AAV serotype 2 vectors preferentially integrate into active genes in mice

Hiroyuki Nakai¹, Eugenio Montini^{2,4}, Sally Fuess¹, Theresa A Storm¹, Markus Grompe^{2,3} & Mark A Kay¹

Recombinant adeno-associated virus serotype 2 (rAAV2) is a promising vector for gene therapy because it can achieve longterm stable transgene expression in animals and human subjects after direct administration of vectors into various target tissues¹. In the liver, although stable transgene expression primarily results from extrachromosomal vector genomes², a series of experiments has shown that vector genomes integrate into host chromosomes in hepatocytes^{3–5} at a low frequency². Despite the low integration efficiency, recent reports of retroviral insertional mutagenesis in mice⁶ and two human subjects^{7,8} have raised concerns about the potential for rAAV2mediated insertional mutagenesis. Here we characterize rAAV2-targeted chromosomal integration sites isolated from selected or non-selected hepatocytes in vector-injected mouse livers. We document frequent chromosomal deletions of up to

Figure 1 rAAV2 shuttle vectors, strategy for isolating vector-cellular DNA junctions and structures of rAAV2 proviruses isolated from mouse livers. (a) rAAV2 vector maps. ITR, AAV2 inverted terminal repeat; EF1 α P, the enhancer-promoter of the gene encoding eukaryotic translation elongation factor 1 α 1 (*EEF1A1*); lacP, bacterial *lac* operon promoter; β glpA, the poly(A)⁺ of human β-globin gene; FAH, FAH cDNA; B, BamHI; P, Pmel; S, Swal; X, Xbal. (b) Plasmid rescue strategies for isolating proviral genomes from rAAV2transduced mouse livers. A rAAV2 proviral genome integrated in a head-to-tail tandem array is indicated with straight (vector genome) and zigzag (mouse genome) thick lines. A representative form of extrachromosomal circular monomers, which are abundantly present in non-selected livers but not in in vivo selected livers, is indicated with a thick-lined circle. The strategies for isolation of junctions from non-selected hepatocytes transduced with AAV-EF1α-GFP.AOSP and from in vivo selected hepatocytes transduced with AAV-EF1 α -hFAH.AOS are shown above and below the provirus genome, respectively. Incorporation of a Pmel digestion followed by calf intestinal alkaline phosphatase (CIP) treatment greatly facilitated isolation of junctions from nonselected hepatocytes, but only one side of the junctions could be isolated⁴. AO, Amp^r/ori; ITR, AAV2 inverted terminal repeat. (c) Structures of proviral genomes isolated from in vivo selected hepatocytes transduced with AAV-EF1ahFAH.AOS. A representative structure of monomer provirus (P7) and three proviral genomes with various complicated structures (P9, P10 and P11) are shown, with a unit length vector genome to the upper right. Thin lines represent cellular DNA sequences. P9 contained a portion of the rAAV2 vector plasmid backbone sequence (shown with a gray box in the provirus). The internal structure of P10 (indicated with hatched boxes) was undetermined owing to the complexity. P, EF1 aP; FAH, FAH cDNA; AO, Amp^r/ori; AH, a 3' portion of FAH; O, 3' portion of AO; B, BamHI; K, KpnI; P, PmeI; S, SwaI; X, XbaI.

2 kb at integration sites (14 of 14 integrations, 100%; most of the deletions were <0.3 kb) and preferred integration into genes (21 of 29 integrations, 72%). In addition, all of the targeted genes analyzed (20 of 20 targeted genes, 100%) were expressed in the liver. This is the first report to our knowledge on host chromosomal effects of rAAV2 integration in animals, and it provides insights into the nature of rAAV2 vector integration into chromosomes in quiescent somatic cells in animals and human subjects.



¹Departments of Pediatrics and Genetics, Stanford University School of Medicine, 300 Pasteur Dr. Rm G305A, Stanford, California 94305, USA. Departments of ²Molecular and Medical Genetics and ³Pediatrics, Oregon Health & Science University, Portland, Oregon 97201, USA. ⁴Present address: Instituto per la cura e la ricerca del Cancro, Candiolo, Torino 10060, Italy. Correspondence should be addressed to M.A.K. (markay@stanford.edu).

Despite recent advances towards understanding the mechanisms of rAAV2 vector transduction in vivo (in animals), host chromosomal effects of vector integration and target site selection in quiescent somatic cells in animal tissues are not known, owing to the lack of an efficient system that allows for isolation of the infrequently integrated proviruses from the large number of extrachromosomal vector genomes in transduced tissues. Most previous studies analyzing rAAV2 integration used in vitro (in cell culture dishes) systems9-11, in which genetically unstable cell lines were transduced and clonally selected against a marker gene product. We previously established a plasmid rescue technique to retrieve rAAV2 vector-cellular DNA junction sequences as plasmids in bacteria (Fig. 1a,b and Supplementary Note online) and successfully isolated 18 vector-cellular DNA junctions from rAAV2 vector-injected normal C57BL/6 mouse livers without selection⁴. We reported the cellular sequences flanking the rAAV2 proviral genomes⁴ but could not annotate them owing to the lack of satisfactory information of the mouse genome at that time. Now, with the availability of the public mouse genome database¹², the precise integration sites in cellular chromosomes can be determined, to clarify the specificity of integration target site selection.

Vector 5'TTR EFICP FAH Amp' ori 3'TTR Provirus/mouse Provirus Provirus Mouse Mouse gagtgattagttagttagttagtagttagt CCTTAGTGATtgGACTTGGCC-3'TT P1 Mouse gagtgattagttagttagttagttagtagtagt CCTTAGTGATtgfttgttgtagtage P2 Mouse gagtgattagttagttagttagttagttagtagtage Mouse P2 Mouse gagtgattagttagttagttagttagt CCTTAGTGATtgfttgtage Mouse P2 Mouse tdttagttagttagttagttagttagttagt Mouse Mouse Mouse P3 SiTTR ACTTTGG TCGCCGGCC GCGACCAAAGTCGCCCGACC-3'TT P4 Mouse tgttagttagttagttagttagttagttagttagttagt				β			
Provirus Provirus Mouse EF10P—COTGAATTAGGTCATAGGGT_//_CCCTAGTCATGGAGTGGCC_3 / ITT P1 Mouse_gagtgactcagTCATAGGGT_//_CCCTAGTCATGGAGTGGCC_3 / ITT P2 Mouse_gagtgactcagTCATAGGGT_//_CCCTAGTCATGGAGTGGCC_3 / ITT P2 EF10P—GAACTAGAATTGAACCGGTC_//_TTTTGCGGCCTATTGCTA_works Mouse_tggactcagGgggggcc_//_agtgggadggaggggaggggggggggggggggggggggg	Vect	or 5'ITR	EF10P	FAH	Amp ^r	ori 3	'ITR
Mouse P1 Mouse-gagtgactcagrocurate CCCTAGTARTGACTAGGCC-3'IT P1 Mouse-gagtgactcagrocuract CCCTAGTARTGCTTTGCCCATTGCCTAGC JITT P2 Mouse-gagtgactcagrocuract CCCTAGTARTGCTTGCCAC-MOUSE Int7.115829087 P3 Mouse-ctggacgcccTGAACCGGTC-//-TTTTGCTGGCaaagcgactc-Mouse Int7.115829087 P4 Mouse-ctggacgcccTGAACCGGTC-//-TTTTGCTGGCaaagcgactc-Mouse Mouse-ctggacgcccTGAACCGGTC-//-GGAACCAAGCTGGCCGAC-3'ITT Mouse-tgccagac ggacgacgacggacgac P4 Mouse-tgccagac ggaatgagga-/-aagtcccaagcaggaga-Mouse Int9.6422037 P4 Mouse-caacttfgatTCGCGCACGGC-//-GGAACCCCTAtGGCGCAC-3'ITT Mouse-caacttfgatTCGCGCCGCCC-//-GGAACCCCTAtgGcaCtMouse Mouse-caacttfgatTCGCGACGGC-//-GGAACCCCTAtgGcaCtMouse Int1.87860175 Int1.87860175 P5 Mouse-tggacattggatgcagcag-/CCCGGGCGAACCAAAGGTCGCCC-3'ITT Mouse-ggaagaggcCCCCGGAGC-//-CCGGGCGAACCAAGGTCGCCC-3'ITT P6 S'ITR-CACTAGGGTTCCTGGAGGG-//-CCGGGCGAACCAAAGGTCGCCC-3'ITT Mouse-ggaagagggcCCCCGGAGCA-//-CCGGGCGAACCAAGGTCGCCC-3'ITT P7 Mouse-ggaagagggcGCCCCGGAGC-//-CCGGGCGAACCAAAGGTCGCCC-3'ITT Mouse-ggaagaggggggggggggggggggggggggggggggg	Prov	irus/mouse		Provi	rus		
 FF1oP—COTGANTACGTCATAGGGT // CCCTAGTGATGGCC-3'ITT Mouse—gagtgatcagTCATAGGGT // CCCTAGTGATtgtctgtaac—Mouse nt7.1158290731 fnt7.1158290731 fnt4.96882030 frmaced control of the second control	Mous	e				-	
 BF10P—COTCAATTAGGTCATAGGTC—//_CCCTAGTGATGAGGTGGCC_ATTAGGT MOUSE—gagtgattcaGTGATAGGT_//_CCCTAGTGATGGCGCGCCA_ATTAGGTCATGGCGTGATGGCGTGATGGGCGTGATCGGGCGTGATCGGGCGTGATCGGGCGTGATGGGGGCGTGATGGGGGGCGGGGGGGG			**		▼	▼	
 Moisegadtgettagtagtagtagtagtagtagtagtagtagtagtagtag	D1	Mouse—gag	GAATTACGTC	ATAGGG T//- ATAGGG T//-	-CCCTAGTGAT	rGGAGTTGGCC-	-3'ITR -Mouse
 nt7.11582907311 Int7.115829087 P2 Mouse -ctggacgcccAACGGTC // TTTGCTGGCCCTTTGCTGC-Mouse Mouse -ctggacgcccdgcgggccc // agcagctggcaacgcgatc - Mouse Mouse -ctggacgcccagcgggccc // agcagctggcaacgcgaacgcgatc - Mouse Mouse -tgccagateGTCGCCCGGCC // GCACCAAAGGCGCCCCACGC - 3'ITT Mouse -tgccagateGTCGCCCGCGCC // GCACCCAAAGGCGCGCCACGG - 3'ITT Mouse -tgccagateGTCGCCCGCGC // GGAACCCCTACGGCGCGACC - 3'ITT Mouse - caactttgattCCGCACGGC // GGAACCCCTACGGCGCGAC - 3'ITT Mouse - caactttgattCtcaccg // - GGAACCCCTACGGCGCACGGC - 3'ITT Mouse - caactttgattCCGCACGGC // - GGAACCCCTACGGCGCACG - 3'ITT Mouse - caactttgattCtCGCACGGC // - GGAACCCCTACGGCGCACG - 3'ITT Mouse - caactttgattCtCGCACGGC // - GGAACCCCTACGGCGCAC - Mouse mt4.01486731 BF10P - GGCCGCCTCTGAACTAGAAT // - CCCGGGCAAACGgGCGCC - 3'ITT Mouse - tggaccttcGAACTAGAAT // - CCCGGGCAACCGggccatc - Mouse mt1.87860123 F11L 87860123 Mouse - ggaacttcGAACTAGAAT // - CCGGGCGACCAAGGTCGCCC - 3'ITT Mouse - ggaacttgcgtCacgacaag // - acdgGCCACCAAAGGTCGCCC - 3'ITT Mouse - ggaacttgcgtCacgacgaa // - acdgGCCACCAAAGGTCGCCC - 3'ITT Mouse - ggaacttgcgtCacgacgaa // - acdgGCACCAAAGGTCGCCC - 3'ITT Mouse - ggaagggGCCCCGGGCA // - CCGGGCGACCAAAGGTCGCCC - 3'ITT Mouse - ggaaggggGCCCCGGGCA // - CCGGGCGACCAAAGGTCGCCC - 3'ITT Mouse - ggaaggggGCCCCGGCCA // - CCGGGCGACCAAAGGTCGCCC - 3'ITT Mouse - ggaaggggGCCCCGGCCA // - CGGGCGACCAAAGGTCGCCC - 3'ITT Mouse - ggaaggggGCCCCGGCCCA // - CGGGCGCCCAAAGGTCGCCCC - 3'ITT Mouse - ggaaggggGCCCCGGCCCA // - CGGGCGCCCCAAGGCCCCCGCCG - Mouse mt1.356554841	••	Mouse—gag	tg a ctcagca	cgcaact	-cgcaacttat	tgtctgtaac-	-Mouse
 P2 EFIOP—GAACTAGAATTGAACCGGTG—// TTTTGCTGGCCTTTGGTCA—ori Mouse—ctggacgcccTGAACCGGTG—// TTTTGCTGGCCTTTGGTCA—ori Mouse—ctggacgccccggggggcc// agcagcggacatc_Mouse nt4.96820081 fnt4.9682003 P3 Mouse—tgccagac TCGCCCGGCC// GCGACCAAAGCagcgcaga Mouse nt9.642270571 fnt9.64226937 fnt9.64226937 P4 AGGAGGTCGTCCGCACGGC// GGAACCCCTAGTGATGAAGT—3'ITT Mouse—caactttgattattcCaGCGC// GGAACCCCTAGTGATGAAGT—3'ITT Mouse—caactttgattattcCaGCGC/// GGAACCCCTAGTGATGAAGT—3'ITT Mouse—caactttgattattcCaGCGC/// GGAACCCCTAGTGATGAAGT—3'ITT Mouse—caactttgattattccacg/// tgtgatagtggtcatct Mouse nt4.401486731 fnt4.4015781 fnt4.401486731 fnt4.4015781 P5 Mouse—tggaccttcGAACTGAAT—// CCCGGGCAAAGCCGGGCGC—3'ITT Mouse—tggaccttcGAACTGAAT—// CCCGGGCAAACGTGGCCC—3'ITT Mouse—ggaacatgcctCCTGGAAGGC/// CGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgcctCCTGGAAGGC/// CGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgcctCCTGGAAGGC/// CGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgcctCCTGGAAGGC/// CGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgccCCCGGGCA/// CCGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgccCCCGGGCA/// CCGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgccCCCGGGCA/// CGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgccCCCGGGCA/// CCGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgccCCCGGGCA/// CCGGGCCACCAAAGTGGCCC—3'ITT Mouse—ggaacatgccCCCGGGCA/// CCGGGCCACCAAAGTGCGCC—3'ITT Mouse—ggaacatgccCCCGGGCA/// CCGGGCCACCAAAGTGCGCC—3'ITT Mouse—gtaatttatCGGTGCCTA/// CGGTCCTGGCCTTTGCTGGC-3'ITT Mouse—gtaaatttatCGGTGCCTA/// GGTTCCTGGCCTTTTCCTGGCCAAAGTGGCG// Mouse Mouse—ctgtggaacaCGCCCGCGCC// TATTTTTCTAATACATTCA — Mouse Mouse=gtggaaaaaGCTCGCCCGC// TATTTTTCTAATACATTCA — Mouse Mouse=gtggaaaaaGCTCGCCCGCA/// TATTATCATTAATCATTAA — ori Mouse=gtggaaaaaGCTCGCCCGCA/// TTTAAACATGGCGGGTagagctgga Mouse nt5.1263845691 fnt5.126384847 // 111.1074215731 fnt11.107421558 // Amp^C_CCGGGGGGGGGA/// AGGCCCCCGGGCAAGCCC_3'ITT Mouse=ttgcccdAGGGTGCCCCCGC// TTTAACATGGAGGATACCCC-3'ITT Mouse=ttgcccdGGGGGGGGA/// AGGCCCCCGGGCAAGCCC_3'ITT Mouse=ttgcccdAGGGTGCCCCCGC// TTTAACATGGAGAAGCCC_3'ITT Mouse=ttgc		nt7.115	8290731			lint7.115829	087
 P2 Mole of the set o							
Mouse—ctg3acgcccagcggggccc—//_agcagctagcaa3cgcatc—Mouse nt4.968820081 fnt4.96882030 fnt4.96882030 fvt4.968820081 fnt4.96882030 fvt4.968820081 fnt4.96882030 fvt4.968820081 fnt9.64228937 fnt9.64228937 fnt9.64227057fl fnt9.64228937 fnt9.64227057fl fnt9.64228937 fnt4.4018673fl fnt4.40150781 fnt4.40150781 fnt4.40150781 fnt4.4019781 fnt4.40150781 fnt4.40150781 fnt4.4019781 fnt4.40150781 fnt1.878602031 fnt4.40150781 fnt4.40150781 fnt1.878602031 fnt4.40150781 fnt4.40150781 fnt11.878602031 fnt4.40150781 fnt4.40150781 fnt11.87860175 fnt11.87860175 fvouse=ctg3acttcg3acgagcag-//_ctcttaccactg3gccatc—Mouse nut1.878602031 fnt1.87860175 fnt11.87860175 fvouse=ctg3acttcGAACTAGAAT—//~CCCGGGCAACAAGGTCGCCCG_3/ITT Mouse=cgaacatgctCCTCGAAGCA_//~CCGGGCGACCAAGGTCGCCCG_3/ITT Mouse=gaacatgctCCTCGAAGCA_//~CCGGGCGACCAAGGTCGCCCG_3/ITT Mouse=gaacatgctCCGGGCA_//~CCGGGCGACCAAGGTCGCCC_3/ITT Mouse=gaacatgctCCGGGCA_//~CCGGGCGACCAAGGTCGCCC_3/ITT Mouse=gaacatgctCCGGGCA_//~CCGGGCGACCAAGGTCGCCC_3/ITT Mouse=gaacatgctCCGGGCA_//~CCGGGCGACCAAGGTCGCC_3/ITT Mouse=gaagaggggGGCCCGGGCA_//~CCGGGCGACCAAGGTCGCC_3/ITT Mouse=gaagaggggGgGCCGGGCA_//~CCGGGCGACCAAGGTCGCC_3/ITT Mouse=gaagaggggGgGCCGGGCA_//~CCGGGCGACCAAGGTCGCCC_3/ITT Mouse=gtagagggggGgGCCGGGCA_//~CCGGGCGACCAAGGTCGCC_3/ITT Mouse=gtagaggggggggggggggggggggggggggdgcCGGCCGGGCA_//~CCGGGCGACCAAGGTCGCCC_3/ITT Mouse=gtagaggggggggggggggggggggggdgcCGGCCGGGCA_//~CCGGGGCACCAAGGTCGCCCC_3/ITT Mouse=gtagaggggggggggggggggggggdgcdcCGGGCACCAAGGTCGCCGCC_3/ITT Mouse=gtagagggggggggggggggdgdggggggggdgggggggdggggg	P2	Mouse-ctq	gacgcccTGA	ACCGGTG-//-	-TTTTGCTGGC	aaagcgcatc-	-Mouse
 nt4.96820081 Int4.9682030 S'ITR—ACCTTGG_TCGCCCGGCC—//_GCGACCAAAGGTCGCCGCC_3'ITT Mouse—tgccagaG_GTGCCCGGCC_//_GCGACCAAAGGTCGCCCA3GGaga_Mouse nt9.642270571 Int9.642270571 Int9.642270571 Int4.90186-caactttgatTCCGCACGCC_//_GGAACCCCTATGATGGAGT_3'ITT Mouse—caactttgatTCCGCACGCC_//_GGAACCCCTATGGTCGGAGT_3'ITT Mouse—caactttgatTCCGCACGCC_//_GGAACCCCTATGGTCGGCGGT_3'ITT Mouse—cagacttcGACTAGAATAGAAT_//_CCCGGGCAAAAGGTCGCCC_3'ITT Mouse_tcggaccttcGACTAGAAT_//_CCCGGGCGACCAAGGTCGCCC_3'ITT Mouse_ggaactgcgtCCCGGGCA_//_CGGGCGACCAAGGTCGCCC_3'ITT Mouse_ggaactgcgtCtCGGGCGA_//_CGGGCGACCAAGGTCGCCC_3'ITT Mouse_ggaactgcgtCtacagaaa_//_acaggtCtadagaggtCtdc_Mouse nt2.443125381 Int11.87860175 S'ITR_TCACTGAGGCGCCCCGGGCA_//_CCGGGCGACCAAGGTCGCCC_3'ITT Mouse_ggaaggggGCCCCGGGCA_//_CCGGGCGACCAAGGTCGCCC_3'ITT Mouse_ggaaggggggggggggggggggggg_/_Cttacaccctaagcaacg_Mouse nt2.443125381 Int15.67303144 BF10P_TCAATGCGGCGCCCCGGGCCA_//_CCGGGCGACCAAGGTCGCCC_3'ITT Mouse_ggaagagggggggggggggg_/_Cttacaccctaagccaacg_Mouse nt18.673028981 Int15.67303144 BF10P_TCAATGCGCGCCCCGGCCA//_CCGGCCCCCGGCCTTTGCGGG_Mouse nt18.937442911 Int18.85654841 Int1.856564841 Int1.856564841 Int1.856564841 Int1.856564841 Int1.856564841 Int1.856564841 Int1.856564841 Int1.1074215731 Int11.107421568 P10 Mouse_gaaaccCCCCGCCCCCCCC//_CATGCGGGTTGAGCGGACGAAGGCCGCCCCCCCCCCCCC		Mouse-ctg	jĝac ĝ ccca g c	gggc g cc//-	-agcagctagc	aaagc gc atc-	-Mouse
 P3 5'ITR—ACCTTTGG TCGCCCGGCC—//—GCGACCAAAGGTCGCCCGAC—3'ITT Mouse—tgccagac ggactgaga_Mouse nt9.642270571 P4 Mouse—caacttgatTCGCCCGGCC—//—GGAACCCCTAGGAGTGATGGAGT_3'ITT Mouse—caacttgatCtCGCACGGC—//—GGAACCCCTAGTGATGGAGT_3'ITT Mouse—caacttgatCtCGCACGGC—//—GGAACCCCTAGTGATGGAGT_3'ITT Mouse—caacttgatCtCGCACGGC—//—GGAACCCCTAGTGATGGAGT_3'ITT Mouse—cagacttcGACGAGGTCGCCGCAAAGCCCGGGCGAACGGGGCCC_3'ITT Mouse—cagacttcGACTAGAAT—//—CCCGGGCAAACGGGGCCC_3'ITT Mouse—cggaactgccTCCTGGAGGG—//—CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgccTCCTGGAGGG—//—CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgccTCCTGGAGGG—//—CGGGCGACCAAGGTCGCCC_3'ITT Mouse=ggaacatgccTCCTGGAGGG—//—CCGGGCGACCAAGGTCGCCC_3'ITT Mouse=ggaacatgccTCCTGGAGGC_//—CCGGGCGACCAAGGTCGCCC_3'ITT Mouse=ggaagaggGGCCCCGGGCA—//—CCGGGCGACCAAGGTCGCCC_3'ITT Mouse=ggaagaggGCCCCCGGCAA//—CCGGGCGACCAAGGTCGCC_3'ITT Mouse=ggaagaggGCCCCCGGCAA//—CCGGGCGACCAAGGTCGCCC_3'ITT Mouse=ggaagaggGCCCCCGGCA//—CCGGGCGACCAAGGTCGCC_3'ITT Mouse=ggaagaggGCCCCCGGCCA//—CCGGGCGACCAAGGTCGCC_3'ITT Mouse=ggaagaggGCCCCCGGCCA//—CCGGGCGACCAAGGTCGCC_3'ITT Mouse=ggaagaggGCCCCCGGCCA//—CCGGGCGACCAAGGTCGCC_3'ITT Mouse=ggaagaggGCCCCCGGCCA//—CCGGGCGACCAAGGTCGCC_3'ITT Mouse=gaacatgcCCCCGGCCA//—CCGGGCGACCAAGGTCGCC_3'ITT Mouse=gaacatgcCCCCGGCCA//—CCGGCCGCCAAGGTCGCC_3'ITT Mouse=gaacatgcCCCCCGGCCA//—CCGGCCGCCAAGGTCGCC_3'ITT Mouse=gaacatgcCCCCCGGCCA//—CCGGCCGCCACAAGGTCGCC_3'ITT Mouse=gaacatttatCCGGTGCCCA//—CCGGCCGCCACAAGGTCGCCC_3'ITT Mouse=gaacatttatCCGGTGCCCCA//—CCTGCGCGGCTAGCCCAAGCC// Mouse=ggaacaagggggaggagggggggcgc//—tactgggaagggggggggggggggggggggggggggggg		nt4.96	8820081		-	Unt4.968820	30
 Mouse—tgccagac_gatgaga_//agtccaagcagggaga_Mouse nt9.642270571 BF10P—AGGGAGGTCGTCGCGCACGGC_//GGAACCCCTAGTGATGGAGT_3'ITT Mouse_caactttgatTCCGCACGGC_//GGAACCCCTAgtgccatcMouse nt4.401486731 BF10P—GGCCGCTCTAGACTAGAAC_//_GGAACCCCTAggccatcMouse nt4.401486731 BF10P—GGCCGCTCTAGACTAGAAC_//_CCGGGCGAAAGCCCGGGGCT_3'ITT Mouse_cgaacttcgatagaggcag_//_ctottacccatgagccatcMouse nt11.878602031 S'ITR—CACTAGGGTTCCTGGAGGG_//_CCGGCGACCAAAGCCCCG_3'ITT Mouse_ggaacatgcctCCTGGAGGG_//_CCGGCGACCAAGGTCCCCC_3'ITT Mouse_ggaacatgcctCCTGGAGGG_//_CCGGCGACCAAGGTCGCC_3'ITT Mouse_ggaacatgcctCCTGGAGGG_//_CCGGCGACCAAGGTCGCCC_3'ITT Mouse_ggaacatgcctCCTGGAGGG_//_CCGGCGACCAAGGTCGCC_3'ITT Mouse_ggaacatgcctCCGGGCA_//_CCGGCGACCAAGGTCGCC_3'ITT Mouse_ggaacatgcctCCGGGCA_//_CCGGCGACCAAGGTCGCC_3'ITT Mouse_ggaacatgcctCCGGGCA_//_CCGGCGACCAAGGTCGCC_3'ITT Mouse_ggaagaggggggggggggg_/_ctctcaccctaagccaacg_Mouse mt2.443125381 T11.673028981 BF10P—TAGAATGAACCGGTGCCTA_//_GGTTCCTGGCCTTTGCTGG_ori Mouse_gtacatttatCGGTGCCTA_//_GGTTCCTGGCCTTTTGCTGG_ori Mouse_gtacatttatCGGTGCCTA_//_GGTTCCTGGCCTTTTCTAGAGCCGCC_3'ITT Mouse_gtacatttatCGGTGCCTA_//_GGTTCCTGGCCTTTTCTAAACAATCAA-ori Mouse_gtacatttatCGGTGCCTA_//_CGTTCCTGGCCTTTTCTAAATACATCA_Amp^r Mouse_gtagaccgcaccCGCATGGCC_//_TATTTTCTAAATACATCA_Amp^r Mouse_gtggaaccagcaccdgcc//_TATTTTCTAAATACATCA_Amp^r Mouse_gtggaacaaggggggggg_/_ttgga_Mouse nt1.855564841 Mouse_gtagaacaggggcgcc_//_tactggggtagggctgga_Mouse nt5.1263845691 Mouse_gtagaacaggggcgcc_//_TTTAAACATGGTGCGGACTori Mouse_gaaaccctgtGCCTCCCCGT_//_TTTAAACATGGTGCGGACGT_ori Mouse_ggaaccctgtGCCTCCCCCACC_//_ATGGCGGGTTAATCATTAAC_ori Mouse_ggaaccctgtGCCTCCCCCACC_//_ATGGCGGGTTAATCATTAC-ori Mouse_gaaaccctgtGCCTCCCCCCCC_//_ATGGCGGGTTAATCATTAC-ori Mouse_ctgcggaacaaggggacadaaagaaccattggttttcta_Mouse nt19.319635401		5'ITR-ACC	TTTGG TCG	cccg g cc—//—	-GCGACCAAAG	GTCGCCGAC-	-3'ITR
 NOUSE - LGCC3QGC 'GGACGACGACGACGACGACGACGACGACGACGACGACGAC	P3	Mouse-tge	cagacacICG	cccg c cc//-	-GCGACCAAAG	caa g g c gaga-	-Mouse
P4 Mouse—AGGAAGGTCGTCCGCGACGGC—//—GGAACCCCTATGATGATGGAGT—3'ITT Mouse—caactttgattcttccaccg—//—tgtgdatagatggtcacttd—Mouse mt4.401486731 fmt4.40150781 p5 Mouse—tcggaccttcgAACTAGAAT—//—CCCGGGCAAACCCGGGCGCGT—3'ITT Mouse—tcggaccttcgAaCTAGAAT—//—CCCGGGCAAACCGGggcGCCC_Mouse nt11.878602031 fmt11.87860175 p6 S'ITR—CACTAGGGGTCCTGGAGGG—//—CGGGCGACCAAAGGTCGCCCG_3'ITT Mouse—ggaacatgccTCCTGGAGGG—//—CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgccTCCTGGAGGG—//—CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgccTCCTGGAGGG—//—CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgccTCCTGGAGGCA//—CCGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaagaggGGCCCGGGCAA//—CCGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaagaggGaggaggagg//—ctctcaccacaag_mouse nt2.443125381 fnt15.67303114 p7 5'ITR—TCACTGAGCCGCCGGGCAA//—CCGGGCGACCAAAGGTCGCC_3'ITT Mouse—cggaagaggGaggaggaggggg//—ctctcaccacaag_mouse nt18.6564841 fnt15.67303114 p8 Mouse—gtacatttatctgGTGCCTA—//~GGTTCCTGGCCTTTGCGGCTGGG_Mouse nt1.856564841 fnt1.85656205 nt8.937442911 fnt18.8556205 nt8.937442911 fnt18.93744503 p7'ITR—CTCTCTGCGCGCTGCCCGC//—CATGGCGGGTAATCATTAA—ori Mouse—ctgtggaccGCCAGTGGCCG//—CATGGCGGGTAATCATTAA—ori Mouse—ctgtggaccGCCCAGTGGCCG//—CATGGCGGGTAATCATTAA—ori Mouse—gtggaaaaggggcgc//—tattggggtaggagggam nt11.1074215731 fnt11.107421568 p11 Ampr—CACAGGTGCCCGCCGCC//—ATGGCGGGTTAATCATTAA—ori Mouse—gtggaaaaggggcgc//—tattggggtaggagggam nt11.1074215731 fnt11.107421568 p12 BF10P—TTAGGAGGTCGCCCGCCGC//—ATGGCGGGTTAATCATTAAC—ori Mouse—gtggaaaaaggggcgcd//—tattggggtagagctggamMouse nt5.1263845691 fnt19.31963885 p111 Ampr—CACAGGGTCGCCCCCCGC//—ATGGCGGGTTAATCATTAAC—ori Mouse—gaaacctgtcCCCCCCCGC//—ATGGCGGGTTAATCATTAAC—ori Mouse—gaaacctgtcCCCCCCCGC//—ATGGCGGGTTAATCATTAAC—ori Mouse—ttgctggaataggggCGCC//—ATGGCGGGCTAATCATTAAC—ori Mouse—ttgctgggaataggagGCGCGCACG//—ATGGCCGCCCGCGAAAGCCCC-3'ITT Mouse—ctgctggaatagggTCGCCCCCCGC//—ATGGCCGGCTAACCCTTATCATTAAC—ori Mouse—ttgctgctgcdCCCCCCGCGCACG//—ATGGCCGCCCGGCAAACCCCC-3'ITT Mouse—ctgctggaatagggGCGCCCCCCGCG//—ATGGCCGCCCGCGAAACCCC-3'ITT Mouse—ctgctgcdCCCCCACGCACG//—ATGGCCGCCCGGCGAAACCCCC-3'ITT Mouse—ttgctgctgc		nt9 6422	cagae gga 70571	atga g ga//-	-aagteeeaag	10119 642269	37
 P4 Mouse—AGGAAGCTCCTCCCACGGC—//-GAACCCCTAGGAGAGGAGT-3'ITT Mouse—caactttgattCCCCACGCC//-GAACCCCTAGGACCCT-Mouse nt4.401486731 [htt4.40150781 P5 Mouse—cggaacttcGAACTAGAAT—//-CCCGGGCAAACCgagccatc—Mouse Mouse—tgggaccttcggaagagcag_//-ctcttaccactgagccatc—Mouse nt11.878602031 [htt1.87860175 P6 S'ITR—CACTAGGGTTCCTCGAGGGC—//-CCGGGCGACCAAAGGTCGCCC-3'ITT Mouse—ggaacatgccTCCTGGAGGG—//-CGGGCGACCAAAGGTCGCC3'ITT Mouse—ggaacatgccTCCTGGAGGG—//-CGGGCGACCAAAGGTCGCC3'ITT Mouse—ggaacatgccTCCTGGAGGGC—//-CGGGCGACCAAAGGTCGCC3'ITT Mouse—ggaacatgccTCCTGGAGGGC—//-CGGGCGACCAAAGGTCGCC3'ITT Mouse—ggaacatgcCGCCGGGCA-//-CCGGGCGACCAAAGGTCGCC3'ITT Mouse—ggaacatgcCGCCGGGCA-//-CCGGGCGACCAAAGGTCGCC3'ITT Mouse—ggaagaggggggggggggg//-ctctcacccacg_Mouse nt15.673028981 [htt15.6730314 P7 S'ITR—TCACTGAGCCGCCCGGGCA-//-CGGGCGACCCAAGGTCGCCG-3'ITT Mouse—gtacatttatCGGTGCCTA-//-CGGTTCCTGGCCTTTGCGC3'ITT Mouse—gtacatttatCGGTGCCTA-//-GGTTCCTGGCCTTTGCGC3'ITT Mouse—gtacatttatCGGTGCCTA-//-CGGTCCTGGCCCTTTGCGC3'ITT Mouse—gtagatgggggggggggggggggggggggggggggggg		110910122	▼ ▼		•	▼	57
 Mouse—caactttgatTCCCGACGGC—//_GGAACCCCTAtggtcactct—Mouse nt4.401486731 FFIOP—GCCCGCTCTAGAACTAGAAT—//_CCCGGGCAAACCCCGGCGCA-3/IT Mouse—tcggaccttcdgaagagcag—//~ctcttaccactgagcatcd—Mouse nt11.878602031 fnt11.87860175 S'ITR—CACTAGGGGTTCCTGGAGGG—//~CGGGCGACCAAAGGTCGCCC_3/ITT Mouse—ggaacatgcctCCGGGCAGC//~CGGGCGACCAAAGGTCGCCC_3/ITT Mouse—ggaacatgcctCCCGGGCA—//~CGGGCGACCAAAGGTCGCCC_3/ITT Mouse—ggaacatgcctCCCGGGCA—//~CGGGCGACCAAAGGTCGCCC_3/ITT Mouse—ggaacatgcctCCCGGGCA—//~CGGGCGACCAAAGGTCGCCC_3/ITT Mouse—ggaagaggggggggggggggg//~ctctcaacgcaacg—Mouse nt15.673028981 EFIOP—TAGAATTGAACCGGTGCCTA—//~GGTTCCTGGCCACtagccaacg—Mouse nt18.65564841 Mouse—gtgaacatttaCCGGTGCCTA—//~GGTTCCTGGCCTTTGCTGG—ori Mouse—gtgaacatttaCCGGTGCCTA—//~GGTTCCTGGCTGTGCCGagCACCAAAGGTCGCC Mouse—gtgaacatttaCCGGTGCCTA—//~GGTTCCTGGCCTGGCTGG_Mouse nt1.856564841 fnt1.85656205 ori—AGTGGCTGCCCAGTGGCC—//—TATTTTTCTAAATACATCA_Amp^r Mouse—ctgtggaacccggcacctggc—//~tggtccaccatgtttcga_Mouse nt8.937442911 fnt8.93744503 fnt5.1263848647 Mouse—gtggaaaagggggggggggggggggggggggggggggg		EF1αP-AGG	GAGGTCGTC	GCACGGC-//-	-GGAACCCCT	GT G ATGGAG T -	-3'ITR
 nt4.401486731 [nt4.40150781 FF10P—GGCCGCTCTAGAACTAGAAT—// CCCGGGCAAAGCCCGGGCGT_3'ITT Mouse—tcggaccttcgAACTAGAAT—// CCCGGGCAAACtgagccatc_Mouse nt11.878602031 [nt11.87860175 5'ITR—CACTAGGGTTCCTCGGAGGC—// CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgcctCCCGGAGGA// CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgcctCCCGGAGGA// CCGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgcctcacgcaaa // -acaggtctcacgdagdttctc_Mouse nt2.443125381 [nt2.44312597 5'ITR—TCACTGAGGCCGCCCGGGCA// CCGGGCGACCAAAGGTCGCC_3'ITT Mouse—cggaagagggGCCCCGGGCA// CCGGGCGACCAAAGGTCGCC_3'ITT Mouse—cggaagagggGCCCCGGGCA// CCGGGCGACCAAAGGTCGCC_3'ITT Mouse—cggaagagggGCCCCGGGCA// CCGGGCGACCAAAGGTCGCC_3'ITT Mouse—cggaagagggGCCCCGGGCA// CCGGGCGACCAAGGTCGCC_3'ITT Mouse—cggaagagggGCCCCGGGCA// CCGGGCGACCAAGGTCGCC_3'ITT Mouse—cggaagagggGCCCCGGGCA// CCGGGCGACCAAGGTCGCC_3'ITT Mouse—cggaagagggGCCCCGGGCCA// CGGTCCTGGCCTAGGCGG// Mouse nt15.673028981 [nt15.67303114 FF10P—TAGAATTGAACCGGTGCCTA-// GGTTCCTGGCCTAggCCagaatttgg Mouse nt1.856564841 [nt1.85656205 99 ori=AGTGGTGCTGCCCACTGGCC// TATTTTTCTAATACATTCA_MDuse Mouse—ctgtggaacaCCCGCCGCCGC// CATGGCGGGTTATCATAA-ori Mouse—ctgtggaacaaGCCCGCTCGC// CATGGCGGGTAATCATTAA-ori Mouse—gtggaaaaaGCTCGCTCGCC// CATGGCGGGTAATCATTAA-ori Mouse—gtggaaaaaGCTCGCTCGCC// CATGGCGGGTAATCATTAA-ori Mouse—gtggaaaaaGCTCGCTCCCCGT // TTTAAACATGaaaaaccaaa_Mouse nt11.1074215731 [nt11.107421568 P11 Amp^r—CAGATGGTAGCCCTCCCCGT // TTTAAACATGGaaaaccaaa_Mouse nt19.319635401 [nt19.31963585 S'ITR—TCCTCG AGGGGTCGCCCC// AGGCCGCCCGGCCCCCGC // AGGCCGCCCGGCCCCCCCCCC	P4	Mouse—caa Mouse—caa	ictttgat TCC	G C ACGGC//-	-GGAACCCCT7 -tgtgcatage	tg g tcactc t-	-Mouse
 P5 Mouse—tcggaccttcgAACTAGAAT—// CCCGGGCAAAGCCCGGGCGT—3'ITT Mouse—tcggaccttcgAACTAGAAT—// CCCGGGCAAACtgagcctatc—Mouse nt11.878602031 nt11.87860175 P6 5'ITR—CACTAGGGGTCCTGGAGGG—// CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgccTCCTGGAGGG—// CGGGCGACCAAAGGTCGCCC_3'ITT Mouse—ggaacatgcctcacagcaaa// acaggtccaggacttctc_Mouse nt2.443125381 nt2.44312597 P7 5'ITR—TCACTGAGGCCGCCGGGCA—// CCGGGCGACCAAAGGTCGCC_3'ITT Mouse—cggaagggGGCCCCGCGGGCA—// CCGGGCGACCAaGGTCGCC_3'ITT Mouse—cggaagggGGCCCCGGCGCA// CCGGGCGACCAaGGTCGCC_3'ITT Mouse—cggaagggGGCCCCGCGGGCA—// CCGGGCGACCAaGGTCGCC_3'ITT Mouse—cggaagggGGCCCCGGCGCA// CCGGGCGACCAaGGTCGCC_3'ITT Mouse—cggaagggGGCCCGCGGGCA—// CCGGGCGACCAaGGTCGCG_3 Mouse nt15.673028981 nt15.67303114 P8 Mouse—gtacatttatCtgGGTGCTA—// GGTTCCTGGCTTTGTGGCTTGTGGCT P9 ori—AGTGGCTGCTGCCAGTGGCCG// TATTTTTCTAAATACATCA_Amp[*] Mouse—ctgtggaccGCCAGTGGCCG// TATTTTTCTAAATACATCA_Amp[*] Mouse—ctgtggaccGCCAGTGGCCG// TATTTTTTCTAAATACATCA_Amp[*] Mouse—ctgtggaccGCCCAGTGGCCG// CATGGCGGGTTAATCATTAA—ori Mouse—ctgtggaccGCCCAGTGGCCG// CATGGCGGGTTAATCATTAA—ori Mouse—gctggaaaagggggcgcg// ttggtccacactgga_Mouse nt5.1263845691 nt5.126384847 P10 Amp[*] CACATGGTAGCCCTCCCCGT // CATGGCGGGTTAATCATTAA—ori Mouse—gtggaaaagggggcgc// ttggtccacacatgga_Mouse nt5.1263845691 nt5.126384847 P11 Amp[*] CACATGTAACCCCTCCCCGT // TTTAACATGGACGCACT—ori Mouse—gtggaaaaggggCGCCCCCCCGC // ATGGCGGGTTAATCATTAA_Ori Mouse—gtgcggaaaagggGTCCCCCCCGT // TTTAACATGGACGCACT—ori Mouse—ttgctgggaatatggaaa// aggacctttggtLttctag_Mouse nt19.319635801 nt19.31963585 P13 Mouse—ctgctggGTCGCCCCCGCG // ATGGCCGGCTAATCATTAAC_Ori Mouse—ttgcctgcd_ACGGTGGCGCCCCCGGGCGACGCCCCGGCGACGCCCCCGCGGAAAGCCCCCC		nt4.40	1486731		-3-99-	Înt4.401507	81
 PFIGP-GCCGCTCTAGAACTAGAAT-//-CCCGGGCAAAGCCGGCGT-3'ITT Mouse-tcggaccttcggaagacag-//-ctcttacccatgagccatc-Mouse mtll.878602031 full.878602031 full.87656205 full.85656205 full.857422911 full.85456301 full.85456301 full.85456301 full.85456301 full.85456301 full.85456301 full.85456301 full.85456301 full.85456301 full.83742911 full.8384847 <li< th=""><th></th><th></th><th>••</th><th></th><th>•</th><th>•</th><th></th></li<>			••		•	•	
 Mouse_tcggadcttcggaagagcag_//_ctcttacccactgagccatt_Mouse ntll.87860203f S'ITR—CACTAGGGGTTCCTGGAGGG_//_CGGGCGACCAAAGGTCGCC_3'ITT Mouse_ggaacatgcctCCTGGAGGG_//_CGGGCGACCAAAGGTCGCC_3'ITT Mouse_ggaacatgcctCCCGGGCA_//_CGGGCGACCAAGGTCGCC_3'ITT Mouse_cggaagagggGggggggggg_//_ctctcacccctagccaacg_Mouse ntl5.67302898f S'ITR—TCACTGAGGCCGCCGGGCA_//_CGGGCGACCAAGGTCGCC_3'ITT Mouse_cggaagaggagggggggggg_//_ctctcacccctagccaacg_Mouse ntl5.67302898f BFIOP—TAGAATTGAACCGGTGCCTA_//_GGTTCCTGGCCTTTGCTGG_mori Mouse_gtacatttatCCGGTGCCTA_//_GGTTCCTGGCtagactttgg_Mouse ntl.8655484f BFIOP—TAGAATTGAACCGGTGCCTA_//_GGTTCCTGGCCTAGTCAATCATCA_Amp' Mouse_gtacatttatCCGGTGCCCA_//_ACCGGtgGCT_ACCgtggdtagactttgg_Mouse ntl.85656484f P9 ori_AGTGGCTGCCCAGTGGCC_//_TATTTTTCTAAATACATCA_Amp' Mouse_ctgtggaccagcacctggc_//_tggtccaccactgtttctga_Mouse nt8.93744291f P10 Mouse_gtggaaaagggggggggggggggggggdg//_tactggggtagggtgga_Mouse nt5.126384569f P11 Amp'_CAGATGGTAGCCTCCCCGT_//_TTTAACATGGTGCGGACGT_ori Mouse_gaaaccctgtGCCCTCCCCGT_//_TTTAACATGGTGCGGACGT_ori Mouse_gaaaccctgtGCCCTCCCCGT_//_TTTAACATGGTGCGGACGT_ori Mouse_gaaaccctgtGCCCTCCCCGT_//_TTTAACATGGTGCGGACGT_ori Mouse_gaaaccctgtGCCCTCCCCGT_//_TTTAACATGGTGCGGACGT_ori Mouse_gaaaccctgtGCCCTCCCCGT_/TTTAACATGGTGCGGACGT_ori Mouse_gaaaccctgtGCCCCCCCGC_//_ATGCCGGGTTAATCATTAAC_Ori Mouse_gaaaccctgtGCCCCCCCCGT_/_ATGCCGGGTTAATCATTAAC_Ori Mouse_gaaaccctgtgCCCCCCCCCG//_ATGCCGGGTTAATCATTAAC_Ori Mouse_tcgctggaaaaaggggagada_//_agaaccatttggttttctag_Mouse nt11.107421573f P13 Mouse_ctgcdcccdAGGGCCCCCCCGCAC_//_AGGCCGCCCGGCGCAACCCC_3'ITT Mouse_ctgcdcccdAGGGGGTGAAC_//_AGGCCGCCCCGCGCGCAACCCC_3'ITT P140 Dista4417f B140P_GTCATAGGGTTAGGGAGGTC_//_AGGCCGCCCGGCGCAACCCC_3'ITT 	P5	EF1αP-GGC	CGCTCTAGAA	CTAGAAT-//-		GCCCCGGGCGT-	-3'ITR
 ntll.878602031 [Intll.87860175 96 5/ITR—CACTAGGGGTCCTOGAGGG—//-CGGGCGACCAAAGGTCGCCC_3/ITT Mouse—ggaacatgcctCCGGGGGA 97 Mouse—cgaagaggccGCCGGCCGGCAA//-CCGGGCGACCAAAGGTCGCCGC_3/ITT Mouse—cgaagaggcGCCCGGCCGGCAA//-CCGGGCGACCAAAGGTCGCCGC_3/ITT Mouse—cgaagaggcGCCCGGCCGGCAA//-CCGGGCGACCAAAGGTCGCCGC_3/ITT Mouse—cgaagaggcGCCCGGCCAA//-CCGGGCGACCAAAGGTCGCCGC_3/ITT Mouse—cgaagaggcGCCCGGCCAA//-CCGGGCGACCAAAGGTCGCCGC_3/ITT Mouse—cgaagaggcGCCCGGCCAA//-CGGGCCGACCAAAGGTCGGCG_MOuse ntl5.673028981 [Intl5.67303114 98 Mouse—gtacatttatCCGGTGCCTA-//-GGTTCCTGGCCTTTGCGTGG_MOuse Mouse—gtacatttatCCGGTGCCTA-//-GGTTCCTGGCTGTGGC_MOuse Mouse—gtacatttatCCGGTGCCCAA//-GGTTCCTGGCLagactttgg_Mouse ntl.85654841 [Intl.8565484] 99 ori—AGTGGCTGCTCCGCGCGCAAGGTCGCGC//-TATTTTTCTAAATACATTCA-Amp^r Mouse—ctgtggaaccagcactggc=//-tggtccaccatgttttcga-Mouse nt8.937442911 [Intl.856548437 910 ?/ITR—CTCTCTGCGCGCTCGCCGC//-CATGGCGGGTTAATCATTAA-ori Mouse—gctggaaaaagGTCGCTCCCGC//-CATGGCGGGTAagagctgga-Mouse nt5.1263845691 [Int5.126384847 911 Amp^r—CAGATGGTAGCCCTCCCGT-//-TTTAACATGGTGCGGCACGT-ori Mouse—gaaccctgtctCGCACG-//-ATGGCGGGTTAATCATTAA-ori Mouse—gaaccctgtctCCGCACG-//-ATGGCGGGTTAATCATTAA-ori Mouse—gaaccctgtctCGCCCCCCGC//-ATGGCGGGTTAATCATTAA-Ori Mouse—gaaccctgtctCGC//TTTAAACATGGTGCGGCACGT-ori Mouse—gaaaccctgtctCGCCCCCCGC-//-ATGGCGGGTTAATCATTAAC-ori Mouse—gaaaccctgtctCGCCCCCCGC//ATGGCGGGTTAATCATTAAC-ori Mouse—gaaaccctgtctCGCACG-//-ATGGCGGCTTAATCATTAAC-ori Mouse—ttgcctggaagataggggcGCACG-//-ATGGCGGGCTAATCATTAAC-ori Mouse—ttgcctggaagataggGTCGCCCCCCCGC//ATGGCCGGCTAATCATTAAC-ori Mouse—ttgcctggaagataggGTCGTCCGCACG-//-ATGGCCGCCCGGCAAAGCCC-3/ITT 913 Mouse—cttgcccctgGCGCACGACG//ATGGCCGCCCGGCAAAGCCC-3/ITT 9148.941844171 [Int8.94184420 91645401 91745.941844171 91845400 	15	Mouse-tcg	gaccttcgga	ag ag c a g—//-	-ctcttaccca	ctgagccatc-	-Mouse
 P6 5'ITR—CACTAGGGGTTCCTGGAGGG—//—CGGGCGACCAAAGGTCGCCC—3'ITT Mouse—ggaacatgccTCCTGGAGGG—//—CGGGCGACCAagagcttctc—Mouse nt2.443125381 P7 5'ITR—TCACTGAGGCGCCGGGCAA//—CCGGGCGACCAAAGGTCGCC—3'ITT Mouse—cggaagaggGGCCCGGGCGAA//—CCGGGCGACCAAAGGTCGCC—3'ITT Mouse—cggaagaggGagggaggagg//—ctctcacccctaagccaacg—Mouse nt15.673028981 P8 BFloP—TAGAATTGAACCGTGGCTA—//—GGTTCCTGGCTTTTGCTGGMouse matactttatctgGGGCCA//—CCGGGCGACCAagccaacgMouse nt1.856564841 P9 ori—AGTGGCTGCCCAGTGGCCA//—acccagtggctagactttgg—Mouse nt1.856564841 P9 ori—AGTGGCTGCCCAGTGGCCA//—acccagtggctagactttgg—Mouse mt8.937442911 P10 p-TTGGGCGCTGCCCGCCGC//—TATTTTTCTAAATACATCA_Amp[*] Mouse—ctgtggacccGCCAGTGGCG//—TATTTTTCTAAATACATCA_Amp[*] Mouse—ctgtggaccGGCCAGTGGCC//—CATGGCGGGTTAATCATTAA—ori Mouse—gtggaaaaggggcgcgc//—tggccaccatggacttgga—Mouse nt5.1263845691 P11 Amp[*]—CACATGGTAACCCTCCCGC//—CATGGCGGGTTAATCATTAA—ori Mouse—gtggaaaagggggcgcd//—tattggggtagagctgga—Mouse nt1.1074215731 P12 BFloP—TTAGGGAGGTCCTCCCCGC//—ATGGCGGGTTAATCATTAAC—ori Mouse—gtggaaaacGTCGCCCCCCGC//—ATGGCGGGTTAATCATTAAC—ori Mouse—gaaccctgfctCG P13 Mouse—ctgctggagtGGCCCCCCCGCAGT//—ATGGCGGGCTAATCATTAAC—ori Mouse—ctgctggaatagagtGTCGCCCCCCGC//—ATGGCCGGCTAATCATTAAC—ori Mouse—ctgctggagtatgaaca//—acgaccCGCGACGCAAACCGCCCC-3'ITT Mouse—ttgcctggagtatgaaca//—acgaccCGCGGCAAACCCC-3'ITT P13 Mouse—ctgctggagtGCGCCCACGC//—ATGGCCGCCCCGCCGCAAGCCC-3'ITT Mouse—ctgctggaatagaca//—acgaccCCGCGCGCGCAAACCCC-3'ITT P13 Mouse—ctgctgc dccd ACGGGTGGACG//—ATGGCCGCCCCGGCAAACCCCCCGCGAAACCCC-3'ITT P148.941844171 P1510P—GTCATAGGGTAGCGTAGCGCCC//—GGAACCCCTAGCGAGTAGCMACGCC P160P—GTCATAGGGTAGGCGTCACC//—GGAACCCCTAGTGACGAGT-3'ITT 		nt11.87	8602031			lint11.87860	175
 P6 Maria - Chandrot Control and Control a		5/TTDCAC		TCCACCC//_		VARGTCCCCC-	-3/770
 Mouse—ggaacatgcctcacgcaaa—//—acaggctccaggagctccicd_Mouse nt2.443125381 frtrm-TCACTGAGGCCGCCCGGGCA—//—CCGGGCGACCAAAGGTCGCC—3'ITT Mouse—cggaagaggGGCCCGGGCA—//—CCGGGCGACCAaggcaacg—Mouse nt15.673028981 fint15.67303114 FF1dP—TAGAATTGAACGGTGCCTA—//—GGTTCCTGGCCTTTTGCTGG—ori Mouse—gtacatttatCCGGTGCCTA—//—GGTTCCTGGCtagactttgg—Mouse nt1.856564841 fint18.85656205 ori—AGTGGCTGCTGCCAGTGGCC—//—TATTTTTCTAAATACATTCA—Amp^r Mouse—ctgggaacagCCCGCGCGC//—TATTTTTCTAAATACATTCA—Amp^r Mouse—ctgggaacagCCCGCGCGC//—TATTTTTCTAAATACATCA—Mouse Mouse—ctgtggaaccagcacctggc—//—tggtccaccactgttctga—Mouse nt8.937442911 fint8.937442911 Mouse—gtcggaaaagGCCGCCGC//—CATGGCGGGTAATCATTAA—ori Mouse—gctggaaaaagCTCGCTCGC//—CATGGCGGGTAATCATTAA—ori Mouse—gctggaaaaagGTGCCCCCCGT—//—TTTAAACATGGaagagctgga—Mouse nt5.1263845691 fint5.126384647 Amp^r—CAGATGGTAGCCCTCCCCGT—//—TTTAACATGGTGCGGACGT—ori Mouse—gaccggaaaagggggcgC//—tactgggggtagagggaadaa nt11.1074215731 fint1.107421568 FF1dP—TTAGGAGGTCGCCCGCACGC//—ATGGCGGGTAATCATTAACO—ori Mouse—tctgcctggagatctggaaadagagaggagcTCC//—ATGGCGGGTTAATCATTAACO—ori Mouse—tctgcctggagatcccgCTCCCCCC//—ATGGCGGGTAATCATTAACO—ori Mouse—tctgcctggagatccgCCCCCCCCG//—ATGGCGGGTAATCATTAACO—ori Mouse—tctgcctggagatcatgaaa//—agaaccattggttttctag—Mouse nt19.319635601 fittP-TCCTCG fittP-TCCTGG fittP-TCCTGG fittP-TCCTGG fittP-TAGGAGGTCCTCCCCG//—ATGGCGGGTAATCATTAACO—ori Mouse—tctgcctggagatcatgaaa//—agaaccattggttttctag—Mouse nt19.319635601 fittP-TCCTGG fittP-TCCTGG fittP-TCCTGG fittP-TCCTGG fittP-319635401 fittP-31963565 fittP-TCCTGG fittP-TCCTGG fittP-31963560 fittP-319635601 fittP-319635607 fittP-31963565 fittP-TCCTGG fittP-TCCTGG fittP-319635	P6	Mouse—gga	acatgccTCC	TG G AGGG //-	-CGGGCGACCA	ggagcttctc-	-Mouse
P7 Mouse-ctgaagaggCGCCCGGGCA-//-CCGGGCGACCAAAGGTCGCC-3'ITT Mouse-cgaagaggCGCCCGGGCA-//-CCGGGCGACCAAAGGTCGCC-3'ITT Mouse-ctgaagaggCGCCCGGGCA-//-CCGGGCGACCAAAGGTCGCC-Mouse Mouse-cgaagaggGGCCCGGGCCA-//-CGGTCCTGGCCTTTGGCTGG-ori Mouse-gtacatttatCCGGTGCCTA-//-GGTTCCTGGCCTTTGGCTGG-ori Mouse-gtacatttatCCGGTGCCTA-//-GGTTCCTGGCCTTTGGCTGG-ori Mouse-gtacatttatCCGGTGCCTA-//-GGTTCCTGGCCTTTGGCTGG-ori Mouse-gtacatttatCCGGTGCCA-//-accccgtggCtagactttgg-Mouse ntl.8565484f) fntl.8565648 p9 ori-AGTGGCTGCTCCCAGTGGCC-//-TATTTTTCTAAATACATTCA-Amp ^r Mouse-ctgtggaccGCCAGTGGCC-//-TATTTTTCTAAATACATTCA-Amp ^r Mouse-ctgtggaccagcacctggc-//-tggtccaccactgtttctga-Mouse nt8.93744291f fnt8.93744503 p10 ?'ITR-CTCTCTGCGCGCCCGCCGC-//-CATGGCGGGTTAATCATTAA-ori Mouse-gctggaaaaaGCTCGCTCGC-//-CATGGCGGGTTAATCATTAA-ori Mouse-gctggaaaaagggggcagc-//-tatcggggtgaggctgga-Mouse nt5.126384569f fnt5.126384847 p11 Mouse-gaaaccctgtctCCGCCGC-//-ATGGCGGGTTAATCATTAA-ori Mouse-gaaaccctgtctCG Mouse-gaaaccctgtctCGCCCCCGC-//-ATGGCGGGTTAATCATTAA-ori Mouse-gaaaccctgtctCGCCCCCCGC-//-ATGGCGGGTTAATCATTAA-Ori Mouse-gaaaccctgtctCGCCCCCCGC-//-ATGGCGGGCTGGCGCAGC-/-ori Mouse-gaaaccctgtctCGCCCCCCGC-//-ATGGCGGGCTAATCATTAAC-ori Mouse-gaaaccctgtctcg nt11.107421573f fnt11.107421563 p11 Mouse-cttgcctggagatcagaaa-//-aggacGCGCGCCGCGCAAAGCCC-3'ITT Mouse-cttgcctggagatCGCCCCCCCGC-//-ATGGCCGGCTAATCATTACC-ori Mouse-cttgcctggagatcagaaa-//-aggacCCCCCCGGCAAAGCCC-3'ITT Mouse-cttgcctggagatcagaaa-//-AGGCCGCCCCGGCAAAGCCC-3'ITT Mouse-cttgcccctACGCGCCCCCCCCCGC-//-ATGGCCGCCCCGCGCAAAGCCCC-3'ITT Mouse-cttgccccAACGGGTGGAGC-//-AGGCCCCCCCGGCAAAGCCC-3'ITT Mouse-cttgccccAACGGCTGCACCC-//-ATGGCCGCCCCGCCGCAAAGCCC-3'ITT Mouse-cttgcccCAACGGCTGGCACC-//-AGGCCCCCCCGGCAAAGCCC-3'ITT Mouse-cttgccccAACGCCCCCCCCCCGCGCAAAGCCCC-3'ITT Mouse-cttgcccCAACGGCTGGACC-//-AGGCCCCCCCGGCAAAGCCC-3'ITT Mouse-cttgcccCAACGGCTGGACC-//-AGGCCCCCCGCGCAAAGCCC-3'ITT Mouse-cttgcccCAACGGCTGGACC-//-AGGCCCCCCGGCAAAGCCC-3'ITT Mouse-cttgcccCAACGGCTGGACC-//-AGGCCCCCCCGCGCGAAAGCCC-3'ITT		Mouse—gga	acatgcc tc a	ca g caaa—//–	-acaggetcca	lggagcttctc-	-Mouse
 P7 5'TTR—TCACTGAGGCGCCCGGGCA—//—CCGGGCGACCAAAGGCGCC—3'TTT Mouse—cggaagaggGGCCCCGGGCA—//—CCGGGCGACCAAGGCGACGAGGGAGGAGGAGGGGGGGGG		1102.11	5125501			TIC2. 115125	51
 Mouse—cgaaqaggGCCCCGGGCA—//—CCGGGCGACCtaagccaacg—Mouse nt15.673028981 P8 Mouse—gtacatttatCCGGTGCCTA—//—GGTTCCTGGCCTTGTCCTGGC_ Mouse—gtacatttatCCGGTGCCTA—//—GGTTCCTGGCCTGTGCCGG_ mouse—gtacatttatCtgCaacacc—//—acccagtggctagactttgg—Mouse nt1.856564841 P9 ori—AGTGGCTGCCAGTGGCG—//—TATTTTTCTAAATACTTCA—Amp^r Mouse—ctgtggacccaGCCAGTGGCG—//—TATTTTTCTAAATACATCA—Amp^r Mouse—ctgtggacccaGCCAGTGGCG—//—TATTTTTCTAAATACATCA—Amp^r Mouse—ctgtggacccaGCCAGTGGCG—//—CATGCGGGGTTAATCATTAA—mouse Mouse—ctgtggaaccagcacctggc—//—tggtccaccactgtttctga—Mouse nt8.937442911 P10 ?'ITR—CTCTCTGCGCGCCCGCCGC—//—CATGGCGGGTTAATCATTAA—ori Mouse—gctggaaaagggggcgcg//—tactggggtagaggctgga—Mouse nt5.1263845691 P11 Amp^r—CAGATGGTAAGCCTCCCCGT—//—TTTAAACATGGTGCGGACGT—orig Mouse—gaacacctgtGCCCTCCCGT—//—TTTAAACATGGTGCGGACGT—orig Mouse—gaacacctgtGCCCTCCCGG—//—ATGGCGGGTTAATCATTAAC—ori Mouse—gaacacctgtGCCCTCCCCGT—//—TTTAAACATGGTGCGGACGT—orig Mouse—gaacacctgtGCCCTCCCGGT—//—TTTAAACATGGTGCGGACGT—orig Mouse—gaaccctgtGCCCTCCCGGT—//—TTTAAACATGGTGCGGACGT—orig Mouse—gaaccctgtGCCCTCCCGGT—//—TTTAAACATGGTGCGGACGT—orig Mouse—ttgcctggaatacctggagattagaa—Mouse nt11.1074215731 P11 Mouse—ttgcctggggatcgaadgcctgga—Mouse nt19.319635801 P13 Mouse—cttgcctgcdgcTCGCACG—//—ATGGCGGCTCACCCGC-//=ATGGCGGCTCACCCCGC-//=ATGGCGGCCCCCCGCGAAAACCC-3'ITT Mouse—cttgcctggaattagaaa—//—agaacctttggcttactag—Mouse nt18.941844171 P13 Mouse—cttgccdcctAGGGTGAGGC—//—AGGCCCCCCGGCAAAACCCC-3'ITT Mouse—cttgccdcctAGGGTGGAGC—//—GGAACCCCCGGCGAAAGCCC-3'ITT P13 Mouse—cttgccdcctAGGGTGAGCC//—ATGCCGCCCCGCGCAAAACCCC-3'ITT P14 Mouse—cttgccdcctAGGGTGAGAC—//—AGGCCCCCCGGCCGAAAACCCC-3'ITT P15 Mouse—cttgccdcctAGGGTGGAGC—//—AGGCCCCCCGGCCAAACCCC-3'ITT P14 Mouse—cttgccdcctAGGGTGAGCC//—ATGGCGCCCCGGCGAAAACCCC-3'ITT P15 Mouse—cttgccdcctAGGGTGAGCC///—AGGCCCCCCGGCCGAAACCCCC-3'ITT P16 Mouse—cttgccdccctAGGGTGGAGCC///—AGGCCCCCCGCGCGAAAGCCCC-3'ITT P	D 7	5'ITR-TCA	CTGAGGCCGC	CCGGGCA//-	-ccgggcgac	AAAGGTCGCC-	-3'ITR
<pre>Note = Cigagagagagagagagagagagagagagagagagagaga</pre>	Ρ/	Mouse-cgg	aagagggCGC	CCGGGCA//-	-CCGGGGCGACC	taagccaacg-	-Mouse
 P8 EF10P TAGAATTGAACCGGTGCCTA // GGTTCCTGGCCTTTTGCTGG ori Mouse gtacatttatCCGTGCCTA // GGTTCCTGGCtagactttgg Mouse Mouse gtacatttatCCGTGCCTA // GGTTCCTGGCtagactttgg Mouse nt1.85554841 fill fill fill fill fill fill fill fil		nt.15.67	3028981	9 49994 9 //	ecceccec	fnt15.67303	114
P8 EF1αP—TAGAATTGAACCGGTGCCTA—//-GGTTCCTGGCCTTGG_OTGG_OTGG Mouse—gtacatttaictgCGGGGCCGA//GGTTCCTGGCGTGTGG_Mouse Mouse—gtacatttaictgGGGGCGC//-acccGggggtagactttgg_Mouse ntl.856564841 P9 Mouse—ctgtggaccGCGCAGTGGCC_//-TATTTTTCTAATACATCA_Amp ^F Mouse—ctgtggaccGCGCAGTGGCG_//-TATTTTTCTAATACATCA_Amp ^F Mouse—ctgtggaccGCGCAGTGGCG_//-TATTTTTCTAATACATTCA_Amp ^F Mouse—ctgtggacacGCCGCGCTGGC_//-CATGGCGGGTTAATCATTAA—ori Mouse—gctggaaaaagggggcagc_//-tactggggtgaggctgga_Mouse nt5.1263845691 Mouse—gctggaaaaagggggcagc_//-tactggggtgaaggctgga_Mouse nt5.1263845691 Mouse—gctggaaaaagggggcagc_//-tactggggtgaaggctgga_Mouse nt1.1074215731 P11 Mouse—gctggaaaaagGTCGCCCCCCGT_//-TTTAACATGGTGCGGCTCOCT Mouse—gctggaaaaagggggcgcagc_//-tactggggtagagctgga_Mouse nt1.1074215731 P11 Mouse—gctgggaatacgGTCGCCCCCG Mouse—tctgcctggagatatgaaa Mouse—tctgcctggagatagacd Mouse—tctgcctggagatagacdgCGCGCCCCC Mouse—tctgcctgacdGTGGCGCCCCG Mouse—tctgcctgacdGTGCGCCCCCG Mouse—tctgcctgcd_ACGGTGGCGC_//-ATGGCCGCCCGGCAAAGCCC_3'ITT Mouse—cttgccctlaCGGGTGGACG Mouse—cttgcctgccdAGGGTGGCGCCCCCCGGC			**		•	▼	
 Mouse glacattraitergaaaacc // Garceagtggetsgactragactrigg Mouse ntl.85656205 P9 ori AGTGGCTGCCCAGTGGCG // TATTTTTCTAAATACATCA Amp^r Mouse ctgtggaaccaGCCAGTGGCG // TATTTTTCTAAATACGTCA Amp^r Mouse ctgtggaaccagcacctggc // tggtccaccactgttctga Mouse nt8.937442911 P10 ?'ITR CTCTCTGCGCGCCCGCCGC // CATGGCGGGTAATCATTAA ori Mouse gtggaaaaaGCTCGCTCGC // CATGGCGGGTAATCATTAA ori Mouse gtggaaaaagggggagc // tatggggtagaggtgga Mouse nt5.1263845691 P11 Amp^r CAGATGGTAAGCCTCCCGT // CATGGCGGGTAATCATTAA ori Mouse gaaaccctgtGCCTCCCGT // CATGGCGGGTGAGagcaga nt5.1263845691 P11 Amp^r CAGATGGTAAGCCTCCCGT // TTTAAACATGGTGCGGACGT ori Mouse gaaaccctgtGCCTCCCGT // TTTAAACATGGTGCGGACGT ori Mouse gaaaccctgtGCCTCCCGT // TTTAAACATGGTGCGGACGT ori Mouse gaaaccctgtGCCTCCCGCACG // ATGGCGGGTTAATCATTAAC ori nt11.1074215731 P11 Mouse tgctggaggagataggagagagagagagagagagagagag	P8	EF1αP-TAG	AATTGAACCG	GTGCCTA-//-	-GGTTCCTGG	CTTTTGCTGG-	-ori
<pre>ntl.856564841</pre>		Mouse—gta	cattatetg	caacacc-//-	-acccagtgg	tagactt tgg-	-Mouse
P9 ori —AGTGGCTGCTGCCAGTGGCG—// —TATTTTTCTAAATACATTCA—Amp ^r Mouse—ctgtggaccGCCAGTGGCG—// —TATTTTTCTActgtttctga—Mouse nt8.93744291 Mouse Int8.93744291 P10 ?'ITR—CTCTCTGCGCGCTCGCC// —CATGGCGGGTTAATCATTAA—ori Mouse—gctggaaaaaGCTCGCTCGC // —CATGGCGGGTTAATCATTAA—ori Mouse—gctggaaaaaGCTCGCTCGC // —CATGGCGGGGTAATCATTAA—ori Mouse—gctggaaaaaGCTCGCTCGC // —CATGGCGGGGTAATCATTAA—ori Mouse—gctggaaaaaGCTCGCTCCCCT // —TTTAAACTGGTGCGGACGT—ori Int5.126384569 P11 Amp ^r —CAGATGGTAAGCCCTCCCGT // —TTTAAACTGGTGCGGACGT—ori Mouse—gaaccctgtctcg Int5.126384847 P11 Amp ^r —CAGATGGTAAGCCCTCCCGT // —TTTAAACTGGTGCGGACGT—ori Mouse—gaaccctgtctcg Int10.107421568 P12 EF1αP—TTAGGGAGGTCGCCGCACG // —ATGGCGGGGTTGgtttctctag—Mouse Mouse—tctgcctggagatcatgaaa Mouse Mouse—tctgcctggagatcatgaaa P13 S'ITR—TCCTGG AGGGGTGGAG AGGCCGCCCCGGCAAAGCCC—3'ITT Mouse—cttgcccclAGGGTGGAG Mouse Mouse—cttgcc actaagcctaac P13 Mouse—cttgcccclAGGGTGGAGG MagGCCGCCCCGGCAAAGCCC—3'ITT Mouse—cttgcc Mouse ACTAagCCCCCCCCCCCGGCAAAGCCC—3'ITT Mouse—cttgcc Mouse ACTAagCCCCCCCCCCCGGCAAAGCCC—3'ITT Mouse—cttgcc Mouse ACTAagCCCCCCCCCCCCCGGCAAAGCCC—3'ITT Mouse—cttgcc Mouse ACTAagCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		nt1.85	6564841		_	l[nt1.856562	05
P9 Noise - ctgtggaccdCCATGGCC -// TATTTTTCTActgttttiga - Moise Noise - ctgtggaccdgcacctggc -// -tggtccacctgtttttga - Moise nt8.93744291 fttt - fttt - Moise nt8.93744291 fttt - fttt - Moise Noise - gctggaaaaaGCTCGCTCGC -// -CATGGCGGGTTAATCATTAA - ori Moise - gctggaaaaagggggcagc -// - CATGGCGGGTTaATCATTAA - ori Moise - gctggaaaaagggggcagc -// - CATGGCGGGTTaATCATTAA - ori Moise - gctggaaaaagggggcagc -// - CATGGCGGGTTaATCATTAA - ori Moise - gctaaacctgfcCCCCCCCGT -// - TTTAACATGGTGCGGCACCT - ori Moise - gaaaccctgfcCCCCCCCGT -// - ATGGCGGGTTAATCATTAC - ori Moise - tctgcctggaGCTCCGCACG -// - ATGGCGGCTTAATCATTAAC - ori Moise - tctgcctggaGCTCCGCACG -// - ATGGCGGCTAATCATTAC - ori Moise - tctgcctggaGCTCCGCACG -// - ATGGCGGCGCAAGCCC - 3/ ITT P12 P13 Moise - tctgcctggaGCTCCGCACG -// - ATGGCCGCCCCGGCAAGCCC - 3/ ITT Moise - cttgcctgaGGTGGAGG -// - AGGCCGCCCCGGCAAGCCC - 3/ ITT Moise - cttgccctlAGGGTGGAGG -// - AGGCCGCCCCGGCAAGCCC - 3/ ITT Moise - cttgcccctlAGGGTGGAGG -// - AGGCCGCCCCGGCAAGCCC - 3/ ITT P13 Moise - cttgcccctlAGGGTGGAGG -// - AGGCCGCCCCGGCAAGCCC - 3/ ITT Moise - cttgcccctlAGGGTGGAGGC -// - AGGCCGCCCCGGCAAGCCC - 3/ ITT P13 Moise - cttgcccctlAGGGTGGAGG -// - AGGCCGCCCCGGCAAGCCC - 3/ ITT P14 Mint8.94184420		ori-AGT	CACTACTAC	AGTGCC-//-		AATACATTCA-	-Amp ^r
<pre>Mouse_ctgtggacccagcacctggc_//_tggtccaccactgtttctga_Mouse nt8.937442911</pre>	P9	Mouse-ctg	tggacccGCC	AGTGGCG-//-	-TATTTTTCTZ	Actgtttctga-	-Mouse
 P10 Mouse—ctgctggaaaaaGCTCGCTCGC—//—CATGGCGGGTTAATCATTAA—ori Mouse—gctggaaaaaGCTCGCTCGC—//—CATGGCGGGTAATCATAA—ori Mouse—gctggaaaaaGCTCGCTCGC—//—CTGGCGGGGTAATCATAA—ori Mouse—gcaaaccctgfGCCCTCCCGT—//—TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgfGCCCTCCCGT—//—TTTAAACATGGTGCGGACGAA—Mouse Mouse—gaaaccctgfGCCCTCCCGT—//—TTTAAACATGGTGCGGACGAA—Mouse mt11.1074215731 P11 Mouse—ctgcctggaGTCGCCGCACG—//—ATGGCGGCGTAATCATTAAC—ori Mouse—tcgcctggaGTCGCCGCACG—//—ATGGCGGCGTAATCATTAAC—ori Mouse—tcgcctggaGTCGTCCGCACG—//—ATGGCGGCGTAATCATTAAC—ori Mouse—tcgcctggaGTCGCGCACG—//—ATGGCGGCGTAATCATTAAC—ori Mouse—tcgcctggaGTCGCGCACG—//—ATGGCGGCGCTAATCATTAAC—ori Mouse—tcgcctggaGTCGCGCACG—//—AGGCCGCCCGGGCAAAGCCC—3'ITT Mouse—ctgccctlACGGTGGAGG—//—AGGCCGCCCGGGCAAAGCCC—3'ITT Mouse—ctgccctlACGGGTGGAGG—//—AGGCCGCCCGGaagcctaac—Mouse nt8.941844171 EF10P—GTCATAGGGTAGGTAGGGTGAGGTC—//—GGAACCCCTAGTGATGGAGT—3'ITT 		Mouse-ctg	tggacccagc	acctggc//-	-tggtccacca	1ctgtttc t ga-	-Mouse
P10 ?'ITR—CTCTCTGGGGCTCGCTCGC—//—CATGGCGGGTTAATCATAA—ori Mouse—gctggaaaaagggggcagc—//-tactgggggtagaggctgga—Mouse nt5.1263845691 Int5.126384847 P11 Mouse—gctggaaaaccgtgCCCCCCGT—//-TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtCCCCCCGT—//-TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtCCCCCCCGT—//-TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtCCCCCCCGT—//-TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtCCCCCCCGT—//-TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtGCCCCCCCGT Int11.107421568 P12 BF10P—TTAGGGAGGTCGCCCCCACG—//-ATGGCGGGTTGgttttctag—Mouse Mouse—tctgcctggagatcatgaaa magaacctttggttttctag—Mouse Mouse—tctgcctggagatcatgaaa magaacctttggttttctag Mouse—ttgcccggagatcatgaaa magaacctttggttttctag Mouse—ttgcccgclAGGGTGGAG—//-AGGCCGCCCCGGGCAAAGCCC—3'ITT Mouse—cttgcccclAGGGTGGAGG Mouse—cttgccccacd Mouse—cttgcc actagcctaac Mouse nt8.941844171 Int8.94184420 Int8.94184420 EF10P—GTCATAGGGTAGGTGAGGTGAGCC—//-GGAACCCCCAGTAGGAGT—3'ITT Int8.94184420		1100.95				▼	105
P11 Mouse—gctggaaaaagggggcagc—//—CATGGCGGGTagaggctgga=Mouse nt5.126384569¶ Int5.126384847 P11 Mouse—gctggaaaaagggggcagc—//—TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtGCCCCCCGT—//—TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtGCCCCCCGT—//—TTTAAACATGGTGCGGACGT—ori Mouse—gaaaccctgtGCCCCCCGT—//—TTTAAACATGGTGCGGACGT—ori Int11.107421573¶ P12 Mouse—tctgcctggaGGTCCGCACG—//—ATGGCGGGTTAATCATTAAC—ori Mouse—tctgcctggaGTCCGCACG—//—ATGGCGGGTTAATCATTAAC—ori Mouse—tctgcctggagatagaaa—//—agaaccattggttttctag—Mouse nt19.31963540¶ P13 Mouse—cttgccccdgagagatagaaa—//—agacccccCGGCAAAGCCC—3'ITT Mouse—cttgcccdAGGGTGGAGG—//—AGGCCGCCCGGCAAAGCCC—3'ITT Mouse—cttgcccdAGGGTGGAGG—//—AGGCCGCCCGGagagctaac—Mouse nt8.94184417¶ P14 Mouse—cttgcccaac	D10	?'ITR-CTC	TCTGCGCGCT	CGCTCGC-//-	-CATGGCGGG	TAATCATTAA-	-ori
<pre>nt5.126384569ff fnt5.126384847 P11 Amp^r CAGATGGTAAGCCTCCCGT // TTTAAACATGGTGCGGACGT ori Mouse gaaaccctgtGCCTCCCGT // TTTAAACATGGTGCGGACGAA Mouse nt1.107421573ff flnt1.107421568 P12 Mouse tcgcctgaGGTCGTCCGCACG // ATGGCGGGTTAATCATTAAC ori Mouse tcgcctggagatatgaaa // agaaccattggttttctag Mouse nt19.31963580ff P13 Mouse ttgcctgcaCGTCGGACG // ATGGCCGCCCGGCAAACCC -3 'ITT Mouse ttgcctggagatatgaaa // agaaccattggttttctag Mouse nt19.31963580 P13 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGCAAACCC -3 'ITT P13 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGCAAACCC -3 'ITT P14 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGCAAACCC -3 'ITT P15 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGCAAACCC -3 'ITT P16 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGCAAACCC -3 'ITT P17 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGCAAACCC -3 'ITT P18 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGCAAACCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAAACCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAAACCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAACCCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAACCCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAACCCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAACCCC -3 'ITT P19 Mouse cttgcc ac acGGGTGAACCCCCGGGCCGAaCCCCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAACCCCC -3 'ITT P19 Mouse cttgcc ac acGGGTGGAC // AGGCCGCCCGGGCAACCCCCA3C // AGGCCGCCCGGACCCCGACCCCGCGCGACCCCGACCCCGACCCCGCCG</pre>	FIU	Mouse-gct	ggaaaaaaGCT	CGCTCGC-//-	-CATGGCGGGI	ag a ggotgg a- ag a ggotgg a-	-Mouse
P11 Amp ⁻ —CAGATGGTAAGCCCTCCCGT—//—TTTAAACATGGTGCGGACGT—ori Mouse—gaaacctgtGCCCTCCCGT—//—TTTAAACATGaaaaaccaaa—Mouse ntl1.1074215731 P12 EFlop—TTAGGGAGGTCCTCCGCACG—//—ATGGCGGGTTAATCATTAAC—ori Mouse—tcgcctggaGCGCCGCACG—//—ATGGCGGGTTgtttttctag—Mouse nt19.319635401 P13 S'ITR—TCCTGG AGGGGTGGAG—//—AGGCCGCCCGGGCAAAGCCC—3'ITT Mouse—tcgcctgaGGTGGAGG—//—AGGCCGCCCGGGCAAAGCCC—3'ITT Mouse—tcgcctgaGGTGGAGG—//—AGGCCGCCCGGGCAAAGCCC—3'ITT Mouse—tcgcctgaCGCTGAGGGTGGAG—//—AGGCCGCCCGGGCAAGCCC—3'ITT Mouse—tcgcctlaCGGTGGAGG—//—AGGCCGCCCGGacAaGCCCG—3'ITT Mouse—tcgcctlaCGGTGGAGG—//—AGGCCCCCCGGacAaGCCGGAGGMG nt8.941844171 P13 EFlop—GTCATAGGGTTAGGGATGAGGCC—//—GGACCCCTAGTGATGAGGT_3'ITT		nt5.126	3845691			fnt5.126384	847
P11 Amp — CALATGTAAGCCCTCCCGT —// — TTTAACATGGTGCGAGCTGCGTGCGTGGAGAGACCGTAGTGTAGCCGTAGCGGTGT// — TTTAACATGGTGGGGAGAGAGAAGCGGAGAMOUSE MOUSE — gaaaaccctgtctcg dtggaaaacccaaa — Mouse MOUSE — gaaaaccctgtctcg dtggaaaacccaaa — Mouse Mouse — gaaaccctgtctcg dtggaaaacccaaa — Mouse ntll.107421573 mlntll.107421568 P12 Mouse — tctgcctggaCGTCCCCCACC —// — ATGGCGGCGTAATCATTAC— > ori Mouse — tctgcctggaCGTCCGCACC —// — ATGGCGGCGTGgttttctag — Mouse ntl9.319635401 ntl9.319635401 Intl9.31963585 5'ITR — TCCTGG _ AGGGGTGGAG —// — AGGCCGCCCCGGGCAAAGCCC — 3'ITT Mouse — cttgcccctlAGGGGTGGAG —// — AGGCCGCCCCGGacAAGCCC — 3'ITT Mouse — cttgcc _ acctlaGGGTGGAG —// — AGGCCGCCCCGGacAAGCCC — 3'ITT Mouse — cttgcc _ acctlaGGGTGGAG —// — AGGCCGCCCCGGacAAGCCC — 3'ITT Mouse — cttgcc _ acctlaGGGTGGAG —// — AGGCCCCCCGGacAad — Mouse nt8.941844171 Int8.94184420 EF1αP — GTCATAGGGTAGGAGTAGCC — // — GGAACCCCTAGTGAGGTGAGT — 3'ITT		1	V		•		
Mouse Gtcgaaaaccccdjctcg Mouse Mouse Mntll.107421573 P12 Mouse TTAGGGGGGTCGCCGCACG Matter and the second	P11	Mouse-gaa	accctqtGCC	CTCCCGT-//- CTCCCGT-//-	-TTTAAACATC	GTGCGGACGT- aaaaaccaaa-	-Ori -Mouse
<pre>ntll.10/4215/3 ntll.10/421568 P12 Mouse—tctgcctgaGGTCGCGCACG—//—ATGGCGGGTTAATCATTAAC—ori Mouse—tctgcctggaGGTCCGCACG—//—ATGGCGGGTTAGTGLttctag—Mouse ntl9.31963540 mtll9.31963585 5/1TR—TCCTGC AGGGTGGAG—//—AGGCCGCCGGGCAAAGCCC—3/ITT Mouse—cttgcctaCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG</pre>		Mouse-gaa	accctgtctc	<u>g</u>	ctcg	aaaaaccaaa-	-Mouse
P12 EF1αP—TTAGGGAGGTCGTCCGCACG—//—ATGGCGGGTTAATCATTAAC—ori Mouse—tctgcctggaGGTCGCGCACG—//—ATGGCGGGGTggttttctag—Mouse nt19.319635401 Mouse Int19.31963585 P13 5'ITR—TCCTGG AGGGGTGGAG—//—AGGCCGCCCGGGCAAAGCCC—3'ITF Mouse—cttgccctlAGGGTGGAG—//—AGGCCGCCCGGGCAAAGCCC—3'ITF Mouse—cttgcc Mouse nt8.941844171 EF1αP—GTCATAGGGTAGGGAGGTC—//—GGAACCCCTAGTGATGGAGT—3'ITF		nt11.107	4215731		-	∥nt11.10742	1568
<pre>P12 Mouse—tctgcctggaCGTCGCACG—//—ATGGCGGGTTggttttctag—Mouse Mouse—tctgcctggagacatgaaa—//—agaacatttggttttctag—Mouse nt19.319635401 Int19.31963585 5'ITR—TCCTGG_AGGGGGGGGG_//—AGGCCGCCCGGGCAAAGCCC_3'ITT Mouse—cttgccctlAGGGTGGAG—//—AGGCCGCCCGGagcctaac—Mouse nt8.941844171 Int8.94184420</pre>	540	ΕΓΙ αΡΤΤΑ	GGGAGGTCGT	CCGCACG-//-	-ATGGCGGGT	AATCATTAAC-	-ori
<pre>nousectrgcctrggggtctargaaa // agaacectrggtctcagMouse nt19.319635401</pre>	P12	Mouse-tct	gcctggaCGT	CCGCACG-//-	-ATGGCGGGT	ggttttctag-	-Mouse
5'ITR—TCCTGC AGGGGTGGAC—//—AGGCCGCCGGGCAAAGCCC3'ITF P13 Mouse—cttgcdccctlaGGGGTGGAC—//—AGGCCGCCCGCtaagcctaac Mouse—cttgcc actaagcctaac Mouse—cttgcc actaagcctaac Mouse—cttgcc actaagcctaac Mouse—cttgcc actaagcctaac Mouse—cttgcc actaagcctaac Mouse—cttgcc actaagcctaac Mouse actaagcctaac		nt19.31	.9635401	catyada //-	ayaaccatti	fnt19.31963	585
5'ITR—TCCTGG_AGGGGGGGGGGG—//—AGGCCCCGGCCAAGCCC_3'ITI P13 Mouse—cttgcc_cctAGGGGGGGGGGGGG/_/_AGGCCCGCCGGCCAAGCCCGaccAAGCCC_3'ITI Mouse—cttgcc_ac _accaccacc_mouse nt8.941844171 Înt8.94184420 EF1αP—GTCATAGGGTAGGGGGTAGGGGGTC—//_GGAACCCCTAGTGATGGAGT_3'ITI						v	
Mouse Ctrgcc Ctrgcc Address Mouse nt8.941844171 Int8.94184420 Int8.94184420 EF1ap GGAACCCCTAGTGATGGAGT 'Int8.94184420	D13	5'ITR-TCC	TGG AGG	GGTGGAG-//-	-AGGCCGCCCC	GGCAAAGCCC-	-3'ITR
nt8.94184417 înt8.94184420 EF10P-GTCATAGGGTTAGGGAGGTC-//-GGAACCCCTAGTGATGGAGT-3'ITT	113	Mouse—ctt	.gcc <u>ecu</u> AGG	GG1GGAG-//-	-AGGCCGCCCG	taageetaa e -	-Mouse
EF1αP—GTCATAGGGTTAGGGAGGTC—//—GGAACCCCTAGTGATGGAGT—3'ITH		nt8.941844	171			lint8.941844	20
EFINE-GICAIAGGGIIAGGGAGGIC-//-GGAACCCCIAGIGAIGGAGI5'III				CCACCTC_ //		CTCATCCACT	-3/ 7772
P14 Mouse—ttaaaaattaTAGGGAGGTC—//—GGAACCCCTAtttttattta	P14	Mouse—tta	aaaattaTAG	GGAGGTC-//-	-GGAACCCCT	tttttattta-	-Mouse
Mouse—ttaaaaatta <u>ttta</u> — <u>ttta</u> ttttattta—Mouse nt4.1474951341 Int4 147495139		Mouse—tta nt4 147	aaaatta <u>ttt</u> 4951341	<u>a</u>	<u>ttt</u> a	nt4,147495	-Mouse

The bioinformatic information obtained from our previous study allowed for characterization of only one side of the vectorcellular DNA junctions owing to the nature of the plasmid rescue strategy (Fig. 1b). Obtaining the whole proviral vector genome together with both the 5' and 3' junction sequences is particularly important to establish chromosomal deletions or rearrangements associated with rAAV2 integration in animals. To this end, we used an in vivo hepatocyte selection system based on a hereditary tyrosinemia type I (HTI) mouse model. HTI is an inherited fatal metabolic hepatorenal disease caused by deficiency of fumarylacetoacetate hydrolase (FAH; ref. 13). Because there is a selective repopulation of stable genetically modified FAH-positive hepatocytes in HTI mouse livers¹⁴, it allows for *in vivo* clonal selection of hepatocytes with integrated rAAV2 vector genomes expressing FAH⁵, resulting in dilution of all the extrachromosomal circular rAAV2 vector genomes by cell division. With injection of a human FAH-expressing rAAV2 vector (Fig. 1a) into HTI mouse livers followed by in vivo selection, we isolated a total of 14 whole proviral vector genomes with both junction sequences, using a plasmid rescue strategy (Fig. 1b). Detailed restriction enzyme mapping and sequencing of proviral genomes identified complicated structures in three cases (Fig. 1c) and frequent partial and sometimes complete deletions of the viral ITRs (Table 1). By aligning the vector sequences obtained from our past and present studies with the sequences of the integration sites obtained from the mouse genome database, we characterized each integration event (Fig. 2 and Table 1). We identified deletions of cellular genomes at all the integration sites (14 of 14 integrations, 100%), ranging from 2 bp to ~0.3 kb in most cases. A deletion of 2.1 kb occurred in an integration event, but there was no large rearrangement or translocation of chromosomes. We observed nucleotide insertions of 1-4 bp of unknown origin in three cases. We found no important homologies between vector and cellular DNA but frequently found patch homologies up to 4 bp around the integration sites. The number of nucleotides shared by both vector and cellular DNA sequences within the 10-bp region inside from the breakpoints was significantly higher than the expected number calculated from a random model (a two-tailed binomial test, P = 0.0002), confirming that rAAV2 integration was influenced by microhomology. The G/C contents in a 200-bp window around each breakpoint ranged from 33% to 61% (average 48%, s.d. 7%), showing no general trend compared with the mean G/C content of the mouse genome (42%; ref. 12).

With the information from a total of 29 integration sites identified in rAAV2-injected mouse livers without selection⁴ and *in vivo*-selected rAAV2-transduced hepatocytes, we began to elucidate how rAAV2 vectors might select target sites for integration in genetically stable somatic cells in animals. The results are summarized in **Table 1**.

Figure 2 Sequences of rAAV2 vector-mouse cellular DNA junctions. Sequences of rAAV2 vector (top line), provirus with flanking mouse genomic DNA (middle line) and the mouse genome (bottom line) around 5' and 3' vector-cellular DNA junctions are aligned. Red upper-case letters represent vector genome and green lower-case letters represent mouse genome. The locations of each junction are indicated with an arrowhead and nucleotide positions in the mouse genome (nt; the first number indicates a chromosome number, followed by a nucleotide position number obtained from the NCBI database). Blue letters indicate nucleotides shared by vector and mouse genomic sequences. Nucleotides in a box in P3 and P13 show a nucleotide insertion at the junction. Two arrowheads at a junction indicate that the breakpoint should be located between the arrowheads but the exact location cannot be determined because of microhomology. Underlined sequences show an overlap between 5' and 3' junctions. The origin of the left ITR of P10 was not determined.

Table 1 Structures of rAAV2 provinal genomes and host chromosomal eff	Table 1	Structures of rAAV	2 proviral gei	nomes and host	chromosomal	effects
---	---------	--------------------	----------------	----------------	-------------	---------

ID	Selection	Provirus			Chromosome			Microhomology ^d	G/C content (%) at	Target
		Structure	Deletic 5' junction	on (bp) ^b 3' junction	Number/ band	Deletion (bp)	Insertion ^c (bp)		breakpoints ^e 5' 3'	
P1	Yes	Monomer	Δ173	Δ129 (ITR)	7 / F3	Δ13	0	8 / 20	45 47	Gene, intron, reversed
P2	Yes	Monomer	Δ221	Δ230	4 / C6	Δ21	0	7 / 20	52 53	Intergenic
P3	Yes	Monomer	$\Delta 77$ (ITR, flop)	$\Delta 71$ (ITR, flop)	9/C	Δ119	+2	7 / 20	55 52	Gene, intron, reversed
P4	Yes	Monomer	Δ194	Δ134 (ITR)	4 / A5	Δ2107	0	6 / 20	39 38	2 genes:(i) complete deletion, forward; (ii) intron & exon, reversed
P5	Yes	Monomer	Δ211	∆72 (ITR, flip)	11/C	Δ27	0	9 / 20	48 48	Gene, intron, reversed
P6	Yes	Monomer	$\Delta 141$ (ITR)	$\Delta74$ (ITR, flop)	2/C1	$\Delta 58$	0	7 / 20	50 51	Gene, intron, forward
P7	Yes	Monomer	$\Delta41$ (ITR, flop)	$\Delta 75$ (ITR, flop)	15 / D3	Δ215	0	7 / 20	61 51	Intergenic (hit a Uni Gene cluster ⁱ)
P8	Yes	Monomer	Δ225	Δ241	1/C5	Δ278	0	7 / 20	46 48	Gene, intron, reversed
P9	Yes	Complicated	Δ4399	Δ1972	8 / C5	Δ211	0	7/ 20	51 49	Intergenic (hit a Uni Gene cluster ⁱ)
P10	Yes	Complicated	$\Delta 21$ (ITR, flip) ^f	Δ160	5 / F	Δ277	0	9 / 20	49 57	Gene, intron, forward/reversed
P11	Yes	Complicated	∆3882	Δ190	11/E1	Δ4	0	3/8	43 44	Gene, intron, forward
P12	Yes	Monomer	Δ192	Δ159	19/C3	Δ44	0	7 / 20	38 35	Intergenic
P13	Yes	Monomer	Δ147	∆78 (ITR, flip)	8/C5	Δ2	+4	1/4	49 49	Intergenic
P14	Yes	Monomer	Δ183	Δ134 (ITR)	4 / E2	Δ4	0	3/ 8	33 33	Gene, intron, forward
J16	No	ND		$\Delta 141 (ITR)^k$	11/B1		(0)	4/10	43	Gene, intron, forward
J104	No	Complicated ^k	$\Delta 557^{k}$		11/A1		(0)	1/10	54	Gene, intron, forward/reversed
J121	No	ND		$\Delta 114 \; (ITR)^k$	10/C1		(0) ^k	2/10	57	Gene, intron, reversed ^k
J134	No	ND		$\Delta 76$ (ITR, flop) ^k	6 / E2		(0)	4/10	39	Intergenic
J166	No	ND		∆135 (ITR) ^k	14/E4		(0)	6/10	39	Gene, exon, reversed
J175	No	Complicated ^k		$\Delta 2$ (ITR, flip) ^k	4/E2		(0)	3/10	55	Gene, intron, forward
J192	No	ND		$\Delta 78$ (ITR, flop) ^k	NA ^g		(0) ^k	2/10	51	Gene, transcribed region, reversed ^k
J216	No	ND		$\Delta 118 \ (ITR)^k$	6/C3		(+1)	3/10	50	Intergenic
J236	No	ND		$\Delta 124 \ (ITR)^k$	11/E1		(0)	3/10	49	Gene, intron, reversed
J270	No	ND		$\Delta 175^{k}$	7 / A3		(0)	5/10	54	Gene, exon, forward
J278	No	ND		$\Delta 77$ (ITR, flip) ^k	NA ^g		(0)	3/10	50	Intergenic
J288	No	ND		$\Delta74$ (ITR, flop) ^k	4/B1		(0)	3/10	36	Gene, intron, reversed
J299	No	ND		$\Delta 106 (ITR)^k$	14/B		(0)	3/10	52	Gene, intron, forward
J305	No	ND		$\Delta 107 \ (ITR)^k$	15/E1		(0)	4/10	54	Gene, intron, reversed
J313	No	ND		$\Delta 106 \ (ITR)^k$	2 / F3		(0)	2/10	55	Gene, exon, reversed
Total (r	nean ± s.d.))						136 / 400 ^h	(48 ± 75	 Selection (Yes), hit genes: 9/14^j
										Selection (No), hit genes: 12/15 ^j

^aProviral genomes were isolated from in vivo selected HTI mouse hepatocytes (Yes) or from C57BL/6 mouse hepatocytes without selection (No). ^bThe number of nucleotides that were deleted at each end of the vector genome is shown. (ITR, flop) and (ITR, flip) indicate that a portion of the ITR sequence remained at the junction and was identified with flip or flop orientation. (ITR) indicates that a portion of the ITR sequence remained, but the ITR orientation could not be determined because the length of the ITR remnant was less than 64 bp. CThe number in parentheses represents nucleotide insertion at only one junction; therefore, it does not necessarily mean the actual amount of nucleotide insertion at each integration site. ^dThe number of nucleotides that were shared by both vector and cellular DNA sequences within a 10bp stretch inside of each breakpoint (the deleted side of the cellular sequences) was counted. When information for both 5' and 3' junctions were available, we combined them. eG/C contents of a 200-bp window around 5' and 3' breakpoints of cellular genomes are shown. This may be another 3' junction, undetermined because of complicated head-to-head provirus structure. ^BNA, not applicable because of the sequence redundancy of the target region in the mouse genome. J192 and J278 targeted the 45s pre rRNA gene and its intergenic spacer, respectively. ^hThe number of shared nucleotides is higher than expected with a statistical significance (a two-tailed binomial test, *P* = 0.0002). When rAAV2 integration targeted an EST that belonged to a UniGene cluster but was not identified as a gene by either the NCBI Map Viewer or the Ensembl browsers, we separately described it in parentheses. iThe P values against a random integration model calculated by a two-tailed binomial test are P = 0.0001 and P = 0.002 under non-selective and selective conditions, respectively, with a predicted probability of hitting a gene as 0.25, and P = 0.003 and P = 0.1 under non-selective and selective conditions, respectively, with a probability of hitting a gene as 0.41. A comparison of frequency of hitting a gene by χ^2 test and Fisher's exact probability test showed no statistical difference between non-selective and selective conditions ($\chi^2 = 1.451 < \chi_1^2(0.05) = 3.841$; Fisher's two-tailed probability, P = 0.385 > 0.05). ^{kT}hese data have been published previously⁴. Three of 18 junction sequences in the previous study⁴ did not match with sequences in the mouse genome database. ND, not determined.

Total, hit genes: 21/29

Although the integration sites seemed to be distributed on mouse chromosomes with no significant bias, rAAV2 integration was favored in genes regardless of whether they were analyzed in normal liver or after selective repopulation *in vivo* (12 of 15 (80%) under a non selective condition; 9 of 14 (64%) under a selective condition; 21 of 29 (72%) overall frequency). This bias was statistically significant (a two-tailed binomial test, P = 0.0000001 and P = 0.0006, with a predicted probability of hitting a gene as 0.25 and 0.41, respectively; see **Supplementary Note** online for details). Both exons and introns were disrupted with rAAV2 integrations, with a higher incidence in introns.

There seemed to be no bias for orientation of rAAV2 proviral genomes relative to gene transcription. Notably, web-based public data-bases^{15,16} and our RT–PCR analysis confirmed that 20 of 20 (100%) target genes that we analyzed were expressed in the liver, and expression of approximately half of these genes was upregulated in the liver (**Table 2** and **Fig. 3**).

Although the number of integration events analyzed was relatively small and the results may be somewhat biased by the procedures for provirus isolation, our study showed that rAAV2 preferentially integrated into active genes when delivered directly

Table 2 Expression of rAAV2-targeted genes in the liver	
---	--

ID	Targeted gene	Expression in the liver							
	Name	UniGene ID	Web-based database		RT_PCR ^d			Summarv ^e	
		(other IDs ^a)	Liver as READ a cDNA source		Gene Expression Atlas ^c (media	BALB/c	C57BL/6	HTI	Cummary
P1	RIKEN cDNA 2410027J01	Mm.26928	Yes	NI	NI	3.92	3.95	4.24	Expressed (Up)
P3	LOC208011	(XM164986)	NI	NI	NI				ND
P4	RIKEN cDNA 2010003002	Mm. 1103	Yes	0.345	548 (~740))			Expressed
P4	Expressed sequence AW105885	Mm.39006	Yes	NI	NI	0.88	0.92	0.80	Expressed
P5	RIKEN cDNA 1200011M11	Mm.23257	NI	0.654	NI	0.39	0.43	0.56	Expressed
P6	Kynureninase (L-kynurenine hydrolase)	Mm.105278	Yes	NI	NI	Upregulated	Upregulated	Upregulated	Expressed (Up)
P8	DNA segment, Chr 1, ERATO Doi 757, expressed	Mm.27888	NI	0.616 1.261	NI				Expressed (Up)
P10	Epimorphin	Mm.3003	Yes	NI	20 (20)				Expressed
P11	Testis expressed gene 2	Mm.245663	Yes	0.429 1.260	NI				Expressed (Up)
P14	RIKEN cDNA 1300002F13	Mm.21679	Yes	NI	3985 (~200))			Expressed (Up)
J16	UDP-N-acetyl-alpha- D-galactosamine:polypeptide N-acetylgalactosaminyltransfera:	Mm.11360 se 9	NI	NI	74 (~80)	0.31	0.20	0.53	Expressed
J104	Oxoglutarate dehydrogenase (lipoamide)	Mm.30074	Yes	0.413	707.5 (~1	100)			Expressed
J121	Procollagen, type XVIII, alpha 1	Mm.4352	Yes	0.602	2574.5 (20) 691.5 (20))			Expressed (Up)
J166	Similar to Eukaryotic translation initiation factor 4B (eIF-4B)	(XM139255)	NI	NI	NI				ND
J175	Period homolog 3 (Drosophila)	Mm.10723	NI	NI	1641.5 (~95	50)			Expressed
J192	45s pre rRNA, 28s rRNA transcribed region	(X82564)	Known	to be univ	versally expres	ssed			Expressed
J236	ATP-binding cassette, sub-family A (ABC1), member 8	Mm.138955 a	Yes	1.669	NI				Expressed (Up)
J270	RIKEN cDNA 1700023M09	Mm.41511	Yes	0.277	213.5 (~25	50)			Expressed
J288	Aldolase 2, B isoform	Mm.218862	Yes	0.258	5634 (~400))			Expressed (Up)
J299	RIKEN cDNA A130034K24	Mm.212365	NI	NI	306 (~450))			Expressed
J305	RIKEN cDNA 1810044A24	Mm.31995	NI	0.488	NI				Expressed
J313	Inosine triphosphatase (nucleoside triphosphate pyropho	Mm.21399 osphatase)	Yes	-0.265	133.5 (~19	90)			Expressed

^aIn the case where rAAV2-targeted transcripts do not belong to UniGene clusters, other available IDs are listed in parentheses: XM164986 and XM139255 (RefSeq) and X82564 (GenBank accession number). ^bSee ref. 15. Values are log-transformed ratios. Negative values indicate downregulated and positive values indicate upregulated expression, compared to a reference RNA sample (mRNA from E17.5 embryos). ^cGene Expression Atlas of Genomics Institute of the Novartis Research Foundation¹⁶. Values are average differences. Median values are indicated in parentheses. ^dValues are ratios of the amount of RT–PCR product from a target gene to that from *Gapd*, normalized with the ratio from a universal reference RNA. Values <1 and 1> represent downregulated and upregulated expression of the target gene, respectively, compared to the reference RNA sample normalized with *Gapd* expression. In one case, upregulation of a target gene transcript was obvious without coamplification with *Gapd* transcript, indicated as 'Upregulated'. ^eTarget genes that are apparently upregulated in the liver are indicated with 'Up'.

Two target UniGene clusters that NCBI Map Viewer or the Ensembl browser do not consider as genes (Mm.12505 in P7 and Mm.136889 in P9) are expressed in the liver based on the fact that the ESTs were retrieved from mouse livers. ND, not determined; NI, no information is available in web-based public databases.

Figure 3 RT–PCR analysis for expression of rAAV2-targeted genes in mouse livers. Total liver RNA was extracted from adult BALB/c, C57BL/6 and HTI mouse strains, and expression of each targeted gene was analyzed by RT-PCR. We separated 15 μ l of the RT–PCR products on a 2.0% agarose gel and stained it with ethidium bromide. (a) Coamplification of each targeted gene transcript and Gapd transcript. Lane 1, universal reference RNA; lane 2, BALB/c mouse liver RNA; lane 3, C57BL/6 mouse liver RNA; lane 4, HTI mouse liver RNA; lane 5, RT-minus negative control containing 0.1 µg universal reference RNA; lane 6, template-minus negative control. The positions of target transcript and Gapd transcript are indicated with an arrow and an arrowhead, respectively. (b) RT-PCR amplification of Mm.105278. Lanes 1 and 5, universal reference RNA; lanes 2 and 6, BALB/c; lanes 3 and 7, C57BL/6; lanes 4 and 8, HTI mouse; lane 9, RT-minus negative control containing 0.5 µg universal reference RNA; lane 10, template-minus negative control. RT products corresponding to 0.5 µg and 0.05 µg RNA are used for lanes 1-4 and lanes 5-8, respectively. (c) RT–PCR amplification of Gapd transcript. Lanes are the same as in **a**. M, HaeIII-digested Φ X 174 DNA fragments. The gel images are inverted.

into experimental animals, regardless of how the proviruses were

isolated. Similar unpredictable imperfect structures of rAAV2 proviral genomes and chromosomal deletions at integration sites in G418-selected, genetically unstable HeLa cells were recently reported¹¹, but there was no statistical analyses to establish if there was preference for integration into intragenic regions. There is considerable evidence that target site selection by retroviruses and retrotransposons is non-random. Although there are several conflicting results, local DNA structures and surrounding environment including chromatin structures and transcription factors influence the choice for a target site¹⁷⁻²⁴. It has been recently shown that HIV-1 selectively integrates into active genes²⁵. It is important to note that retroelements and rAAV2 proviral sequences integrate by different mechanisms. Retroelements use their encoded integrase to catalyze integration, whereas rAAV2 vector DNA integration is totally dependent on host cellular proteins. Thus the possibility remains that rAAV2 vectors preferentially integrate into chromosomal regions that are already broken¹¹. Nevertheless, the preferred integration into active genes may be a common propensity of certain kinds of integrating elements including rAAV2. Although rAAV2 vectors integrate at a low efficiency, the current results will need to be considered in risk/benefit considerations until the consequences for vector integration are more fully understood.

METHODS

rAAV2 shuttle vectors. We produced rAAV2 shuttle vectors AAV-EF1 α -GFP.AOSP and AAV-EF1 α -hFAH.AOS (Fig. 1a) based on plasmids pAAV-EF1 α -GFP.AOSP and pAAV-EF1 α -hFAH.AOS2 as described in Supplementary Note online. Both vectors carried the bacterial gene encoding β -lactamase (ampicillin resistance gene or Amp^r) and the ColE1 plasmid origin of replication (ori), allowing for retrieval of vector genome sequences in bacteria.

Strains of mice and animal husbandry. All the animal experiments were done according to the guidelines for animal care at Stanford University and Oregon Health & Science University. We purchased female C57BL/6 mice 6–8 weeks old from Jackson Laboratory. HTI mice were the FAH^{Δexon5} strain previously described¹³ and inbred at the Department of Animal Care, Oregon Health & Science University. We gave HTI mice drinking water containing 2-(2-nitro-4-trifluoro-methylbenzoyl)-1,3-cyclohexanedione (NTBC; Swedish Orphan AB) at a concentration of 7.5 mg l⁻¹. FAH-negative hepatocytes accumulate the toxic metabolite fumarylacetoacetate (FAA) and die in a cell-autonomous manner, but oral administration of NTBC reduces FAA, allowing normal hepatic function and preventing hepatocellular damage. For *in vivo* selection of HTI hepatocytes with integrated rAAV2 vector genomes, we withdrew NTBC from drinking water.



Portal vein injection, in vivo selection and hepatocyte transplantation. We carried out portal vein injection of AAV-EF1α-GFP.AOSP into C57BL/6 mice and sample collection as previously described⁴. We injected adult male HTI mice on NTBC with 3.0×10^{11} particles of AAV-EF1 α -hFAH.AOS into the portal vein (n = 8) and then divided the mice into two groups (n = 4 each). After being kept on NTBC for six weeks (enough time to establish stable hepatocyte transduction with rAAV2), we withdrew NTBC from the mice in Group 1, but continued to give the mice in Group 2 water containing NTBC for an additional eight weeks. To further select for integrated vector genomes and dilute nonintegrated vector genomes, after eight weeks with (Group 1) or without (Group 2) in vivo selection (14 weeks after injection), we isolated hepatocytes from vector-injected mice by a two-step collagenase perfusion and injected one million hepatocytes in 100 µl of appropriate medium into the portal vein of recipient HTI mice on NTBC as previously described¹⁴. We used liver DNA from a Group 1 recipient isolated after a 7-month in vivo selection to isolate rAAV2 proviral genomes. The outcomes of hepatocyte transplantation into recipient HTI mice in Groups 1 and 2 are summarized in Supplementary Note online.

Isolation of proviruses and mouse genomes around integration sites. Vectorcellular DNA junctions from AAV-EF1 α -GFP.AOSP-injected C57BL/6 mouse livers and detailed procedures for the isolation were previously reported⁴ and are concisely explained in Figure 1b. The strategy for isolating whole proviral vector genomes together with 5' and 3' vector-cellular DNA junctions from AAV-EF1 α -hFAH.AOS-transduced HTI mouse hepatocytes was basically the same as our previously published method with minor modifications²⁶. The detailed procedures are found in **Supplementary Note** online.

Construction of restriction enzyme maps of isolated plasmids containing the whole proviral genome. We digested each rescued plasmid with *KpnI*, *XbaI*, *PmeI* or *Bam*HI, alone or in any possible combination, to draw draft maps (see **Fig. 1a**). A *KpnI* site does not exist in the vector but must reside only once in each rescued plasmid. Combining the sequence information, we identified that 11 of 14 proviral genomes resulted from rAAV2 monomer integration with various terminal deletions of the vector genome. The remaining three proviral genomes required additional restriction enzyme digestion and sequencing of subcloned fragments in pBluescript II KS⁻. When we obtained plasmid clones that were identical based on their restriction maps and sequencing data, we considered those as a single integration event, not individual different integration events.

Sequencing of junctions. We carried out sequencing using an ABI PRISM 377 DNA Sequencer (PE Applied Biosystems). The detailed methods for sequencing are available in Supplementary Note online.

Bioinformatics. Isolated mouse cellular DNA sequences were BLAST searched against the public mouse genome database through the National Center for Biotechnology Information (NCBI) and Ensembl browsers. We did a targeted gene search based on the chromosomal localization of each integration site using NCBI Map Viewer and Mouse Contig View of Ensembl Mouse Genome Browser. We assessed the transcriptional activity of each targeted gene using web-based public databases and browsers: NCBI UniGene, SOURCE²⁷, READ¹⁵ and Gene Expression Atlas of Genomics Institute of the Novartis Research Foundation (GNF)¹⁶. Additional information about the bioinformatic analyses is available in **Supplementary Note** online.

RT–PCR of rAAV2-tageted gene transcripts. Among 22 rAAV2-targeted genes, we analyzed expression of 5 genes by RT–PCR. We extracted total liver RNA from an adult female C57BL/6 mouse and an adult male HTI mouse. We purchased total liver RNA from BALB/c mice and a mouse universal reference total RNA from Clontech. We coamplified each target transcript and the gene encoding glyceraldehyde-3-phosphate dehydrogenase (*Gapd*) in the same tube. We calculated the ratio of the PCR product from a target transcript to that from *Gapd* for each sample and normalized it to the ratio from the reference RNA. Thus, the normalized ratio of the universal reference RNA is always 1.0 regardless of the target transcripts, and the normalized ratio of each target transcript in samples can be used to assess as an increase (the ratio >1.0) or a decrease (the ratio <1.0) in the amount of the target transcript in the liver compared to that in the reference sample. Additional information about the RT–PCR analysis is available in **Supplementary Note** online.

Statistics. We assessed the statistical significance of the bias for or against preferential integration into genes by a two-tailed binomial test. In humans, transcription units are estimated to account for about 25–33% of the genome^{28,29}. Based on the difference in the length of the genome and the number of transcripts between the human and the mouse, we estimated the proportion of transcription units to the whole genome (gene density) in the mouse as 0.25–0.41 (see **Supplementary Note** online). We assessed the influence of the presence of *in vivo* selective pressure on integration target site selection by rAAV2 using the χ^2 test and Fisher's exact probability test. We excluded from the analyses two integrations that targeted the 45s pre rRNA gene (J192) and the 45s pre rRNA gene intergenic spacer (J278) because of the redundancy of the ribosomal RNA genes in the genome. The number of shared nucleotides between aligned rAAV2 vector and mouse genomic sequences is one of the indicators for microhomologies. We tested a null hypothesis that there is no bias for or against base sharing by a two-tailed binominal test (see **Supplementary Note** online).

URLs. NCBI mouse genome BLAST search, http://www.ncbi.nlm.nih.gov/ genome/seq/MmBlast.html; NCBI Map Viewer, http://www.ncbi.nlm.nih.gov/ mapview/static/MVstart.html; Ensembl mouse genome database, http://www. ensembl.org/Mus_musculus; NCBI UniGene http://www.ncbi.nlm.nih.gov/ UniGene; SOURCE, http://source.stanford.edu/; READ http://read.gsc.riken. go.jp/; Gene Expression Atlas of Genomics Institute of the Novartis Research Foundation, http://expression.gnf.org/cgi-bin/index.cgi.

Accession numbers. The five genes that we analyzed by RT–PCR are UniGene cluster IDs Mm.26928, Mm.39006, Mm.23257, Mm.105278 and Mm.11360.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank J. Park for data analysis. This work was supported by grants from the National Heart, Lung, and Blood Institute of the US National Institutes of Health to M.A.K.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 15 January; accepted 16 May 2003 Published online 1 June 2003; doi:10.1038/ng1179

- Kay, M.A. *et al.* Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. *Nat. Genet.* 24, 257–261 (2000).
- Nakai, H. *et al.* Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction *in vivo. J. Virol.* 75, 6969–6976 (2001).
- Miao, C.H. et al. The kinetics of rAAV integration in the liver. Nat. Genet. 19, 13–15 (1998).
- Nakai, H., Iwaki, Y., Kay, M.A. & Couto, L.B. Isolation of recombinant adeno-associated virus vector-cellular DNA junctions from mouse liver. *J. Virol.* 73, 5438–5447 (1999).
- Chen, S.J., Tazelaar, J., Moscioni, A.D. & Wilson, J.M. *In vivo* selection of hepatocytes transduced with adeno-associated viral vectors. *Mol. Ther.* 1, 414–422 (2000).
- Li, Z. et al. Murine leukemia induced by retroviral gene marking. Science 296, 497 (2002).
- Marshall, E. Clinical research. Gene therapy a suspect in leukemia-like disease. Science 298, 34–35 (2002).
- Marshall, E. Gene therapy. Second child in French trial is found to have leukemia. Science 299, 320 (2003).
- Rutledge, E.A. & Russell, D.W. Adeno-associated virus vector integration junctions. J. Virol. 71, 8429–8436 (1997).
- Yang, C.C. *et al.* Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration *in vivo* and *in vitro. J. Virol.* **71**, 9231–9247 (1997).
- Miller, D.G., Rutledge, E.A. & Russell, D.W. Chromosomal effects of adeno-associated virus vector integration. *Nat. Genet.* 30, 147–148 (2002).
- Waterston, R.H. et al. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562 (2002).
- Grompe, M. *et al.* Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. *Genes Dev.* 7, 2298–2307 (1993).
- Overturf, K. *et al.* Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinaemia type I. *Nat. Genet.* **12**, 266–273 (1996).
- Miki, R. *et al.* Delineating developmental and metabolic pathways *in vivo* by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. *Proc. Natl. Acad. Sci. USA* 98, 2199–2204 (2001).
- Su, A.I. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA* 99, 4465–4470 (2002).
- 17. Vijaya, S., Steffen, D.L. & Robinson, H.L. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J. Virol.* **60**, 683–692 (1986).
- Sandmeyer, S.B., Hansen, L.J. & Chalker, D.L. Integration specificity of retrotransposons and retroviruses. *Annu. Rev. Genet.* 24, 491–518 (1990).
- Scherdin, U., Rhodes, K. & Breindl, M. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. 64, 907–912 (1990).
- Chalker, D.L. & Sandmeyer, S.B. Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev.* 6, 117–128 (1992).
- Pryciak, P.M. & Varmus, H.E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 69, 769–780 (1992).
- Muller, H.P. & Varmus, H.E. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* 13, 4704–4714 (1994).
- Pruss, D., Bushman, F.D. & Wolffe, A.P. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA* 91, 5913–5917 (1994).
- Leclercq, I. *et al.* Host sequences flanking the human T-cell leukemia virus type 1 provirus *in vivo. J. Virol.* 74, 2305–2312 (2000).
- Schroder, A.R. et al. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110, 521–529 (2002).
- Nakai, H. et al. Helper-independent and AAV-ITR-independent chromosomal integration of double-stranded linear DNA vectors in mice. *Mol. Ther.* 7, 101–111 (2003).
- Diehn, M. et al. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res. 31, 219–223 (2003).
- Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
- 29. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).