

Supplementary Table2: Flanking repeat elements of *IGHV* genes used for identifying orthologous and paralogous genes between humans and chimpanzees.
See footnote for detailed information.

Human			Chimpanzee		
<i>IGHV</i> †	5' Repeats	3' Repeats	<i>IGHV</i> †	5' Repeats	3' Repeats
<i>H1</i>	L1M2_5, L1M4B	L1ME_ORF2, LTR17	<i>C1</i>	L1M2_5, L1M4B	L1ME_ORF2, LTR17
<i>H2</i>	AluSg, AluSx	L1MCB_5, L1MCB_5	<i>C2</i>	AluSg, AluSx	L1MCB_5, L1MCB_5
<i>H3</i>	L1A_OC, L1MC3	MLT1H, MLT1H	<i>C3</i>	L1A_OC, L1MC3	MLT1H, Gypsy-11-LTR
<i>H4</i>	L1M2_5, MLT1B	AluSp, MER89I	<i>C4</i>	AluY, MLT1B	AluSp, AluY
<i>H5</i>	L1PREC1, L1PREC2	MER65A, L1MEC_5	<i>C5</i>	L1MD2, L1PREC2	MER65A, AluSq
			<i>C7</i>	L1PB4, L1PREC2	MER65A, L1MCA_5
LS	—	—	<i>C6</i>	L1P_MA2	LTR131_MD, L1PA16
<i>H6</i>	L1PREC2, MER65A	L1MEC_5, AluY	<i>C8</i>	L1PREC2, MER65A	L1MCA_5, AluY
<i>H7</i>	MER11A, CfERVF2	L1MC2, L1MC3	<i>C9</i>	MER11A, CfERVF2	L1MC2, L1MC3
<i>H8</i>	MLT2C2, ATCOPIA26I	Hat-2_SD, L1ME2	<i>C10</i>	MLT2C2, ATCOPIA26I	Hat-2_SD, L1ME2
<i>H9</i>	MER41B, MLT2C2	L1ME1, L1MCB_5	<i>C11</i>	MER41B, MLT2C2	L1ME1, L1MCB_5
<i>H10</i>	L1MC1, L1MC2	ATCOPIA90, L1PREC2	<i>C12</i>	L1MC1, L1MC2	ATCOPIA90, L1PREC2
<i>H11</i>	LTR16D, MLT2C2	L1PA16, L1P_MA2	<i>C13</i>	LTR16D, MLT2C2	AluJo, L1MB5
<i>H12</i>	L1PB4, L1PREC2	L1P_MA2, L1HS	NI	—	—
<i>H13</i>	AluSc, MLT2C2	RETROSAT2LTR, Copia10-PTR_I	<i>C23</i>	L1P_MA2, MLT2C2	RETROSAT2LTR, LTR16D
<i>H36</i>	L1P_MA2, MLT2C2	RETROSAT2LTR, L1ME1	<i>C27</i>	L1PREC2, MLT2C2	RETROSAT2LTR, L1ME1
<i>H41</i>	MLT2C2, RONIN2_LTR	ERV5_5I_RN, L1ME1			
<i>H14</i>	L1PREC2, L1MC3	L1M2B_5, L1MCB_5	NI	—	—
<i>H15</i>	MER65A, ATGP5A_I	L1ME1, MER4A1_LTR	NI	—	—
<i>H16</i>	L1MD1, L1MC2	FATIMA_TM_I, L1M4B	NI	—	—
<i>H17</i>	L1ME2, Mariner1_AO	L1ME2, L1ME1	NI	—	—
<i>H18</i>	L1PREC1, L1ME2	MER65A, MER65A	NI	—	—
<i>H19</i>	MER65A, WHAM3_TM_LTR	L1MA7, AluYb3a1	NI	—	—

H20	DNA9TA1_DR, AluSz	L1MC2, L1MC3	NI	—	—
H21	L1ME_ORF2, L1MC3	MLT1H2, L1MC_5	C42	L1MC1, L1MC3	L1MC_5, RatERVL
H56	L1MC1, L1MC3	L1MC_5, MERVL_2A			
H22	L1MC1, L1MC2	ERV44_MD_I, L1MCB_5	C43	L1MC1, L1MC2	L1MCB_5, MER89I
H57	L1MC1, L1MC2	L1MCB_5, MER89I			
H23	L1MA9, MSTA	RETROSAT2LTR, LTR6A	NI	—	—
H24	LTR16D, MLT2C2	DMCR1A, AluSc	C41	LTR16D, MLT2C2	RETROSAT2LTR, Polinton-1_TV
H54	LTR16D, MLT2C2	Polinton-1_TV, L1ME1			
H25	MER41B, MLT2C2	L1ME1, Polinton-2_CB	C14	MER41B, MLT2C2	L1ME1, Polinton-2_CB
H59	MLT2C2, LTR6A	L1ME1, LTR13A			
H26	AluSz, TAG1	HELITRON4, AluSz	C15	AluSz, TAG1	HELITRON4, AluJo
H27	L1PA4, MER65A	L1MA10, AluJo	C16	L1PA4, MER65A	L1MA10, L1MA10
H28	L1MC2, MER52D	L1MCB_5, L1MCB_5	C17	L1MC2, MER52D	KANGAROO1_VC, L1MCB_5
H29	L1-6_CR, L1MA7	L1ME1, Copia23-PTR_I	C18	L1-6_CR, L1MA7	L1ME2, Copia23-PTR_I
H30	L1PA6, L1MDA_5	L1MD1, TREP72	LS	—	—
H31	L1-BT, L1P_MA2	HELMET2, L1P_MA2	C19	L1PREC1, L1PREC2	HELMET2, L1P_MA2
H32	Gypsy-18-I_DR, AluSz	L1MC2, L1MC3	LS	—	—
H33	MER65A, L1P_MA2	MLT1H_I, L1MCB_5	C20	MER65A, L1P_MA2	MLT1H_I, L1MCB_5
H34	MLT1B, RandI-3	L1ME1, AluSp	C21	L1M2_5, MLT1B	MER61I, MER61I
			C25	AluYi6, MLT1B	AluJo, L1ME1
H35	L1ME2, LTR16D	L1ME_ORF2, L1ME_ORF2	C22	L1ME1, LTR16D	L1ME_ORF2, L1ME_ORF2
H40	L1ME1, LTR16D	L1ME_ORF2, L1ME_ORF2	C26	LTR16D, LTR16D	L1ME_ORF2, L1ME_ORF2
			C86	L1ME1, LTR16D	—
H37	THE1D, SETARIA2	L1-4_DR, LTR16D	C28	LTR52, GYPOT1_I	L1-4_DR, LTR16D
H42	LTR52, SETARIA2	LTR16D, L1ME2			
LS	—	—	C24	RETROSAT2LTR, LTR16D	L1ME2, AluYi6

H38	L1-4_DR, LTR16D	L1ME2, AluYi6	C29	L1-4_DR, LTR16D	Helitron-1_DVir, Gypsy-19-I_DR
H43	SETARIA2, LTR16D	L1ME2, Gypsy-19-I_DR			
H39	AluYi6, MLT1B	L1ME2, AluSp	C30	Gypsy-19-I_DR, MLT1B	AluSp, MER89I
H44	MLT1B, RIRE5- LTR_OS	AluSp, MER89I			
H45	MER89I, MER89I	L1MCB_5, MER89I	C31	MER89I, MER89I	MER89I, MER89I
H46	L1MA9, MSTA	SHAMUDRA_MT, L1MA7	C32	L1MA9, MSTA	SHAMUDRA_MT, Sm2
H47	AluSq, L1MA7	L1-55_XT, L1M4B	C33	AluSq, L1MA7	L1M4B, L1ME_ORF2
H48	MER65A, L1-1_CR	L1ME1, AluJo	C34	MER65A, ENSPM2_OS	L1ME1, AluJo
H49	L1PA6, MER65A	LTR19C, MER4A1	C35	L1PA6, MER65A	L1ME2, MER4A1_LTR
LS	—	—	C36	AluY, LIMC3	L1MCB_5, MLT1CR
H50	L1M3A_5, MLT1B	MER89I, MER89I	C37	L1M3A_5, MLT1B	MER89I, Ogre-PT1_I
H51	MER89I, MER89I	L1PREC2, MSTB	C38	MER89I, Ogre-PT1_I	L1-BT, L1PREC2
H52	LTR16D, AluSq	L1MA10, AluYi6	C39	L1MA1, LTR16D	L1MA10, THER1
H53	L1PA4, LTR5B	L1PREC2, L1PREC2	C40	L1PA4, LTR5B	L1PREC2, L1PREC2
H55	L1ME1, AluSq	L1MB4, L1MB5	LS	—	—
H58	MER65A, MER65A	L1ME2, MER4A1_LTR	C44	MER65A, MER65A	L1ME2, MER4A1_LTR
H60	L1ME2, L1MA7	NonLTR-5_CR, L1ME1	C45	CR1-21_SP, L1MA7	GGLTR10C_I, L1ME1
H61	AluSq, LTR16D	MTCOPIA1_I, L1ME_ORF2	C46	L1ME2, AluSp	L1A_OC, L1ME_ORF2
H62	AluSx, L1P_MA2	MER65A, MER21I	C47	AluSx, L1P_MA2	MER65A, MER21I
H63	IKEROS_HV_I, LTR16D	LTR16D, RETROSAT2LTR	C48	IKEROS_HV_I, LTR16D	MSAT-1B_CR, LTR16D
H64	IKEROS_HV_I, LTR16D	LTR16D, RETROSAT2LTR	C49	IKEROS_HV_I, LTR16D	MSAT-1B_CR, LTR16D
H65	LTR16D, LTR16D	RETROSAT2LTR, MER52D	C50	MSAT-1B_CR, LTR16D	RETROSAT2LTR, MER52D
H66	MER65A, LTR13	L1ME1, MER61I	C51	MER65A, LTR13	LTR66_MD, L1ME1
H67	MER65A, LTR13	L1ME1, MER61I			
H68	MER69C, L1ME2	LTR16D, L1ME2	LS	—	—

H69	L1ME2, LTR16D	L1ME2, Gypsy-19-I_DR	C52	L1-4_DR, LTR16D	L1ME2, Gypsy-19-I_DR
LS	—	—	C53	THE1BR, THE1B	L1ME2, Gret1_I
H70	Gypsy-19-I_DR, MLT1B	ALU, MER89I	LS	—	—
H71	ALU, MER89I	MLT1H2, L1MCB_5	LS	—	—
H72	ALU, MER89I	MLT1H2, L1MCB_5	LS	—	—
H73	AluYA5, AluYf2	L1MC2, L1MC3	LS	—	—
H74	L1ME_ORF2, L1MC3	MLT1H2, L1MCB_5	LS	—	—
H75	L1M3A_5, MLT1B	THE1BR, THE1B	C55	KANGAROO1, MLT1B	THE1BR, THE1BR
H78	L1ME2, MLT1B	THE1BR, THE1B			
H76	THE1BR, THE1B	QUASIMODOI, L1ME1	C57	THE1BR, THE1B	L1ME1, AluY
H79	THE1BR, THE1B	L1ME1, AluY			
LS	—	—	C56	MLT1B, THE1BR	THE1BR, THE1B
H77	RP5S, L1ME2	MLT1B, THE1BR	C54	L1ME2, KANGAROO1	MLT1B, THE1BR
H80	Gypsy-1-I_AN, LTR16D	L1ME_ORF2, L1ME_ORF2	C58	L1ME2, LTR16D	L1ME_ORF2, L1ME_ORF2
H81	LTR22C_LTR, TABARE	ENS1_Cis, L1ME2	C59	MER65A, LTR22C_LTR	L1ME2, Polinton-1_CB
H82	L1ME3A, UCON13	RTVL-Ib_I, L1MA8	C60	L1ME3A, UCON13	TABARE, RTVL-Ib_I
H83	L1PA6, MER65A	L1ME2, MER61I	C61	L1PA6, MER65A	ERV12_MD_LTR, L1ME2
H84	L1MC1, L1MC2	L1MCB_5, ATCOPIA19I	C62	L1MC1, L1MC2	L1MCB_5, ATCOPIA19I
H85	AluSz, L1PB4	MER34B_I, L1ME_ORF2	C63	AluSz, L1PB4	L1ME_ORF2, L1MD3
H86	MER45C, L1MC1	AluY, L1MC3	C64	MER45C, RONIN2_LTR3	L1MC3, AluY
H87	L1MC3, L1MA2	L1MC2, L1MC3	C65	L1MC3, L1MA2	L1MC2, L1MC3
H88	LTR22B, L1MC3	AluY, L1MCB_5	C66	LTR22B, L1MC3	AluY, L1MCB_5
H89	L1MC1, L1MC2	L1PREC2, L1P_MA2	C67	L1MC1, L1MC2	L1PREC2, L1P_MA2
H90	L1PB4, L1P_MA2	MLT1G1, LTR16D	C68	L1PB4, L1P_MA2	MLT1G1, LTR16D
H91	MER41B, MLT2C2	AluJb, L1ME2	C69	MER41B, MLT2C2	L1MA10, AluJb
H92	L1ME2, L1MA7	Polinton-1_PI, MSTA	C70	L1ME2, L1MA7	MSTA, LTR32

<i>H93</i>	HERV19I, L1MA7	L1M4B, PB1	<i>C71</i>	L1ME2, L1MA7	L1M4B, PB1
<i>H94</i>	ATCOPIA30_I, MLT2C2	MUDSOLT1, L1MB7	<i>C72</i>	ATCOPIA30_I, MLT2C2	MUDSOLT1, L1MB7
<i>H95</i>	L1M2C_5, L1M2C_5	L1MA9, MER110I	LS	—	—
<i>H96</i>	MER110I, MER65A,	L1MA9, L1ME2	LS	—	—
<i>H97</i>	MER65A, SZ-22_I	L1MA9, CR1-4_SP	<i>C73</i>	L1PA15, MER65A	L1MA9, RICKSHA
<i>H98</i>	CR1-4_SP, AluYa5	L1MA9, L1ME2			
<i>H99</i>	MER25, L1ME_ORF2	AluSx, L1-BT	<i>C74</i>	L1P_MA2, LTR16D	MER65A, MER4BI
<i>H100</i>	L1HS, AluSx	AluSp, LTR16D	LS	—	—
<i>H101</i>	AluSp, LTR16D	L1ME2, Sadhu1-1	<i>C75</i>	AluSp, LTR16D	L1ME2, ATHILA4C_I
<i>H102</i>	AluSp, RandI-3	MER89I, MER89I	<i>C76</i>	MLT1B, AluSp	MER89I, MER89I
<i>H103</i>	MER89I, MER89I	ERV43_MD_I, LTR15	<i>C77</i>	MER89I, MER89I	MARY1_TM, L1MCB_5
<i>H104</i>	AluSz, AluYd8	MLT1I, BMSAT1	<i>C78</i>	AluSz, AluJo	MLT1I, BMSAT1

Footnote: From a large number of repetitive elements identified from the entire *IGHV* locus, only two flanking repeat elements for each of the 5' and 3' ends of *IGHV* gene are shown. In some orthologs or paralogs, the closest 5' and 3' flanking repeat elements are not completely identical. Hence, the orthologous or paralogous relationship of these *IGHV* genes were verified by phylogenetic analysis and by similarities of adjacent gene segments and/or nearby repeat elements. For example, the adjacent repeat elements at the 5' and 3' ends of *H4* and *C4* *IGHV* genes are not completely identical. But both *H4* and *C4* belong to clan II and are phylogenetically closest. Furthermore, the adjacent *IGHV* genes of *H4* and *C4* are orthologous and the surrounding repeat elements are also quite similar. The organization of surrounding repeat elements of *H4* gene is L1M3DE_5 -- AluSc -- GOLEM_A -- L1M2_5 -- MLT1B -- ***H4*** -- AluSp -- MER89I -- MER89I -- L1PREC2 -- L1PA15 -- L1ME_ORF2, whereas the organization for the *C4* gene is L1M3DE_5 -- AluSc -- GOLEM_A -- L1M2_5 -- AluY -- MLT1B -- ***C4*** -- AluSp -- AluY -- MER89I -- L1PREC2 -- L1PA15 -- L1ME_ORF2. Therefore, the orthologous relationship between *H4* and *C4* *IGHV* genes is supported by the high sequence similarity of the

genes, the similarity of repetitive elements surrounding them, and the similarities in adjacent *IGHV* genes. The *IGHV* gene lost in either human or chimpanzee lineage is indicated by “LS”. The orthologous genes that could not be found because of the incomplete sequence are denoted by “NI” (nonidentified).

† Gene names are given by the letters ‘H’ (for human) and “C” (for chimpanzee) followed by a number, which indicates the genomic location numbered from the *IGHD* gene.