the pathogenic B19V to test this hypothesis, we established that B19V viral DNA integration occurs throughout the human genome of infected cells. To our knowledge, this study represents the first direct genome wide characterization of autonomous parvovirus integration. At a minimum, these events are mutagenic, and with AAV as an example, rescue of integrated intact parvoviral genomes may serve as a form of viral latency.

B19V integrants were preferentially associated with intronic regions of genes. Although they were found equally represented in the entire transcriptional range of genes, they were positively correlated with H3K9me3 chromatin markers and negatively correlated with all other chromatin markers screened. B19V integrants revealed a highly disperse integration profile with a single computationally identified genomic hotspot, the CAMTA 1 tumor suppressor gene locus. Furthermore, through computational analysis of sequences within 50 bp of integrants we discovered a 41 bp integration consensus sequence. Within this consensus sequence is a core octanucleotide (CTGTa/gTCC). The distribution of core consensus motifs within transcription units highly correlates with that observed for B19V integration itself, demonstrating a marked bias toward introns. Homology to the core octanucleotide sequence was found once in the B19V genome, 5’ to the NS1 P6 promoter TATA like element. The B19V core homology overlaps with an SP1/3 binding motif upstream of the P6 TATA domain. Gel shift assays have shown NS1 binding to AGGGCGGA octanucleotide GC elements upstream of the P6 promoter (Gareus et al., 1998; Raab et al., 2001; Raab et al., 2002). Additionally, within the B19V ITR, NS1 binding motifs have been identified as GC rich imperfect palindromic repeats of CCGGCGGC (Ganaie et al., 2017; Guan et al., 2009; Sanchez et al., 2016). Recent studies have found NS1 can complex with STAT5 to facilitate replication (Ganaie et al., 2017). This genomic configuration of NS1 interacting with DNA and cellular factors is similar to that of AAV Rep binding domains (McCarty et al., 1994).

The AAV2 P5 promoter includes a Rep protein binding domain adjacent to the P5 TATA box and Rep has been shown to interact with a variety of cellular proteins including TATA binding protein (Hermonat et al., 1998; Nash et al., 2009; Pereira et al., 1997; Pereira and Muzyczka, 1997). This interaction influences transcription, replication, and integration associated with the P5 promoter. Interactions between parvovirus replication proteins, cellular factors, the P5 promoter, and origin sequence elements have been proposed to direct AAV2 viral genomes to chromosomal sites (Huser et al., 2014; Janovitz et al., 2013; Janovitz et al., 2014). We anticipate similar interactions between NS1, cellular factors, host and viral genomes must be functioning to direct B19V DNA to integration sites within the human chromosome during B19V infections.

Although there are functional similarities in the parvoviral replication proteins, the integration correlates that we discovered for B19V infections are quite distinctive. In contrast to AAV integrations, where hotspots were a readily identifiable feature (Janovitz et al., 2013; Janovitz et al., 2014), B19V integrants were widespread and only one B19V hotspot was identified. The number of B19V and AAV5 specific integration events was very similar (4×10^4 vs 9.3×10^3 respectively) but 127 AAV5 hotspots were detected. Additionally, the pattern of genomic integration correlates for B19V was highly distinctive when compared to those established for AAV2 or 5 (Janovitz et al., 2013; Janovitz et al., 2014). Several differences between AAV and B19V integrations studies may contribute to the observed integrations patterns. AAV are dependoviruses and B19V is an autonomous virus. Infection of CD36+ EPCs by B19V is productive and cytopathic whereas AAV infections are nonproductive and generally not cytopathic. Active replication and viral mediated cytotoxicity may influence genome targeting. However in one study, AAV2 low throughput integration mapping data was acquired under productive replicative/cytopathic conditions (co-infection with Ad5 helper virus) (Drew et al., 2007). AAV2 integrants identified under productive infection conditions were largely consistent with those established in the absence of AAV2 replication. Therefore we would argue that active replication and a cytopathic environment are not major features that distinguish B19V integration targeting. Studies characterizing AAV integration have been carried out on transformed cell lines such as HeLa or MRC-5 lung fibroblasts whereas the use of primary erythroid progenitor cells with B19V studies are a significant experimental difference. We strongly believe cell specific differences in epigenetic genomic marking and nuclear transcription factors may impact integration targeting. From our observations with AAV2, AAV5, and B19V integration studies, we believe the driving force for integration targeting is the interaction of nonstructural proteins with specific DNA sequences that represent DNA binding and nicking sites. Amino acid sequence identity differences between parvovirus nonstructural proteins that confer DNA binding specificity, nicking substrate specificity, and interactions with cellular proteins are proposed to function as the major viral determinants for integration site targeting.

Although B19V infects a variety of tissue and cell types, including: erythroid CD34+ cells (Schmidt-Lucke et al., 2015), hepatic tissue (Bihari et al., 2013; Bonvicini et al., 2008; Thammasri et al., 2013), synovial tissue (Hokynar et al., 2000), tonsillar tissue (Pyoria et al., 2017), and myocardial endothelial cells (von Kietzell et al., 2014), infection of these cell types does not lead to a productive virus infection. Nonproductive infection by B19V may

be functionally similar to AAV infections in the absence of helper virus. In nonproductive AAV2 infections, low levels of Rep expression are sufficient to mediate stable AAV2 integration. We predict B19V infection of nonproductive cell lines generates low to modest levels of NS1, and these modest levels of NS1 are able to carry out B19V DNA integration. Future studies will determine if unproductive B19V infections generate an integration phenotype that is consistent with the correlates established in CD36+ erythroid progenitor cells. If integration of B19V is a form of latency, the parameters required for reactivation or rescue need to be defined. B19V transmission is generally considered to occur through the respiratory route. One possible mechanism of reactivation could be through superinfection with a helper virus. HHV6, EBV, and Adenovirus are candidate helper viruses that have been associated with inflammatory myocarditis (along with B19V) or tonsilar lymph node tissue. Studies have indicated that adenovirus infection can provide some helper function for transfected B19V DNA (Guan et al., 2009; Pozzuto et al., 2011; Winter et al., 2012). It is possible that both AAV and B19V may under go rescue following coinfection by helper viruses in vivo.

Conclusions

We present the first direct evidence that B19V infection of erythroid progenitor cells disrupts the human genome and facilitates viral DNA integration. Integration occurs in a genome wide manner and is predicted to depend on the endonuclease activity of NS1. Analysis of integration patterns revealed strong nonrandom correlations with gene intronic regions, H3K9me3 sites, and the identification of 41 base pair consensus sequence with an octanucleotide core motif.

Acknowledgments

T.J. was supported by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the National Institutes of Health under award number T32GM07739 to the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD Program. E.F.P. received support from the WR Hearst Foundation and PHS grant RO1 AI094050. This work was supported in part by the Intramural Research Program of the NIH, NHLBI.

References

Adamson-Small LA, Ignatovich IV, Laemmerhirt MG, Hobbs JA. Persistent parvovirus B19 infection in non-erythroid tissues: possible role in the inflammatory and disease process. Virus Res. 2014; 190:8–16. [PubMed: 24998884]

Bihari C, Rastogi A, Saxena P, Rangegowda D, Chowdhury A, Gupta N, Sarin SK. Parvovirus b19 associated hepatitis. Hepat Res Treat. 2013; 2013:472027. [PubMed: 24232179]

Bonvicini F, Filippone C, Manaresi E, Zerbini M, Musiani M, Gallinella G. HepG2 hepatocellular carcinoma cells are a non-permissive system for B19 virus infection. J Gen Virol. 2008; 89:3034–3038. [PubMed: 19008390]

Brown KE, Anderson SM, Young NS. Erythrocyte P antigen: cellular receptor for B19 parvovirus. Science. 1993; 262:114–117. [PubMed: 8211117]

Bua G, Manaresi E, Bonvicini F, Gallinella G. Parvovirus B19 Replication and Expression in Differentiating Erythroid Progenitor Cells. PLoS One. 2016; 11:e0148547. [PubMed: 26845771]

Chen AY, Kleiboeker S, Qiu J. Productive parvovirus B19 infection of primary human erythroid progenitor cells at hypoxia is regulated by STAT5A and MEK signaling but not HIFalpha. PLoS Pathog. 2011; 7:e1002088. [PubMed: 21698228]

Chiorini JA, Wiener SM, Yang L, Smith RH, Safer B, Kilcoin NP, Liu Y, Urcelay E, Kotin RM. The roles of AAV Rep proteins in gene expression and targeted integration. Curr Top Microbiol Immunol. 1996; 218:25–33. [PubMed: 8794243]

Consortium EP. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 2011; 9:e1001046. [PubMed: 21526222]

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell. 2009; 4:80–93. [PubMed: 19128795]

Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics. 2011; 27:3423–3424. [PubMed: 21949271]

Dobec M, Juchler A, Flaviano A, Kaeppeli F. Prolonged parvovirus b19 viremia in spite of neutralizing antibodies after erythema infectiosum in pregnancy. Gynecol Obstet Invest. 2007; 63:53–54. [PubMed: 16940737]

Drew HR, Lockett LJ, Both GW. Increased complexity of wild-type adeno-associated virus-chromosomal junctions as determined by analysis of unselected cellular genomes. J Gen Virol. 2007; 88:1722–1732. [PubMed: 17485532]

Ganaie SS, Zou W, Xu P, Deng X, Kleiboeker S, Qiu J. Phosphorylated STAT5 directly facilitates parvovirus B19 DNA replication in human erythroid progenitors through interaction with the MCM complex. PLoS Pathog. 2017; 13:e1006370. [PubMed: 28459842]

Gareus R, Gigler A, Hemauer A, Leruez-Ville M, Morinet F, Wolf H, Modrow S. Characterization of cis-acting and NS1 protein-responsive elements in the p6 promoter of parvovirus B19. J Virol. 1998; 72:609–616. [PubMed: 9420265]

Guan W, Wong S, Zhi N, Qiu J. The genome of human parvovirus b19 can replicate in nonpermissive cells with the help of adenovirus genes and produces infectious virus. J Virol. 2009; 83:9541–9553. [PubMed: 19587029]

Hermonat PL, Santin AD, Batchu RB, Zhan D. The adeno-associated virus Rep78 major regulatory protein binds the cellular TATA-binding protein in vitro and in vivo. Virology. 1998; 245:120–127. [PubMed: 9614873]

Hokynar K, Brunstein J, Soderlund-Venermo M, Kiviluoto O, Partio EK, Konttinen Y, Hedman K. Integrity and full coding sequence of B19 virus DNA persisting in human synovial tissue. J Gen Virol. 2000; 81:1017–1025. [PubMed: 10725428]

Huser D, Gogol-Doring A, Chen W, Heilbronn R. Adeno-associated virus type 2 wild-type and vector-mediated genomic integration profiles of human diploid fibroblasts analyzed by third-generation PacBio DNA sequencing. J Virol. 2014; 88:11253–11263. [PubMed: 25031342]

Huser D, Gogol-Doring A, Lutter T, Weger S, Winter K, Hammer EM, Cathomen T, Reinert K, Heilbronn R. Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome. PLoS Pathog. 2010; 6:e1000985. [PubMed: 20628575]

Im DS, Muzyczka N. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell. 1990; 61:447–457. [PubMed: 2159383]

Ishikawa A, Yoto Y, Tsugawa T, Tsutsumi H. Quantitation of human parvovirus B19 DNA in erythema infectiosum and aplastic crisis. J Med Virol. 2014; 86:2102–2106. [PubMed: 24962467]

Janovitz T, Klein IA, Oliveira T, Mukherjee P, Nussenzweig MC, Sadelain M, Falck-Pedersen E. High-throughput sequencing reveals principles of adeno-associated virus serotype 2 integration. J Virol. 2013; 87:8559–8568. [PubMed: 23720718]

Janovitz T, Oliveira T, Sadelain M, Falck-Pedersen E. Highly divergent integration profile of adeno-associated virus serotype 5 revealed by high-throughput sequencing. J Virol. 2014; 88:2481–2488. [PubMed: 24335317]

Kerr JR. Pathogenesis of parvovirus B19 infection: host gene variability, and possible means and effects of virus persistence. J Vet Med B Infect Dis Vet Public Health. 2005; 52:335–339. [PubMed: 16316396]

Kerr JR, Barah F, Mattey DL, Laing I, Hopkins SJ, Hutchinson IV, Tyrrell DA. Circulating tumour necrosis factor-alpha and interferon-gamma are detectable during acute and convalescent parvovirus B19 infection and are associated with prolonged and chronic fatigue. J Gen Virol. 2001; 82:3011–3019. [PubMed: 11714978]

Kimura H. Histone modifications for human epigenome analysis. J Hum Genet. 2013; 58:439–445. [PubMed: 23739122]

Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani DF, Casellas R, Nussenzweig MC. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell. 2011; 147:95–106. [PubMed: 21962510]

Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, McLaughlin S, Muzyczka N, Rocchi M, Berns KI. Site-specific integration by adeno-associated virus. Proc Natl Acad Sci U S A. 1990; 87:2211–2215. [PubMed: 2156265]

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645. [PubMed: 19541911]

Kurtzman GJ, Cohen BJ, Field AM, Oseas R, Blaese RM, Young NS. Immune response to B19 parvovirus and an antibody defect in persistent viral infection. J Clin Invest. 1989; 84:1114–1123. [PubMed: 2551923]

Kurtzman GJ, Ozawa K, Cohen B, Hanson G, Oseas R, Young NS. Chronic bone marrow failure due to persistent B19 parvovirus infection. N Engl J Med. 1987; 317:287–294. [PubMed: 3037373]

Lindblom A, Isa A, Norbeck O, Wolf S, Johansson B, Broliden K, Tolfvenstam T. Slow clearance of human parvovirus B19 viremia following acute infection. Clin Infect Dis. 2005; 41:1201–1203. [PubMed: 16163641]

Lou S, Luo Y, Cheng F, Huang Q, Shen W, Kleiboeker S, Tisdale JF, Liu Z, Qiu J. Human parvovirus B19 DNA replication induces a DNA damage response that is dispensable for cell cycle arrest at phase G2/M. J Virol. 2012; 86:10748–10758. [PubMed: 22837195]

Luo Y, Kleiboeker S, Deng X, Qiu J. Human parvovirus B19 infection causes cell cycle arrest of human erythroid progenitors at late S phase that favors viral DNA replication. J Virol. 2013; 87:12766–12775. [PubMed: 24049177]

Luo Y, Qiu J. Human parvovirus B19: a mechanistic overview of infection and DNA replication. Future Virol. 2015; 10:155–167. [PubMed: 26097496]

Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics. 2011; 27:1696–1697. [PubMed: 21486936]

Maggin JE, James JA, Chappie JS, Dyda F, Hickman AB. The amino acid linker between the endonuclease and helicase domains of adeno-associated virus type 5 Rep plays a critical role in DNA-dependent oligomerization. J Virol. 2012; 86:3337–3346. [PubMed: 22205752]

McCarty DM, Pereira DJ, Zolotukhin I, Zhou X, Ryan JH, Muzyczka N. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J Virol. 1994; 68:4988–4997. [PubMed: 8035498]

Moffatt S, Yaegashi N, Tada K, Tanaka N, Sugamura K. Human parvovirus B19 nonstructural (NS1) protein induces apoptosis in erythroid lineage cells. J Virol. 1998; 72:3018–3028. [PubMed: 9525624]

Munakata Y, Saito-Ito T, Kumura-Ishii K, Huang J, Kodera T, Ishii T, Hirabayashi Y, Koyanagi Y, Sasaki T. Ku80 autoantigen as a cellular coreceptor for human parvovirus B19 infection. Blood. 2005; 106:3449–3456. [PubMed: 16076874]

Nabae K, Satoh H, Nishiura H, Tanaka-Taya K, Okabe N, Oishi K, Matsumoto K, Hasegawa T. Estimating the risk of parvovirus B19 infection in blood donors and pregnant women in Japan. PLoS One. 2014; 9:e92519. [PubMed: 24658180]

Nash K, Chen W, Salganik M, Muzyczka N. Identification of cellular proteins that interact with the adeno-associated virus rep protein. J Virol. 2009; 83:454–469. [PubMed: 18971280]

Oliveira TY, Resch W, Jankovic M, Casellas R, Nussenzweig MC, Klein IA. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J Immunol Methods. 2012; 375:176–181. [PubMed: 22033343]

Ozawa K, Kurtzman G, Young N. Productive infection by B19 parvovirus of human erythroid bone marrow cells in vitro. Blood. 1987; 70:384–391. [PubMed: 3038211]

Pereira DJ, McCarty DM, Muzyczka N. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. J Virol. 1997; 71:1079–1088. [PubMed: 8995628]

Pereira DJ, Muzyczka N. The cellular transcription factor SP1 and an unknown cellular protein are required to mediate Rep protein activation of the adeno-associated virus p19 promoter. J Virol. 1997; 71:1747–1756. [PubMed: 9032303]

Pozzuto T, von Kietzell K, Bock T, Schmidt-Lucke C, Poller W, Zobel T, Lassner D, Zeichhardt H, Weger S, Fechner H. Transactivation of human parvovirus B19 gene expression in endothelial cells by adenoviral helper functions. Virology. 2011; 411:50–64. [PubMed: 21236463]

Pyoria L, Toppinen M, Mantyla E, Hedman L, Aaltonen LM, Vihinen-Ranta M, Ilmarinen T, Soderlund-Venermo M, Hedman K, Perdomo MF. Extinct type of human parvovirus B19 persists in tonsillar B cells. Nat Commun. 2017; 8:14930. [PubMed: 28374737]

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]

Raab U, Bauer B, Gigler A, Beckenlehner K, Wolf H, Modrow S. Cellular transcription factors that interact with p6 promoter elements of parvovirus B19. J Gen Virol. 2001; 82:1473–1480. [PubMed: 11369893]

Raab U, Beckenlehner K, Lowin T, Niller HH, Doyle S, Modrow S. NS1 protein of parvovirus B19 interacts directly with DNA sequences of the p6 promoter and with the cellular transcription factors Sp1/Sp3. Virology. 2002; 293:86–93. [PubMed: 11853402]

Sanchez JL, Romero Z, Quinones A, Torgeson KR, Horton NC. DNA Binding and Cleavage by the Human Parvovirus B19 NS1 Nuclease Domain. Biochemistry. 2016; 55:6577–6593. [PubMed: 27809499]

Schmidt-Lucke C, Zobel T, Schrepfer S, Kuhl U, Wang D, Klingel K, Becher PM, Fechner H, Pozzuto T, Van Linthout S, Lassner D, Spillmann F, Escher F, Holinski S, Volk HD, Schultheiss HP, Tschope C. Impaired Endothelial Regeneration Through Human Parvovirus B19-Infected Circulating Angiogenic Cells in Patients With Cardiomyopathy. J Infect Dis. 2015; 212:1070–1081. [PubMed: 25805750]

Servant-Delmas A, Lefrere JJ, Morinet F, Pillet S. Advances in human B19 erythrovirus biology. J Virol. 2010; 84:9658–9665. [PubMed: 20631151]

Sol N, Le Junter J, Vassias I, Freyssinier JM, Thomas A, Prigent AF, Rudkin BB, Fichelson S, Morinet F. Possible interactions between the NS-1 protein and tumor necrosis factor alpha pathways in erythroid cell apoptosis induced by human parvovirus B19. J Virol. 1999; 73:8762–8770. [PubMed: 10482630]

Streitz M, Noutsias M, Volkmer R, Rohde M, Brestrich G, Block A, Klippert K, Kotsch K, Ay B, Hummel M, Kuhl U, Lassner D, Schultheiss HP, Volk HD, Kern F. NS1 specific CD8+ T-cells with effector function and TRBV11 dominance in a patient with parvovirus B19 associated inflammatory cardiomyopathy. PLoS One. 2008; 3:e2361. [PubMed: 18523634]

Takahashi T, Ozawa K, Takahashi K, Asano S, Takaku F. Susceptibility of human erythropoietic cells to B19 parvovirus in vitro increases with differentiation. Blood. 1990; 75:603–610. [PubMed: 2404522]

Tewary SK, Zhao H, Deng X, Qiu J, Tang L. The human parvovirus B19 non-structural protein 1 N-terminal domain specifically binds to the origin of replication in the viral DNA. Virology. 2014; 449:297–303. [PubMed: 24418564]

Thammasri K, Rauhamaki S, Wang L, Filippou A, Kivovich V, Marjomaki V, Naides SJ, Gilbert L. Human parvovirus B19 induced apoptotic bodies contain altered self-antigens that are phagocytosed by antigen presenting cells. PLoS One. 2013; 8:e67179. [PubMed: 23776709]

von Kietzell K, Pozzuto T, Heilbronn R, Grossl T, Fechner H, Weger S. Antibody-mediated enhancement of parvovirus B19 uptake into endothelial cells mediated by a receptor for complement factor C1q. J Virol. 2014; 88:8102–8115. [PubMed: 24807719]

Wan Z, Zhi N, Wong S, Keyvanfar K, Liu D, Raghavachari N, Munson PJ, Su S, Malide D, Kajigaya S, Young NS. Human parvovirus B19 causes cell cycle arrest of human erythroid progenitors via deregulation of the E2F family of transcription factors. J Clin Invest. 2010; 120:3530–3544. [PubMed: 20890043]

Weigel-Kelley KA, Yoder MC, Srivastava A. Recombinant human parvovirus B19 vectors: erythrocyte P antigen is necessary but not sufficient for successful transduction of human hematopoietic cells. J Virol. 2001; 75:4110–4116. [PubMed: 11287560]

Weigel-Kelley KA, Yoder MC, Srivastava A. Alpha5beta1 integrin as a cellular coreceptor for human parvovirus B19: requirement of functional activation of beta1 integrin for viral entry. Blood. 2003; 102:3927–3933. [PubMed: 12907437]

Winter K, von Kietzell K, Heilbronn R, Pozzuto T, Fechner H, Weger S. Roles of E4orf6 and VA I RNA in adenovirus-mediated stimulation of human parvovirus B19 DNA replication and structural gene expression. J Virol. 2012; 86:5099–5109. [PubMed: 22357277]

Wong S, Zhi N, Filippone C, Keyvanfar K, Kajigaya S, Brown KE, Young NS. Ex vivo-generated CD36+ erythroid progenitors are highly permissive to human parvovirus B19 replication. J Virol. 2008; 82:2470–2476. [PubMed: 18160440]

Young NS, Brown KE. Parvovirus B19. N Engl J Med. 2004; 350:586–597. [PubMed: 14762186]

Zarate-Perez F, Bardelli M, Burgner JW 2nd, Villamil-Jarauta M, Das K, Kekilli D, Mansilla-Soto J, Linden RM, Escalante CR. The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization. PLoS Pathog. 2012; 8:e1002764. [PubMed: 22719256]

Zhou X, Zolotukhin I, Im DS, Muzyczka N. Biochemical characterization of adeno-associated virus rep68 DNA helicase and ATPase activities. J Virol. 1999; 73:1580–1590. [PubMed: 9882364]

Highlights

- B19V infection of CD36+ EPCs leads to DNA damage and B19V DNA integration
- IC-Seq revealed 40,000 B19V integration sites distributed throughout the human genome
- B19V integrants were preferentially found in gene intronic regions
- A consensus sequence with an octanucleotide core motif was computationally identified

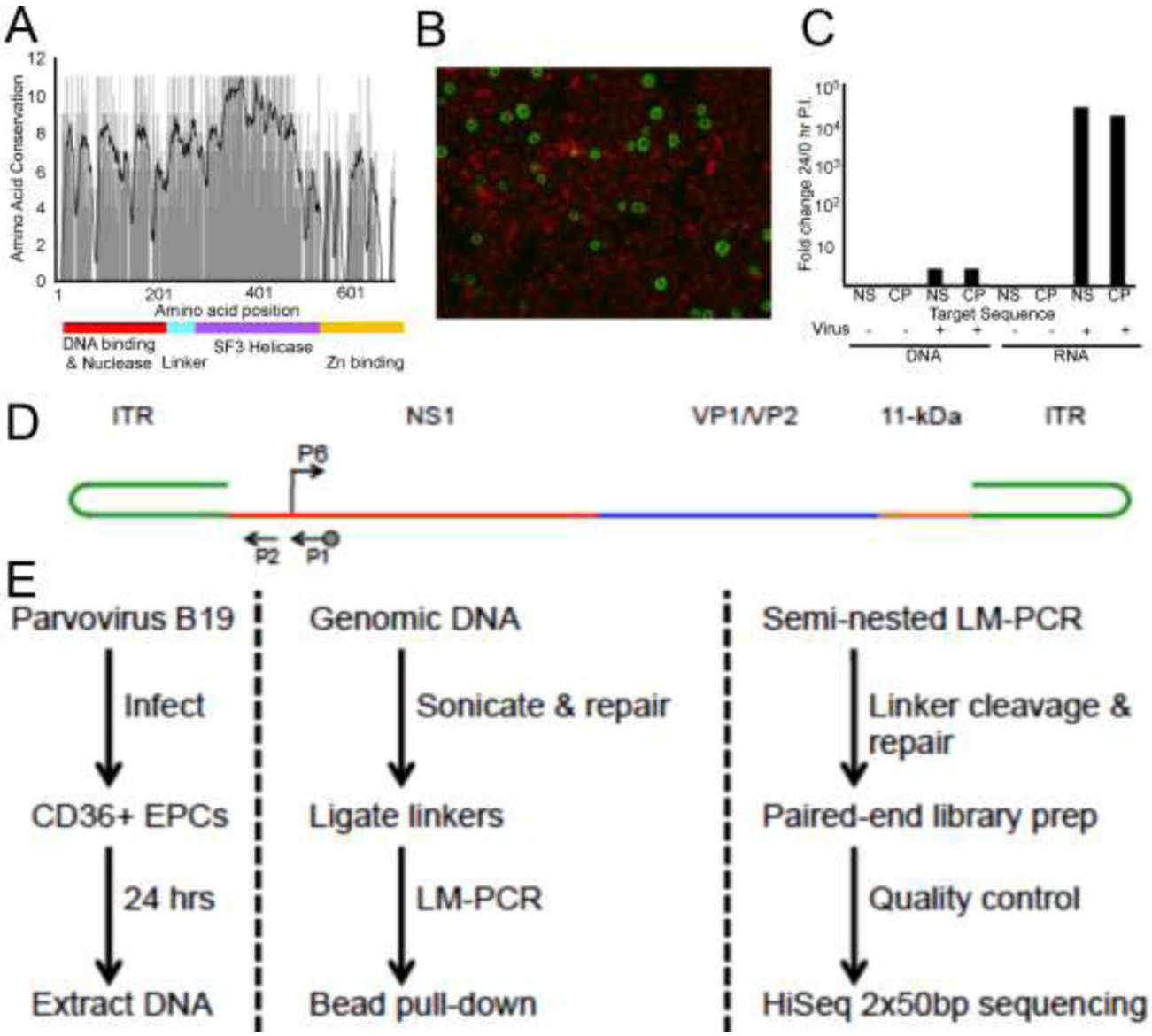

Figure 1. Parvovirus B19 infection of human primary CD36+ erythroid progenitor cells and integration-capture sequencing schematic

(A) Amino acid sequence homology between the major replication proteins of parvovirus B19V, AAV-2, and AAV-5. Black line on graph indicates 10-bp moving average of conservation (11 corresponds to 100% conserved). Colored line indicates functionally active domains established for parvovirus replication proteins. (B) Day 12 $CD36^{+}$EPCs infected with 2000 ge B19V/cell and harvested 24 hours post inoculation, B19V capsid proteins were detected by fluorescein immunofluorescence with Evans blue counterstaining. (C) Total DNA or RNA harvested at 0 and 24 hours post inoculation with 2000 ge B19V/cell quantified by qPCR or RT-qPCR. Fold change (24/0hr) in B19V NS1 and Capsid region CT values; normalization to cellular actin. (D) Schematic representation of B19V genome organization. Arrows indicate the location of primers used for nested PCR. Biotinylated

primer is indicated by dot on arrow end. B19V P6 promoter, arrow indicates direction of transcription. (E) IC-Seq protocol flow diagram (protocol details in materials and methods).

A

Sample	Sequences	Integrations	Unique Int.
B19A	2.20E+08	6.40E+04	1.74E+03
B19B	1.90E+08	5.00E+05	3.90E+04
B19 (2)	4.11E+08	5.64E+05	4.08E+04

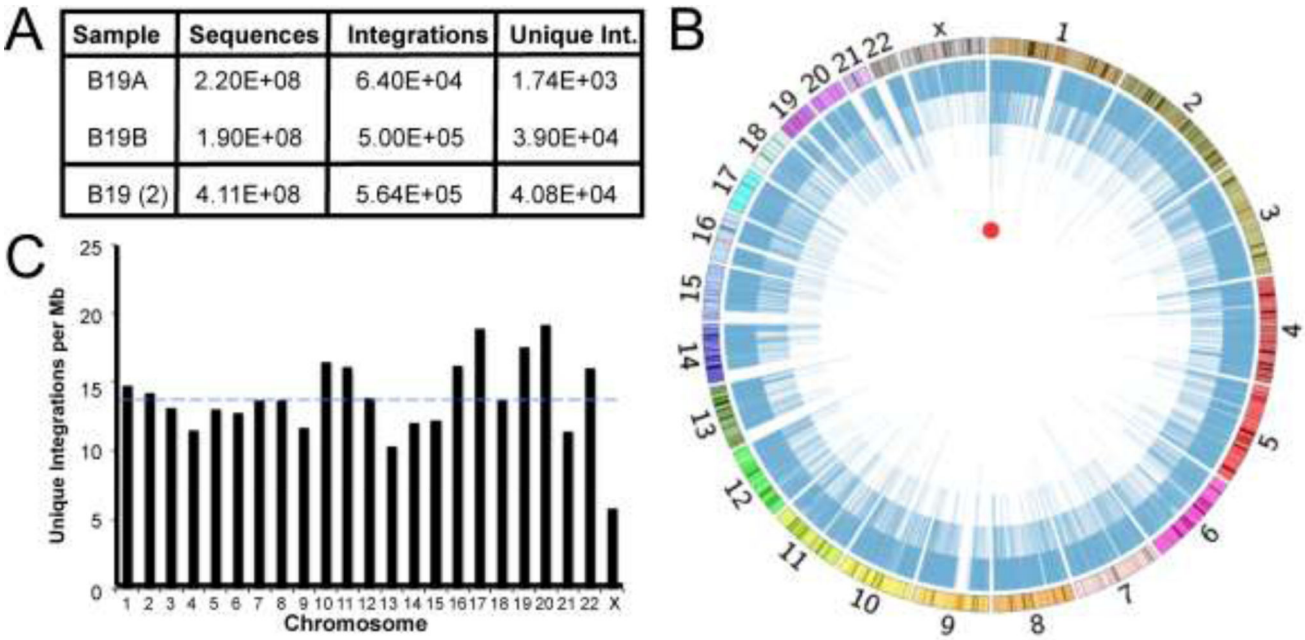

Figure 2. Genome-wide distribution of Parvovirus B19 integration in human erythroid precursors

(A) Summary of integrant-capture sequencing data acquired for B19A and B19B samples with the merge of both data sets. (B) Circos plot presenting genome-wide display of all unique B19V integration events. Chromosomal size and banding phenotype is presented in the external multi-colored ring. The chromosomal position of each unique B19V integration event is indicated as a blue line, where the 40,000 unique integration sites create the internal blue ring. The height of each blue bar corresponds to the read frequency for each site. The red circle indicates the single computationally determined integration hotspot located in CAMTA1 on chromosome 1. (C) Unique B19V integrations per mappable megabase for each chromosome. Dashed blue line represents the mean number of integrants/Mb for all chromosomes.

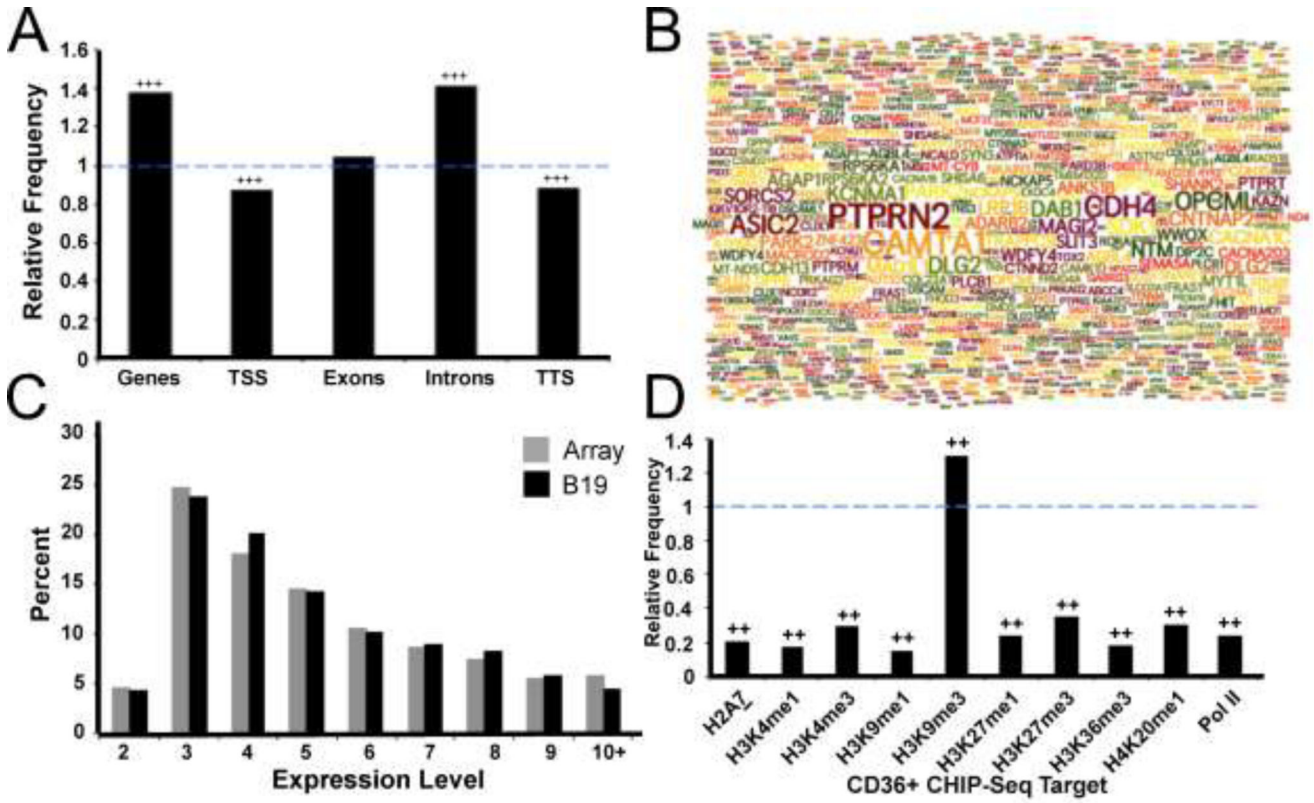

Figure 3. Parvovirus B19 integration is associated with specific gene regions and histone markers

(A) Relative frequency of B19V integration in genes and specific gene regions; fold enhancement of observed data compared to a random distribution model. The dashed blue line indicates the relative frequency expected in a random model. Presence of +++ represents $p < 0.001$ using a permutation test.

(B) Word cloud of the top 1,000 genes which attracted the highest numbers of B19V integration events; the size of the gene name represents the number of unique integrations received.

(C) Percent of B19V integration events occurring within CD36+ erythroid progenitor expression level gene groups (black), compared to the distribution expected based on the gene group array frequency (grey).

(D) Enrichment, displayed as relative frequency, of B19V integrations in several histone marker subtypes and Pol II using CD36+ erythroid precursor CHIP-Seq data. Dashed blue line represents expected frequency based on a random model. Presence of ++ indicates $p < 0.001$ using a permutation test.

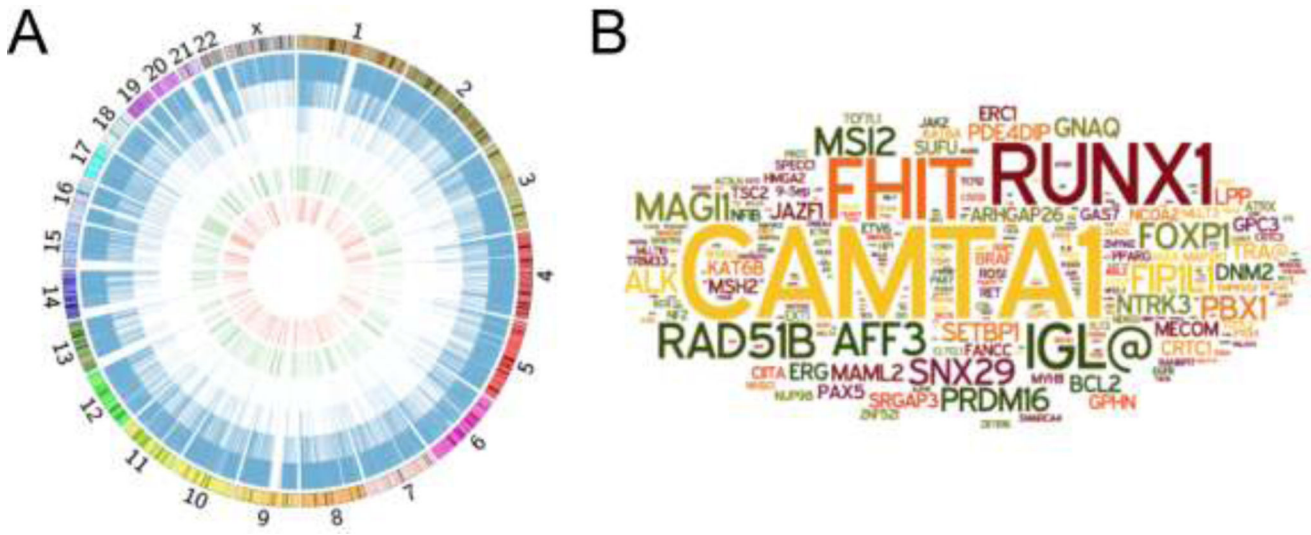

Figure 4. Oncogenes targeted by Parvovirus B19 integration

(A) Circos plot presenting total B19V integration events and those located in cancer associated genes. Genome-wide view of unique B19V integration events (blue bars), known genes relating to cancer induction (green), and those cancer-related genes targeted by B19V integration (red). Chromosomal size and banding phenotype is displayed in the external ring. (B) Word cloud diagram of the 287 genes involved with cancer induction that attracted B19V integration events; the size of the gene name correlates to the number of unique integrations.

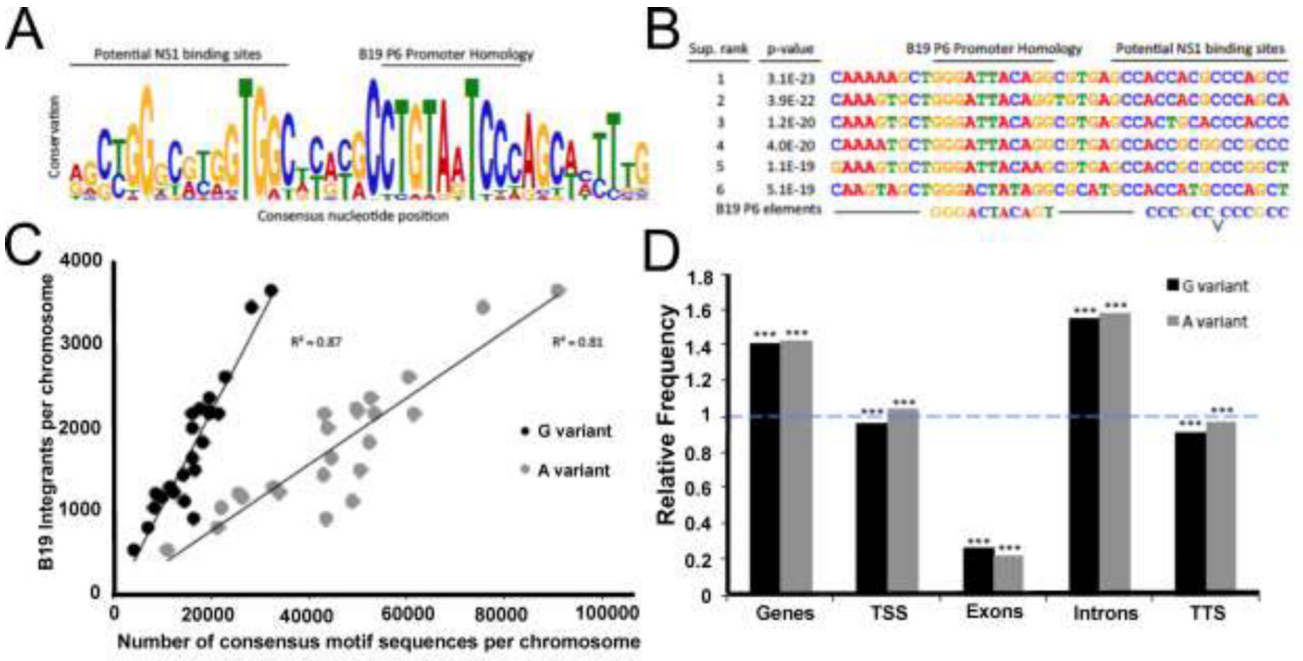

Figure 5. A large consensus DNA motif was discovered proximal to Parvovirus B19 genomic integration events and resembles the viral promoter

(A) Sequence LOGO diagram of the forty-one base pair consensus motif computationally identified in association with B19V integration (p-value = 1.7E-100). Total column stack height indicates information content, in bits, while individual letter size denotes the probability of the given nucleotide occurring in that position. Labeled bars display possible B19V P6 promoter functional correlation.

(B) The sequences of the top six supported B19V consensus motifs associated with integration events. Sites are ranked by p-value, as determined by the probability of a random sequence generating an equal or greater match score. Labeled bars indicate possible B19V P6 promoter correlation.

(C) Integrations per chromosome as a function of the highly conserved eight base pair motif core sequences per chromosome. The term variant refers to the least fixed nucleotide position in the core, most commonly the “G variant” (CTGTAGTC) or the “A variant” (CTGTAATC) shown in grey. Simple linear regression and R^2 value is presented.

(D) Enrichment of “G variant” (black) and “A variant” (grey) eight base pair motif core sequences in genes and specific gene regions, displayed as relative frequency. Dashed blue line denotes expected frequency based on a random model. *** indicates $p < 0.001$ using a permutation test.