

Supplementary Information Online

Supplementary Table 1 online

Proviral integration site sequences detected by LAM-PCR in SCID-X1 patients.

Supplementary Table 2 online

Statistical analysis of common integration sites (CIS).

Supplementary Table 3 online

Summary of all common integration sites (CIS) detected among the 9 patients.

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
1	CD14	15	100	17q21.31	FLJ30656		704				yes	42619874	no	no
1	CD14	18	52	20q11.21	BCL2L1			408	1		yes	31055248	no	no
1	CD15	15	74	11q13.1	FKSG44					7714		64964066	no	no
1	CD15	15	213	6p25.3	SEC5L1		177142					815251	no	no
1	CD15	18	102	8q23.1	ANGPT1			207932	6			108258906	no	no
1	CD19	18	131	21q22.12	RUNX1			20781	1			35321210	yes	no
1	CD19	18	47	10p11.23	MAP3K8			23661	5			30750486	no	no
1	CD19	18	78	12p11.23	ARNTL2					41757		27506490	no	no
1	CD19	18	109	12p13.32	CCND2	PBL 38 mon	3170				yes	4249921	yes	no
1	CD19	18	43	1p35.3	EPB41			42060	1			28943200	no	no
1	CD19	18	171	1q42.3	TOMM20					148444		232165365	no	LINE
1	CD19	18	386	22q13.31	CELSR1			25130	1			45120178	no	no
1	CD19	18	84	2p21	HAAO		80535					43074785	no	LINE
1	CD19	18	337	4q35.1	DCTD					44717		184462431	no	no
1	CD19	18	23	5q31.2	CXXC5		868				yes	139055579	no	no
1	CD19	18	154	6q27	MGC35308					53664	yes	166575355	no	no
1	CD19	18	92	9q33.2	GSN			3049	1			119409501	no	SINE
1	CD3	15	152	11p13	LMO2		1260					33879405	yes	no
1	CD3	15	61	11p15.4	ART5					13356		3610688	no	SINE
1	CD3	15	71	11q14.1	FLJ23441					6843		77866709	no	SINE
1	CD3	15	183	14q11.2	DAD1		714				yes	21048970	no	LINE
1	CD3	32	108	14q23.1	PRKCH		2861					59775508	no	LINE
1	CD3	15	192	14q24.1	RAD51L1					87903		67140558	yes	no
1	CD3	32	78	17p12	FLJ45831					209149		14662317	no	no
1	CD3	32	68	1p11.2	NOTCH2			55979	2			119902435	no	no
1	CD3	32	98	1p13.2	ST7L			595	1		yes	112460173	no	no
1	CD3	15	101	1p31.2	PDE4B			537536	6			66166317	no	no
1	CD3	15	285	5q23.3	IRF1		7641					131910284	no	no
1	CD3	24	62	6p21.1	TBN		31416					42033699	no	no
1	CD3	15	24	9q34.3	EGFL7		44569					134870101	no	no
1	CD3	32	68	12q22	SOCS2	PBL 38 mon				23700		92496146	no	LTR
1	PBL	38	32	11q23.3	MGC13053			198	Exon1		yes	117843374	no	no
1	PBL	38	63	12p13.32	CCND2		2116				yes	4251083	yes	no
1	PBL	38	39	12q14.3	HMGA2			56728	3			64561186	yes	no
1	PBL	38	118	13q31.1	SLITRK5		788831					85234040	no	no
1	PBL	38	218	14q24.3	C14orf43			5576	1			72211457	no	no
1	PBL	38	109	17p13.1	DULLARD			784	2		yes	7355040	no	no
1	PBL	38	313	17q25.3	SEPT9			68931	1			75982203	yes	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	3' of RefSeq Intron Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
1	PBL	38	86	18q21.33	BCL2			68896	1		59066003	yes	SINE
1	PBL	38	45	19p13.11	FLJ20422					13621	19077810	no	no
1	PBL	38	173	1p11.2	NOTCH2			60328	2		119898329	no	no
1	PBL	38	84	1p32.3	SSBP3		59146				54300761	no	no
1	PBL	38	70	1q22	MAPBPIP		224			yes	153241173	no	no
1	PBL	38	30	22q12.2	DEPDC5		3693				30470930	no	LINE
1	PBL	38	116	22q12.3	APOL3		836				34829626	no	LTR
1	PBL	38	225	2p15	COMMD1			218063	2		62325406	no	no
1	PBL	38	53	2q21.2	GPR39		70156				133315229	no	no
1	PBL	38	33	4q35.1	PRDM5		52792				122355626	no	no
1	PBL	38	41	6p21.1	RUNX2			4732	1		45441901	no	no
1	PBL	38	78	6p21.31	PACSL1			60841	2		34541601	no	no
1	PBL	38	30	6p23	JARID2		342796				15011710	no	SINE
1	PBL	38	112	7q11.23	FKBP6			332269	Exon8		72184669	no	no
1	PBL	38	21	9q33.2	TRAF1		5843				119071061	no	no
1	PBMC	15	61	17q24.1	FLJ32065		9490				63531277	no	no
1	PBMC	15	23	1p32.3	SSBP3		2854				54244668	no	no
1	PBMC	15	176	9q32	TNFSF8		27404				113096219	no	no
2	CD3	13	33	10p12.2	PIP5K2A	CD3 38 mon		86713	1		22920777	no	no
2	CD3	13	79	10q25.1	KIAA1754		78808				105837058	no	no
2	CD3	13	119	11p12	LOC119710					900018	37545144	no	no
2	CD3	13	26	11p15.4	SWAP70			53147	3		9703109	no	no
2	CD3	38	356	11q13.1	PLCB3					460	63810848	no	no
2	CD3	13	166	11q13.2	SUV420H1					14470	67684397	no	SINE
2	CD3	38	210	11q22.3	CASP1			18463	1		104491412	no	LINE
2	CD3	13	180	12p13.31	TNFRSF1A		707				6322229	no	SINE
2	CD3	38	216	12p13.32	CCND2		57826				4195373	yes	no
2	CD3	13	72	12q21.1	GLIPR1		7464				74153337	no	LINE
2	CD3	13	91	12q24.11	FLJ40142		11798				108980018	no	no
2	CD3	38	365	12q24.31	GPR81		6190				121658845	no	SINE
2	CD3	38	84	13q12.3	MGC2599		68943				28748528	no	no
2	CD3	13	146	13q14.3	DLEU7			32886	1		49183190	no	simple repeats
2	CD3	13	84	13q22.1	KLF12					68610	71989540	no	no
2	CD3	38	185	14q24.3	C14orf4		1961				75487036	no	no
2	CD3	13	46	14q32.33	KIAA0125		58608				104303680	no	CNP
2	CD3	13	69	15q14	NOLA3				5336	yes	32344639	no	no
2	CD3	38	60	15q21.3	LIPC		73375				56366856	no	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	3' of RefSeq Intron Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
2	CD3	13	204	15q22.31	MYCPBP		27765				63828169	no	no
2	CD3	13	49	16p12.1	NSMCE1	CD3 38 mon	18306				27253706	no	LINE
2	CD3	13	57	16p12.3	THUMPD1		5738				20725409	no	LTR
2	CD3	38	214	16p12.3	SMG1		19648				18923939	no	no
2	CD3	38	53	16q24.1	KIAA0182		58882			yes	85369521	no	no
2	CD3	38	17	17p11.2	RAI1			73153	2		17858506	no	no
2	CD3	38	45	17q21.32	SCAP1			1898	1		46980292	no	no
2	CD3	13	121	17q23.2	DGKE	CD3 38 mon	52118				55334073	no	LINE
2	CD3	13	314	17q23.2	MSI2			183334	2		55992216	yes	no
2	CD3	13	68	18q12.3	PIK3C3					32878	37946309	no	no
2	CD3	38	254	18q23	NFATC1			127623	8	yes	75382372	no	no
2	CD3	13	135	19p13.11	CLONE 24945					8972	17994881	no	SINE
2	CD3	13	23	19p13.13	RFX1			16130	2		13961824	no	LINE
2	CD3	13	192	19p13.3	MOBLAK			4971	1		2042298	no	no
2	CD3	13	36	19q13.2	R29124_1		22202				46752225	no	LTR
2	CD3	13	64	19q13.2	GSK3A		1567			yes	47440143	no	SINE
2	CD3	13	227	19q13.32	AP2S1		11380			yes	52057311	no	no
2	CD3	13	104	19q13.43	ZNF444			1042	1	yes	61345410	no	SINE
2	CD3	13	168	1p11.2	NOTCH2			53563	2		119905094	no	no
2	CD3	41	39	1p33	AGBL4			423099	5		48458204	no	SINE
2	CD3	38	63	1p35.2	PTP4A2		23527				31929955	no	no
2	CD3	38	237	1p35.2	PTP4A2		17201				31923629	no	SINE
2	CD3	38	56	1q24.2	ATP1B1		65921				166197403	no	LINE
2	CD3	38	236	1q32.2	CD34			418	1	yes	205167796	no	no
2	CD3	13	86	1q44	SDCCAG8			136390	12		240502590	no	no
2	CD3	13	115	1q44	KIAA1720		1205			yes	245936204	no	SINE
2	CD3	13	49	20p12.3	PLCB4			16084	1		9098010	no	LINE
2	CD3	13	207	20q11.21	FLJ33706					51077	31914585	no	no
2	CD3	13	251	20q11.23	C20orf77			2314	1		37349796	no	SINE
2	CD3	13	25	20q13.12	PRKCBP1		2028				46672924	no	no
2	CD3	38	96	20q13.12	PKIG			81278	2		43927143	no	no
2	CD3	38	90	21q21.1	USP25					363346	16537594	no	no
2	CD3	13	60	21q22.12	RUNX1			38042	1		35303949	yes	no
2	CD3	13	77	21q22.12	RUNX1			23039	1		35318998	yes	no
2	CD3	38	56	21q22.12	RUNX1		454940				35796931	yes	no
2	CD3	13	189	22q11.1	MGC57211	CD3 38 mon		13225	1		15663911	no	no
2	CD3	13	97	22q12.2	MGC26710					4544	28902158	no	LINE
2	CD3	38	78	22q13.1	GRAP2			60928	4		38601053	no	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	3' of RefSeq Intron Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements	
2	CD3	38	49	22q13.2	PACSN2			9278	1		41632814	no	LTR	
2	CD3	38	530	22q13.31	GTSE1		1472			yes	44904077	no	no	
2	CD3	13	271	2p13.3	AAK1		706			yes	69846089	no	no	
2	CD3	13	142	2q37.3	CMKOR1		22342				237742851	yes	LINE	
2	CD3	13	376	3p13	RYBP				1671		72345493	no	no	
2	CD3	38	115	3q13.12	CD47		29136				109159939	no	SINE	
2	CD3	13	96	3q13.13	ESRRBL1		1724				109263948	no	no	
2	CD3	13	281	3q25.2	MME		318827				155799545	no	SINE	
2	CD3	38	28	3q26.32	TBL1XR1		140400				178376353	no	no	
2	CD3	38	214	3q29	ACK1			6409	1		196956067	no	LINE	
2	CD3	13	37	4q12	KIT			2183	1		55441886	yes	no	
2	CD3	13	39	4q28.1	SPRY1					13395	124797123	no	no	
2	CD3	13	46	5p13.1	PTGER4		346506				40379027	no	no	
2	CD3	38	381	5q15	LRAP		3414				96285927	no	no	
2	CD3	13	128	5q34	MAT2B			1603	1	yes	162912729	no	no	
						CD3 38 mon, 13 mon								
2	CD3	32	175	6p21.1	TBN		33936				42031179	no	no	
2	CD3	13	55	6p21.32	HLA-DQB1					52987	32620717	no	LINE	
2	CD3	13	316	6p22.2	LRRC16		106388				25512560	no	no	
2	CD3	13	57	6p25.1	RREB1	CD3 38 mon	215418				6837845	no	no	
2	CD3	13	93	6q15	ORC3L			12219	3		88307658	no	no	
2	CD3	38	210	6q15	BACH2			65936	3		90936123	no	SINE	
2	CD3	38	65	6q22.31	MAN1A1		91084				119742586	no	LTR	
2	CD3	13	76	6q22.31	TCBA1	CD3 38 mon		159551	1		124265416	no	no	
2	CD3	13	52	6q24.3	STXBP5		251500				147253945	no	LINE	
2	CD3	13	116	6q26	AGPAT4			777	1		161603548	no	no	
2	CD3	38	71	7p15.2	NFE2L3		50250				25883882	no	no	
2	CD3	13	250	7p15.3	IMP-3		3411			yes	23255670	no	no	
2	CD3	38	167	7q22.1	ZNF38		36402				99222412	no	SINE	
2	CD3	13	62	7q32.1	METTL2					27584	yes	127724131	no	no
2	CD3	13	144	7q35	CNTNAP2			538947	1		145744632	no	no	
2	CD3	38	48	7q36.1	HIMAP2			363	1		149774898	no	no	
2	CD3	38	140	8q21.13	PKIA		488959				78989339	no	DNA	
2	CD3	13	130	9p13.1	SHB			70718	2		37988490	no	SINE	
2	CD3	13	266	9q34.11	CDK9		2155				125922227	no	no	
2	CD3	13	162	9q34.11	CDK9		1728			yes	125922654	no	no	
2	CD3	13	160	Xp11.3	ZNF41		1118			yes	46389084	no	no	
2	CD3	13	82	Xp22.11	EIF2S3		34015				23400476	no	no	

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
2	CD3	38	61	Xp22.22	FAM9C		46611					12470837	no	no
2	CD3	13	44	Xp22.22	FAM9C	CD3 38 mon	46404					12470630	no	no
2	CD3	13	151	Xq21.2	CHM			794	1			84073630	no	no
2	PBL	38	96	10q25.2	PDCD4		8172					112288011	no	no
2	PBL	38	115	22q12.2	MGC26710	CD3 13 mon				3934		28901548	no	simple repeats
2	PBL	38	64	2p14	MEIS1					11545		66785969	no	no
2	PBL	38	329	3p21.31	RASSF1			2002	1			50335377	no	LINE
4	CD15	13	43	11p15.1	LDHA	PBL 41 mon	9521					18370899	no	simple repeats
4	CD3	13	119	10p11.22	NRP1					51336		33419097	no	no
4	CD3	24	55	10q21.3	TRIP8			33176	1			64340374	no	no
4	CD3	17	124	10q22.1	PALD			4550	1			71587730	no	no
4	CD3	6	103	10q22.1	PALD			18980	1			71602160	no	SINE
4	CD3	13	56	10q23.33	CYP2C19			35477	4			96222536	no	LINE
4	CD3	24	102	10q24.2	NKX2-3	CD3 17 mon	904				yes	100956393	no	SINE
4	CD3	24	123	10q24.33	SH3PXD2A					124058		105182831	no	LINE
4	CD3	24	28	11p11.2	KAI1			9286	1			44560803	no	LINE
4	CD3	24	103	11p13	LMO2	CD3 37 mon; CD19 31 mon; PBL 34 mon; CD34 pre		2019	1			33876126	yes	LINE
4	CD3	13	86	11p13	LMO2	CD3 6 mon	39518					33917663	yes	no
4	CD3	6	36	11p15.4	SWAP70	CD 37 mon		2585	1			9652547	no	LINE
4	CD3	13	30	11p15.4	OR52N2		3785					5802090	no	SINE
4	CD3	6	84	11q12.1	MGC15937					54091		58553967	no	no
4	CD3	6	20	11q12.1	MGC15937					53626		58553502	no	SINE
4	CD3	17	110	11q24.1	ASAM		69154					122672813	no	LTR
4	CD3	17	126	12p11.21	C12orf 14			210	1		yes	31370130	no	no
4	CD3	24	92	12p11.23	ITPR2	CD3 17mon	36345					26913743	no	SINE
4	CD3	24	27	12p13.2	DKFZp761P1010			4878	1			10713028	no	no
4	CD3	13	71	12p13.2	ETV6	CD3 24 mon		151156	2			11845211	yes	no
4	CD3	13	24	12p13.32	CCND2		2494				yes	4250705	yes	no
4	CD3	24	51	12p13.32	CCND2		2138				yes	4251016	yes	no
4	CD3	6	234	12q13.11	MGC16044					127086		46043794	no	no
4	CD3	37	90	12q13.11	HDAC7A		16574					46495764	no	SINE
4	CD3	13	56	12q23.3	CHST11			224428	2			103577945	no	SINE
4	CD3	6	62	13q12.11	FLJ25952					32171		19716339	no	SINE

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
4	CD3	17	346	13q14.3	SUGT1		955				yes	51024047	no	LINE
4	CD3	24	97	14q13.2	PSMA6		1551					33749966	no	no
4	CD3	17	135	14q21.3	ARF6					76037		48427809	no	no
4	CD3	6	69	14q22.1	PTGER2		2293					50768858	no	DNA
4	CD3	13	46	14q24.1	KIAA0247	CD3 6 mon		91465	2			68159864	no	no
4	CD3	24	21	14q24.3	C14orf4					68330		75412599	no	LINE
4	CD3	24	219	15q21.3	TCF12			364338	16			55291227	yes	no
4	CD3	6	27	15q22.31	HERC1		1004				yes	61842909	no	no
4	CD3	24	124	15q24.2	DKFZP434H132		21396					73188670	no	no
4	CD3	6	60	15q24.2	MDSO18					4228		73063111	no	SINE
4	CD3	17	71	16p12.3	THUMPD1		3018					20722689	no	LTR
4	CD3	17	70	16p12.3	NOMO2		14640					18463096	no	no
4	CD3	17	261	16q22.3	ZNRF1			72041	1			74884510	no	LTR
4	CD3	37	73	16q24.2	BANP			11050	1			87777720	no	DNA
4	CD3	6	78	16q24.3	CDT1		13628					88600704	no	SINE
4	CD3	6	62	17p13.1	MDS006			2241	1			10803779	no	no
4	CD3	6	67	17q11.2	DKFZp434K1421		369				yes	28589018	no	no
4	CD3	6	25	17q11.2	SSH2			204108	2			28198473	no	no
4	CD3	24	76	17q11.2	NF1			213553	15			29781233	yes	no
4	CD3	24	76	17q11.2	EVI2B			5421	1			29781233	yes	no
4	CD3	17	100	17q11.2	TIAF1			35615	2			27613379	no	simple repeats
4	CD3	6	211	17q21.2	KCNH4		307				yes	40706768	no	no
4	CD3	13	108	17q21.31	M17S2			111	1			41695767	no	no
4	CD3	13	51	17q21.31	FMNL1			5110	1			43779824	no	no
4	CD3	6	113	17q21.32	HOXB3		2742					47099772	no	no
4	CD3	24	108	17q21.33	NME1			181	1			49705741	no	no
4	CD3	6	103	17q24.1	RGS9		65788				yes	63669135	no	SINE
4	CD3	6	115	18p11.31	DLGAP1			286728	4			3585407	no	no
4	CD3	6	63	18q11.2	CABLES1			72242	4			19060018	no	no
4	CD3	37	245	18q21.33	BCL2					46996		58892552	yes	LINE
4	CD3	6	47	19p13.3	ZBTB7		10196					4028012	no	no
4	CD3	6	51	19q12	UQCRFS1					728827		33661180	no	no
4	CD3	24	126	19q13.11	CEBPA		6560					38476216	yes	no
4	CD3	6	95	19q13.13	ZNF573		206447					43168487	no	no
4	CD3	24	114	1p13.3	VAV3		84033					108305456	no	no
4	CD3	6	158	1p34.2	HIVEP3			166304	1			41621151	no	simple repeats

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	3' of RefSeq Intron Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
4	CD3	6	56	1p34.2	CAP1			18016	1		39937757	no	SINE
4	CD3	13	142	1p36.11	RUNX3		103021				24710557	no	no
4	CD3	24	171	1p36.11	RUNX3		103050				24710586	no	no
4	CD3	6	43	1p36.11	WASF2		34931				27456288	no	LINE
4	CD3	24	70	1p36.11	LOC149420				1514		26058237	no	SINE
4	CD3	37	53	1p36.23	RERE			398401	11		8198670	no	no
4	CD3	24	119	1q23.1	MNDA		17576				156000384	no	SINE
4	CD3	24	39	1q23.2	COPA				7600		157468254	no	no
4	CD3	24	243	1q25.1	TNFSF4		357				170416204	no	no
4	CD3	6	80	1q32.1	PTPRC			16199	1		195912681	no	no
4	CD3	13	82	20p11.23	C20orf72			3614	2		17948375	no	no
4	CD3	24	82	20p11.23	SNX5		3855				17948375	no	no
4	CD3	6	20	20p13	FKBP1A			716	2	yes	1368029	no	no
4	CD3	24	39	20q13.12	STK4			27849	5		44306398	no	SINE
4	CD3	24	176	20q13.2	ZNF217		156089				53041147	no	LINE
4	CD3	24	69	20q13.2	ZNF217		9295			yes	52894353	no	no
4	CD3	24	147	21q22.3	ITGB2		5738				45202744	no	no
4	CD3	6	98	21q22.3	C21orf25			3744	1		42237170	no	no
4	CD3	6	81	22q13.31	CELSR1		65287				45210595	no	SINE
4	CD3	6	109	22q13.33	BRD-1			12853	1		48424935	no	no
4	CD3	17	245	2p14	AFTIPHILIN				174388	yes	64969058	no	no
4	CD3	6	64	2p14	AFTIPHILIN				174611	yes	64969281	no	no
4	CD3	6	44	2p16.1	BCL11A				353066		60299770	yes	no
4	CD3	6	53	2p25.1	LOC339789		158101				8648389	no	LINE
4	CD3	24	76	2q11.2	MGAT4A		33590				98997920	no	no
4	CD3	37	206	2q14.3	HS6ST1		157596				129327115	no	SINE
4	CD3	13	75	2q22.3	ARHGAP15			324700	7		144422432	no	no
4	CD3	6	51	2q31.3	ITGA4		148082				182375672	no	no
4	CD3	6	155	2q34	ZNFN1A2		928			yes	214218528	no	no
4	CD3	13	226	2q37.1	CAB39		53303				231726901	no	no
4	CD3	24	228	2q37.3	LRRFIP1			10875	1		238898554	no	no
4	CD3	37	55	3p22.3	EPM2AIP1				62826		36927028	no	LTR
4	CD3	13	179	3p23	CKLFSF8			82507	1		32334446	no	no
4	CD3	37	72	3q12.3	NFKBIZ				215460		103116226	no	no
4	CD3	24	73	3q12.3	NFKBIZ				231828		103132594	no	no
4	CD3	24	59	3q26.2	MDS1			323690	2		170378699	yes	no
4	CD3	6	150	4p14	RHOH			2280	1		40098062	yes	no
4	CD3	13	31	4p15.33	C1QTNF7			70564	1		15163059	no	LINE

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
4	CD3	13	27	4q13.3	IL8		30500					75042026	no	SINE
4	CD3	24	96	4q21.1	FLJ10849			369	1		yes	78329659	no	no
4	CD3	6	67	4q31.3	GAJ			636	1		yes	154844064	no	no
4	CD3	17	54	4q32.1	GUCY1A3	CD34 pre		17827	3			157183332	no	no
4	CD3	37	22	4q35.2	FLJ36180					632099		190396638	no	no
4	CD3	37	72	5q11.2	SNAG1					225482		54057036	no	LTR
4	CD3	17	148	5q31.2	CXXC5			23716	1			139080163	no	no
4	CD3	13	125	5q33.1	RPS14		5743					149863565	no	SINE
4	CD3	6	103	5q35.2	RNF44		1466				yes	175946810	no	no
4	CD3	24	149	5q35.3	GNB2L1		240				yes	180780798	no	no
4	CD3	13	200	6p21.33	MSH5		427				yes	31811631	no	SINE
4	CD3	13	127	6p22.3	SCA1			63768	2			16805898	no	no
4	CD3	6	35	6p22.3	SCA1			60581	2			16809085	no	no
4	CD3	17	67	6q13	C6orf155					48082		72071665	no	no
4	CD3	6	139	6q13	CD109			31676	3			74433101	no	no
4	CD3	6	31	6q21	RTN4IP1					57102		107008320	no	no
4	CD3	6	58	6q25.3	ZDHHC14			265515	6			157977358	no	no
4	CD3	24	46	7p15.3	IMP-3			123811	6			23128456	no	no
4	CD3	24	79	7p22.1	FBXL18					34771		5180533	no	no
4	CD3	24	81	7p22.1	FBXL18					5947	yes	5209257	no	no
4	CD3	37	102	7q21.3	DLX5					56633		96204810	no	no
4	CD3	24	44	7q32.2	MEST		13423					129665626	no	LINE
4	CD3	37	48	7q36.1	PRKAG2			37016	1			150928520	no	no
4	CD3	6	63	7q36.1	LR8					3278		149876853	no	no
4	CD3	24	58	7q36.3	PTPRN2			735884	10			157061389	no	no
4	CD3	24	58	7q36.3	DNAJB6		35420				yes	156511145	no	no
4	CD3	24	149	8p21.3	DOK2					50648		21737676	no	no
4	CD3	6	45	8p23.1	ANGPT2			128767	9			6405247	no	no
4	CD3	6	45	8p23.1	FLJ12847			2925	1			6405247	no	no
4	CD3	6	41	8q23.3	CSMD3		305588					114711413	no	DNA
4	CD3	24	48	8q24.22	SIAT4A			52011	2			134488740	no	SINE
4	CD3	6	87	9p13.3	NPR2		16540					35765866	no	no
4	CD3	24	63	9p21.1	B4GALT1		1900					33159131	no	LINE
4	CD3	6	77	9p24.1	PDCD1LG1		399				yes	5440160	no	no
4	CD3	6	76	9q34.13	C9orf59			2076	1			129425875 (-)	no	no
4	CD3	17	84	Xq12	MSN	CD3 13 mon		31771	1			63786178	yes	SINE
4	CD3	24	30	Xq28	CXorf12					3580		151720123	no	SINE
4	CD3	17	154	12p13.32	CCND2		27642					4225557	yes	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements	
4	CD3	6	85	12q13.2	MYL6B			1837	2		54834439	no	no	
4	CD34	pre	186	10p15.1	COPEB		43970				3825425	yes	no	
4	CD34	pre	157	10q22.1	CDH23	CD3 13 mon		356565	27		72857859	no	no	
4	CD34	pre	117	10q26.13	KIAA0140			91855	4		125915164	no	no	
4	CD34	pre	69	10q26.2	PTPRE			156670	9		129336395	no	no	
4	CD34	pre	25	11q13.2	FBXL11			839	1		66663697	no	no	
4	CD34	pre	31	11q21	MAML2			25608	1		182308602	no	no	
4	CD34	pre	109	12p13.31	CLECSF8		49766				8507637	no	DNA	
4	CD34	pre	63	12q12	SFRS2IP				195082	yes	44406377	no	no	
4	CD34	pre	38	12q13.2	KIAA0748		4526				53658507	no	no	
4	CD34	pre	177	12q21.33	ATP2B1		99641				88651953	no	SINE	
4	CD34	pre	92	13q13.1	13CDNA73			83569	3		30487006	no	no	
4	CD34	pre	48	13q13.1	13CDNA73			79211	3		30482648	no	no	
4	CD34	pre	41	13q21.32	PCDH9		209203				65811667	no	no	
4	CD34	pre	93	13q22.3	EDNRB		128549				76476214	no	no	
4	CD34	pre	45	13q22.3	PAM			514152	22		75436948	no	no	
4	CD34	pre	186	13q34	RASA3		11033				112827230	no	DNA	
4	CD34	pre	170	14q32.11	C14orf102			31383	5		88756937	no	no	
4	CD34	pre	139	14q32.12	MJD		262			yes	90563257	no	no	
4	CD34	pre	157	15q12	ATP10A		234280				23890511	no	no	
4	CD34	pre	22	15q15.1	KIAA1259			27640	3		39096757	no	no	
4	CD34	pre	52	15q21.3	TCF12			67488	1		54994377	yes	no	
4	CD34	pre	39	15q22.2	RBA-8b			42785	1		61240335	no	no	
4	CD34	pre	44	15q26.3	ALDH1A3					2169	yes	99194262	no	no
4	CD34	pre	38	16p12.3	XYLT1			129135	2		17401925	no	no	
4	CD34	pre	41	16p13.13	ATF7IP2		70530				10418740	no	LINE	
4	CD34	pre	81	16q22.1	NFAT5			1842	1	yes	69377027	no	no	
4	CD34	pre	67	16q22.3	ZNRF1			75331	1		74888129	no	LINE	
4	CD34	pre	120	16q24.2	ZFPM1		21986			yes	88244681	no	no	
4	CD34	pre	314	16q24.2	ZFPM1			4026	1		88270693	no	no	
4	CD34	pre	97	16q24.2	KLHDC4		12880			yes	87594058	no	SINE	
4	CD34	pre	94	17p13.1	GAS7			43461	1		10258973	yes	no	
4	CD34	pre	908	17p13.1	GAS7			43465	1		10258969	yes	SINE	
4	CD34	pre	22	17q21.32	MYL4			11139	3		45772491	no	no	
4	CD34	pre	65	17q25.1	TREM5					969	73113468	no	LINE	
4	CD34	pre	95	18p11.22	RAB31			120520	5		9818868	no	no	
4	CD34	pre	213	18q11.1	ROCK1		52503				16995361	no	no	
4	CD34	pre	117	18q12.3	CD33L3			1671	1		41659203	no	SINE	

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	3' of RefSeq Intron Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
4	CD34	pre	55	18q21.1	PSTPIP2			58364	4		41845846	no	no
4	CD34	pre	66	19p13.2	TRMT1			11384	16	yes	13076948	no	no
4	CD34	pre	122	19p13.3	RFX2			10956	1		6050598	no	no
4	CD34	pre	65	1p13.3	SARS			813	1	yes	109056080	no	no
4	CD34	pre	96	1p31.1	PRKACB				65289		84192061	no	no
4	CD34	pre	53	1p32.3	OSBPL9			113108	4	yes	51565790	no	no
4	CD34	pre	203	1p34.1	LOC51249			2082	1		45569232	no	SINE
4	CD34	pre	54	1q21.1	MGC46719		1151				143224620	no	no
4	CD34	pre	18	1q23.2	NCSTN			3984	2		157533838	no	no
4	CD34	pre	25	1q25.3	C1orf24			105046	5		182077979	no	no
4	CD34	pre	156	1q32.1	NUCKS				14577		202849271	no	DNA
4	CD34	pre	88	1q32.1	NUCKS				14645		202849203	no	SINE
4	CD34	pre	123	1q32.2	CD34	CD3 24 mon	7003				205175217	no	no
4	CD34	pre	25	1q42.3	IRF2BP2	CD3 13 mon	2896				231832436	no	no
4	CD34	pre	71	1q44	SDCCAG8			31862	5		240398062	no	no
4	CD34	pre	47	20p11.21	SSTR4		112869				22898252	no	no
4	CD34	pre	54	20p11.23	INSM1				266810		20613400	no	no
4	CD34	pre	32	20p12.2	BTBD3			2247	1		11868724	no	no
4	CD34	pre	120	20q11.21	FLJ33706				4619		31961043	no	no
4	CD34	pre	48	21q11.2	SAMSN1		4243				14775183	no	no
4	CD34	pre	109	2p11.2	FLJ90024		711			yes	85805072	no	no
4	CD34	pre	40	2p23.2	BRE			63971	3		28151988	no	no
4	CD34	pre	90	2q12.2	FHL2		39808				105716168	no	SINE
4	CD34	pre	27	2q14.3	TSN				36917		122657491	no	no
4	CD34	pre	29	2q21.3	DKFZP566N034			5888	2		135681454	no	no
4	CD34	pre	121	3p14.1	JWA			11167	1		69076347	no	no
4	CD34	pre	60	3p21.31	LOC389119		16195			yes	49783504	no	SINE
4	CD34	pre	94	3p25.1	MGC15763			22907	4		16304651	no	no
4	CD34	pre	179	3q21.2	MUC13		39025				126013506	no	no
4	CD34	pre	185	3q25.2	MME		69543				156048829	no	no
4	CD34	pre	88	3q27.2	MAGEF1		73991				185824694	no	no
4	CD34	pre	54	4q22.1	LOC285513		8694				90696032	no	no
4	CD34	pre	114	5p13.2	FLJ23577		142005				35521485	no	LINE
4	CD34	pre	77	5p15.2	TRIO			62331	1		14259015	no	SINE
4	CD34	pre	324	5q11.1	FLJ21308				11908		50165969	no	no
4	CD34	pre	59	5q35.1	LOC153222			27408	2		172491701	no	no
4	CD34	pre	147	5q35.1	LCP2			1168	1		169704549	no	no
4	CD34	pre	44	6p12.2	MCM3		33834				52230252	no	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
4	CD34	pre	68	6p22.1	HIST1H2BK					917		27213141	no	LTR
4	CD34	pre	49	6q11.1	KHDRBS2			248991	3			62743977	no	no
4	CD34	pre	47	6q22.31	C6orf170			29088	1			121607114	no	SINE
4	CD34	pre	202	6q24.1	CITED2		265476					139941831	no	SINE
4	CD34	pre	205	6q26	PARK2			898628	6			162159599	no	no
4	CD34	pre	177	6q26	PACRG					19961		163665772	no	no
4	CD34	pre	31	7p14.1	ELMO1			106827	2			37122424	no	no
4	CD34	pre	97	7p21.1	PRPS1L1		196312					18008114	no	no
4	CD34	pre	67	7p22.3	C7orf27					7519		2314334	no	no
4	CD34	pre	132	7q11.21	ZNF679		303600					62796942	no	no
4	CD34	pre	34	7q11.23	FKBP6			211543	1			72063943	no	no
4	CD34	pre	34	7q22.1	CYP3A7		5492					98949710	no	no
4	CD34	pre	76	7q22.1	FBXL13			145273	8			102130273	no	no
4	CD34	pre	76	7q22.1	LRRRC17			16290	1			102130273	no	no
4	CD34	pre	102	7q22.3	LHFPL3	CD3 13 mon				37322		104145532	no	no
4	CD34	pre	164	7q32.2	MEST		32133					129666999	no	no
4	CD34	pre	33	8p12	ASH2L		15271					37965160	no	LTR
4	CD34	pre	166	8p21.2	TRIM35			1056	1		yes	27189688	no	no
4	CD34	pre	39	8q11.23	ST18		56738					53429001	no	no
4	CD34	pre	33	8q24.3	FLJ31164			35212	1			143349110	no	LINE
4	CD34	pre	113	9p13.3	GBA2		7613				yes	35746838	no	no
4	CD34	pre	144	9p21.3	MLLT3			221841	4			20390609	yes	LINE
4	CD34	pre	172	9p24.1	GLDC		36184				yes	6671834	no	SINE
4	CD34	pre	51	9q13	FOXD4b					370870		65792831	no	no
4	CD34	pre	274	9q34.13	C9orf59			7327	1			129420624	no	SINE
4	CD34	pre	28	Xp22.11	ZFX					43906		23635149	no	LINE
4	CD34	pre	31	Xp22.33	PR48			14253	1		yes	275553	no	no
4	PBL	41	227	1q21.3	TSRC1					1551	yes	147751543	no	no
4	PBL	41	32	8q23.3	TRSP1			18865	1			116618975	no	no
5	CD3	13	60	12p13.2	CLEC9A	CD3 31 mon		26043	4			10100586	no	no
5	CD3	13	184	12p13.31	SLC2A14	CD3 31 mon	13937					7930699	no	no
5	CD3	13	59	12q24.33	GALNT9	CD3 31 mon	75881				yes	131260903	no	no
5	CD3	13	62	16q23.1	MAF		167887					79582025	yes	no
5	CD3	13	206	17q21.31	MAP3K14		24924				yes	43868686	no	no
5	CD3	13	78	1p36.31	RPL22	CD3 31 mon	564				yes	5969765	yes	LINE
5	CD3	13	106	20q13.2	ZNF217	CD3 31 mon	156635					53041693	no	LINE
5	CD3	13	188	22q12.3	APOL2		154					34907967	no	no
5	CD3	13	231	2q36.2	DOCK10		196953					226098405	no	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
5	CD3	13	187	6p12.3	C6orf141					5003		49581464	no	no
5	PBL	37	91	10p11.21	ANKRD30A		590500					36828381	no	no
5	PBL	37	46	10q22.1	CDH23			352427	32			72853721	no	SINE
5	PBL	37	74	11p11.2	C1QTNF4			1929	1			47578295	no	no
5	PBL	31	74	11p13	LMO2	CD3 13 mon	2965					33881110	yes	SINE
5	PBL	37	247	11q13.4	ARHGEF17			14650	1			72760627	no	SINE
5	PBL	37	21	11q24.3	ETS1			13954	1			127915859	no	SINE
5	PBL	37	55	12p13.32	CCND2		1910					4251268	yes	no
5	PBL	37	93	12p13.32	CCND2		1834				yes	4251301	yes	no
5	PBL	37	32	12q22	SOCS2					19207		92491653	no	no
5	PBL	37	34	13q14.11	ELF1		36579					39390997	no	no
5	PBL	37	75	13q14.13	LCP1		18748					44573143	yes	LINE
5	PBL	37	31	13q14.3	FLJ13639		21241					50197535	no	LINE
5	PBL	31	26	14q13.2	PSMA6		4319					33747269	no	no
5	PBL	37	96	14q24.3	C14orf4		10156					75495231	no	simple repeats
5	PBL	37	84	14q24.3	C14orf4					22474		75458455	no	SINE
5	PBL	37	21	15q21.1	FBN1					27024		46391219	no	no
5	PBL	37	123	16p12.1	NSMCE1			13569	1			27221831	no	LINE
5	PBL	37	70	16q13	GPR114			1391	1			57353389	no	no
5	PBL	37	59	17q21.2	STAT5A			108	1			40812838	no	no
5	PBL	37	46	18q21.1	SMAD2		68840					43777753	no	LTR
5	PBL	37	71	1p13.3	VAV3		86003					107948215	no	no
5	PBL	37	56	1p31.2	PDE4B			558067	7			66186848	no	no
5	PBL	37	26	1p36.11	DSCR1L2			1030	1		yes	24181585	no	no
5	PBL	37	229	1q21.3	MCL1		1060					147802191	no	SINE
5	PBL	37	140	1q32.1	PTPRC			14539	1			195911021	no	no
5	PBL	37	78	1q44	FLJ20531		6419					245974598	no	LINE
5	PBL	37	106	20p11.21	GGTLA4		37058					24001474	no	no
5	PBL	37	83	2p16.1	FANCL		519117					58962165	no	no
5	PBL	37	222	2q31.1	PPIG		119				yes	170643275	no	no
5	PBL	37	23	2q37.1	SPP2		169592					235434999	no	LINE
5	PBL	37	54	3p14.1	JWA		1512					69065668	no	no
5	PBL	37	53	3q25.1	MBNL1		67499					153239239	no	SINE
5	PBL	37	60	3q25.31	TIPARP		123010					157690282	no	SINE
5	PBL	31	276	3q26.2	MDS1			329842	2			170392640	yes	no
5	PBL	37	157	4q25	FLJ39370		3201					113522297	no	SINE
5	PBL	37	62	5p13.1	PTGER4		267879					40457654	no	LTR

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
5	PBL	37	49	5p13.1	PTGER4		106747					40618786	no	no
5	PBL	37	86	5q23.3	IRF1					19121		131875880	no	no
5	PBL	37	30	5q23.3	SLC22A5					17472		131824991	no	no
5	PBL	37	207	7q21.2	SAMD9		126					92359477	no	no
5	PBL	37	204	9q33.3	NEK6			28331	1			122424625	no	no
5	PBL	37	39	Xp11.23	SLC38A5			1541	1			47373041	no	no
6	PBL	16	466	11q22.3	RDX					201924		109486159	no	no
6	PBL	4	55	12p13.1	ATF7IP	PBL 16 mon	108193					14301685	no	no
6	PBL	16	25	12q14.11	MGC10854		774				yes	108800269	no	DNA
6	PBL	16	384	14q23.1	ACTR10					4252		56695801	no	no
6	PBL	16	64	14q24.2	RGS6			506146	3			70896020	no	LINE
6	PBL	4	22	15q26.2	MCTP2	PBL 16 mon	68796					92502467	no	no
6	PBL	4	161	16p12.2	PRKCB1	PBL 16 mon		45178	2			23859090	no	no
6	PBL	16	74	19p13.11	BST2		1238					17378622	no	no
6	PBL	16	240	19p13.3	OAZ1		3109					2217411	no	no
6	PBL	16	46	1q21.2	TSRC1					2849		147752841	no	no
6	PBL	16	44	22q12.2	MTP18		872				yes	29145402	no	no
6	PBL	16	198	2p13.1	CML2		472					73903477	no	no
6	PBL	16	126	2p21	PRKCE			205227	2			46058804	no	no
6	PBL	16	288	2p21	ZFP36L2					29310		43395931	no	no
6	PBL	16	27	2q33.1	CFLAR			1972	1		yes	202185405	no	no
6	PBL	16	94	2q37.1	MGC35154					18823	yes	232680354	no	no
6	PBL	4	131	3p25.2	RAF1	PBL 16 mon	477				yes	12681155	no	no
6	PBL	16	63	3p26.1	GRM7			429829	2			7307756	no	no
6	PBL	16	91	4q12	SPINK2			705	1		yes	57602730	no	no
6	PBL	16	52	8q24.22	SIAT4A			74390	3			134466361	no	SINE
6	PBL	4	76	9q21.11	FXN					46504	yes	67192575	no	LINE
6	PBL	4	101	9q34.11	SET		3742					126824164	no	SINE
6	Tgd-	16	145	14q13.1	EGLN3	PBL 16 mon	126864					32537189	no	no
7	PBL	11	125	10p12.33	SIAT8F			25669	2			17474591	no	LINE
7	PBL	11	143	10p15.1	COPEB		26808					3808267	yes	no
7	PBL	11	244	11p11.2	DGKZ			5790	1			46324838	no	no
7	PBL	11	24	11q12.1	SERPING1			6562	4			57147080	no	no
7	PBL	11	71	11q13.5	PAK1			53904	1			76857337	no	no
7	PBL	11	177	12p12.1	KRAS2		67141					25362262	yes	no
7	PBL	11	48	12p13.31	MFAP5		30424					8737124	no	no
7	PBL	11	91	12p13.32	CCND2		2273				yes	4250926	yes	no
7	PBL	11	24	12q14.1	FLJ32949					149447		62089513	no	LTR

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
7	PBL	11	106	13q21.32	PCDH9					101327		64573641	no	no
7	PBL	11	200	13q21.33	DACH1			40037	1			70199294	no	no
7	PBL	11	28	13q31.3	C13orf25		1385				yes	89696690	no	no
7	PBL	11	62	14q21.3	C14orf138					62984		48502410	no	no
7	PBL	11	77	15q13.3	C15orf16					44720		29446665	no	LTR
7	PBL	11	35	15q25.1	KIAA1055			26268	2			76059408	no	no
7	PBL	11	64	15q26.1	DKFZp547K1113			35509	1			88310029	no	no
7	PBL	11	74	16p12.3	XYLT1		4183					17535243	no	no
7	PBL	11	99	16q13	DKFZp434I099			2455	1			57506578	no	no
7	PBL	11	39	16q23.2	PLCG2			53195	2			81645810	no	SINE
7	PBL	11	49	17p13.1	EIF5A			859	1		yes	7412393	no	no
7	PBL	11	50	17p13.1	LOC92162		17432					7941518	no	SINE
7	PBL	11	21	17q21.2	CCR7		14667					39109556	no	no
7	PBL	11	28	17q21.31	FMNL1			2182	1			43776896	no	no
7	PBL	11	74	17q21.32	PNPO			1129	1			46494704	no	no
7	PBL	11	21	17q23.2	MRPS23		491				yes	56402528	no	no
7	PBL	11	141	17q23.2	VMP1			77942	5			58337392	no	no
7	PBL	11	66	17q25.3	ACTG1					56718	yes	80120400	no	no
7	PBL	11	55	18q11.2	CABLES1			73246	4			19061022	no	no
7	PBL	11	51	19p13.13	CD97		13690					14339576	no	SINE
7	PBL	11	133	19p13.3	EMR1			2098	1			6840680	no	SINE
7	PBL	11	27	19q13.12	SCN1B			9118	5			40222492	no	no
7	PBL	11	86	19q13.2	SIRT2					61		44060979	no	no
7	PBL	11	27	19q13.2	R29124_1		30013					46744414	no	no
7	PBL	11	77	1p31.1	LPHN2					169360		82049833	no	no
7	PBL	11	77	1p34.1	AKR1A1		5587			19223		45424205	no	no
7	PBL	11	54	1p34.2	CAP1		1542				yes	39918180	no	no
7	PBL	11	92	1p36.12	ELA3A					14020		21822518	no	no
7	PBL	11	289	1q21.2	TSRC1					6791		147756783	no	no
7	PBL	11	267	1q22	LMNA					7980		153334643	no	no
7	PBL	11	60	1q23.1	IRTA2					112730		154587229	no	no
7	PBL	11	117	1q23.2	TAGLN2					5495		157099200	no	no
7	PBL	11	51	1q32.1	PTPRC			14743	1			195911225	no	no
7	PBL	11	39	1q32.3	TRAF5		1328					208578101	no	SINE
7	PBL	11	139	1q42.12	ITPKB			60401	1			223839425	no	no
7	PBL	11	40	20q13.2	ZNF217		122993					53008051	no	no
7	PBL	11	190	2p13.1	DGUOK					11624		74172250	no	no
7	PBL	11	97	2p13.3	FLJ20558					7838	yes	70343715	no	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
7	PBL	11	65	2p13.3	PCBP1					4733		70295599	no	no
7	PBL	11	185	2p14	AFTIPHILIN					155688		64950358	no	no
7	PBL	11	174	2p14	MEIS1			92283	7			66729349	no	no
7	PBL	11	27	2q21.1	PLEKHB2		242				yes	132073436	no	no
7	PBL	11	230	2q37.1	PTMA		44248					232731527	no	no
7	PBL	11	48	3p13	FOXP1			234077	3			71279018	no	no
7	PBL	11	106	3p22.3	ITGA9			41735	~5			37496173	no	no
7	PBL	11	38	3p24.1	SIMP		310943					31235316	no	no
7	PBL	11	55	3p25.3	OGG1			1536	1		yes	9767241	no	no
7	PBL	11	31	3q13.13	ESRRBL1		455				yes	109262604	no	no
7	PBL	11	107	3q26.2	MDS1			260412	1			170441975	yes	no
7	PBL	11	110	4q25	ELOVL6					16036		111413327	no	no
7	PBL	11	62	4q35.1	IRF2					42153		185962652	no	SINE
7	PBL	11	323	5q23.1	TNFAIP8		23357					118744589	no	no
7	PBL	11	107	5q34	MAT2B					839117		163766338	no	LTR
7	PBL	11	73	6p22.3	IBRDC2			13887	2			18509460	no	no
7	PBL	11	97	6p22.3	MYLIP			1598	2		yes	16238894	no	no
7	PBL	11	76	6p25.1	LY86			65473	4			6599406	no	LINE
7	PBL	11	156	6q15	BACH2			109148	4			90892911	no	no
7	PBL	11	54	7p14.1	ELMO1			39869	1			37189382	no	LINE
7	PBL	11	67	7p14.2	KIAA1706			51729	1			35985741	no	no
7	PBL	11	81	7p14.3	PDE1C		205919					32059182	no	no
7	PBL	11	28	7q22.1	CUTL1		74418					100945055	no	SINE
7	PBL	11	60	8p23.1	UNQ2754		450				yes	6680897	no	no
7	PBL	11	29	8q23.1	ANGPT1			152032	2			108314805	no	no
7	PBL	11	140	8q24.3	PTP4A3		18344					142384719	no	DNA
7	PBL	11	130	9q22.33	COL15A1			113433	10			97199461	no	no
7	PBL	11	46	9q31.2	ZNF462		507010					104562373	no	no
7	PBL	11	72	9q32	AKNA			6557	1			112526173	no	no
7	PBL	11	91	9q33.2	GSN		5570					119400882	no	no
7	PBL	11	93	Xp22.12	RPS6KA3		169975					19816151	no	LINE
7	PBL	11	28	Xq27.3	SLITRK4		624104					142040725	no	no
7	Tgd-	16	231	11p15.4	MRV11			54673	1			10625171	no	no
7	Tgd-	16	91	11q13.1	SCYL1		46828				yes	65021115	no	SINE
7	Tgd-	16	74	12p13.2	ETV6			201743	3			11895798	yes	no
7	Tgd-	16	110	13q21.33	DACH1			177229	1			70062103	no	no
7	Tgd-	16	127	16p13.13	MHC2TA			1872	1		yes	10938769	yes	no
7	Tgd-	16	80	17q12	MLLTG		1264				yes	37235506	yes	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
7	Tgd-	16	97	17q21.33	TOB1	CD3 16 mon	3983				yes	49420034	no	no
7	Tgd-	16	111	1q42.3	TOMM20					132916		232180893	no	no
7	Tgd-	16	365	2p13.3	FLJ20558	PBL 11 mon				17292		70334261	no	no
7	Tgd-	16	76	2p21	ZFP36L2	PBL 11 mon				7638		43417603	no	SINE
7	Tgd-	16	90	2p22.2	CYP1B1		1438				yes	38279234	no	no
7	Tgd-	16	26	6p21.1	SUPT3H	PBL 11 mon		53610	2			45338915	no	no
7	Tgd-	16	116	6q24.2	HIVEP2		67688					143144240	no	no
7	Tgd-	16	70	7q34	PRSS1		185685					141712027	no	LTR
7	Tgd-	16	65	9q22.32	PTPDC1					47910		92261455	no	no
8	CD3	10	76	11p13	LMO2		35751					33913896	yes	no
8	CD3	10	23	18p11.31	LPIN2		25430					3027375	no	no
8	CD3	10	74	18q11.2	IMPACT		753				yes	20257844	no	no
8	CD3	10	263	19p13.3	MKNK2		9206				yes	2011078	no	no
8	CD3	10	82	1p32.3	PCSK9					146812		55047253	no	SINE
8	CD3	10	200	20q13.2	SALL4					87067		50998940	no	no
8	CD3	10	34	2p15	B3GNT1					77060		62503458	no	LINE
8	CD3	10	58	3p24.1	TGFBR2			52908	3			30675911	no	no
8														unknown
	CD3	10	53	3q29	FLJ35155			124135	3			196187238	no	repeat
8	CD3	10	21	4q31.21	INPP4B			275236	11			143654923	no	no
8	PBL	10	101	10q11.23	TMEM23			296369	6			51431971	no	no
8	PBL	10	57	10q22.2	C10orf11		44075					76843047	no	no
8	PBL	10	58	11p15.4	STIM1		1772					3839470	no	LINE
8	PBL	10	95	11p15.4	STIM1			38515	1			3879757	no	no
8	PBL	10	36	11q13.5	PAK1			36606	1			76874635	no	no
8	PBL	10	28	11q22.1	PORIMIN			2047	1			101859240	no	SINE
8	PBL	10	103	12p13.33	NINJ2			34891	1			608125	no	LTR
8	PBL	10	50	13q14.2	CHC1L			25090	10			46880227	no	no
8	PBL	10	55	13q14.3	FLJ11712		14929					49266999	no	no
8	PBL	10	231	14q11.2	KIAA0323		8423					22880881	no	no
8	PBL	10	21	14q21.3	RPS29					26691		48013650	no	no
8	PBL	10	51	14q24.3	C14orf166B					49957		75375637	no	no
8	PBL	10	149	15q13.3	KLF13		1817				yes	29333322	no	no
8	PBL	10	55	15q21.2	MGC35274		2283					49748655	no	LINE
8	PBL	10	83	15q24.1	SCAMP2			4323	1			72877164	no	SINE
8	PBL	10	40	17p11.2	DKFZp566O084		1054				yes	21191233	no	LINE
8	PBL	10	37	17p13.2	MGC29671					4453		4602541	no	SINE
8	PBL	10	126	17p13.3	C17orf31			38365	8			2375245	no	SINE

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
8	PBL	10	42	17q23.2	YPEL2					15590		57964812	no	no
8	PBL	10	53	17q24.2	DKFZP586L0724		499				yes	66263663	no	no
8	PBL	10	38	17q25.3	ACTG1					56689	yes	80120429	no	no
8	PBL	10	157	18p11.31	LPIN2			39071	1			2962874	no	no
8	PBL	10	37	19p13.2	FARSLA		186				yes	12905708	no	no
8	PBL	10	50	19p13.3	EMR1			2736	1			6841318	no	no
8	PBL	10	71	19q12	POP4		827				yes	34788214	no	SINE
8	PBL	10	35	1p31.2	MI-ER1			7922	3			66768480	no	no
8	PBL	10	103	1p36.11	RUNX3		105496					24713032	no	no
8	PBL	10	36	1q21.2	CTSS			1655	2			147953232	no	LINE
8	PBL	10	27	1q22	EST1B			421	1			153468987	no	no
8	PBL	10	29	1q22	IL6R		2262					151592199	no	no
8	PBL	10	54	1q42.3	TOMM20					128939		232184870	no	no
8	PBL	10	46	1q42.3	APID4B			52970	1			232479679	no	no
8	PBL	10	149	1q44	HNRPU		391				yes	241974942	no	no
8	PBL	10	33	20q11.21	BCL2L1		252				yes	31055908	no	no
8	PBL	10	59	20q13.12	PRKCBP1			3207	1			46667689	no	no
8	PBL	10	169	20q13.12	PRKCBP1		3428					46674324	no	no
8	PBL	10	137	20q13.2	ZNF217		71903					52956961	no	no
8	PBL	10	35	20q13.2	ZNF217		169521					53054579	no	no
8	PBL	10	113	20q13.2	ZNF217		36936					52921994	no	no
8	PBL	10	340	21q11.2	SAMSN1					3615		14775805	no	no
8	PBL	10	57	21q21.1	NRIP1		95122				yes	15357792	no	no
8	PBL	10	35	21q21.1	NCAM2		793760					20498744	no	no
8	PBL	10	208	22q12.3	MYH9		2355					35058149	yes	SINE
8	PBL	10	57	2p13.1	DGUOK					77989		74238615	no	no
8	PBL	10	35	2p14	MEIS1			15725	6			66652791	no	no
8	PBL	10	95	2q22.3	ZFXH1B		182866					145671551	no	DNA
8	PBL	10	235	2q33.1	LOC91526		26763					198292068	no	no
8	PBL	10	111	3p21.2	FLJ32332			283	1		yes	52238938	no	no
8	PBL	10	151	3p21.2	SRISNF2L		6626					51614139	no	no
8	PBL	10	159	3p22.1	FLJ14566		74560					43182717	no	no
8	PBL	10	322	3p25.2	IQSEC1		99159					13083119	no	no
8	PBL	10	66	3q26.31	FAD104			33841	1			173113140	no	no
8	PBL	10	84	5p13.1	RICTOR			3362	2			39116640	no	no
8	PBL	10	57	5p13.3	TARS		228936					33257463	no	no
8	PBL	10	210	6p21.31	MAPK14			5112	1			36047540	no	no
8	PBL	10	268	6p22.3	SOX4					237681		21944507	no	no

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	Intron	3' of RefSeq Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
8	PBL	10	138	6p22.3	ATXN1		78433					16948133	no	no
8	PBL	10	62	7p15.2	SCAP2					1447		26447515	no	no
8	PBL	10	146	7q11.23	KIAA1505			71156	2			76435174	no	no
8	PBL	10	30	8q13.2	FLJ11267					4673		67677869	no	no
8	PBL	10	99	9q21.13	GCNT1		5815					74525450	no	LTR
8	PBL	10	72	9q32	AKNA			10938	1			112521792	no	SINE
8	PBL	10	172	9q33.2	NDUFA8			1377	1			120296766	no	no
8	PBL	10	118	9q33.3	NEK6			33354	1			122429648	no	no
8	PBL	10	101	Xp11.3	CHST7			3786	1			45482960	no	no
8	PBL	10	89	Xp22.11	EIF2S3		28932				yes	23405559	no	no
8	PBL	10	60	Xp22.22	FAM9C		42068					12466294	no	no
8	PBL	10	54	Xp22.33	SHOX					216106		770603	no	no
8	PBL	10	63	Xq24	SEPT6			11483	1			117569952	yes	SINE
9	PBL	8	50	11q23.3	BCL9L	PBL 12 mon	3011					118322276	no	no
9	PBL	10	27	15q15.2	C15orf25	PBL 6 mon, 12 mon				37221		40613319	no	no
9	PBL	12	36	16q23.1	ADAT1	PBL 6 mon		1381	2			75435326	no	no
9	PBL	6	34	17q21.2	SUI1	PBL 10 mon, 12 mon	10015					40208303	no	SINE
9	PBL	4	451	2p23.3	TP53I3			525	1		yes	24281738	no	no
9	PBL	6	81	2q12.3	MGC10701			902	1		yes	108771150	no	no
9	PBL	7	92	3p13	FLJ10539			79473	10			72739026	no	LINE
9	PBL	6	75	9q33.2	STOM	PBL 7 mon		2594	1			119505996	no	no
9	PBL	10	113	Xq24	SEPT6			18967	2			117567502	yes	LINE
10	CD3	5	207	15q24.2	MDS018					59580		73118463	no	SINE
10	CD3	5	55	16q22.1	MGC11335			440	1		yes	67529017	no	no
10	CD3	5	99	1p36.11	WASF2			1803	1			27419554	no	no
10	CD3	5	20	20q13.2	ATP9A			71704	5			50998626	no	no
10	CD3	5	166	21q22.13	DYRK1A			3280	1			37663535	no	no
10	PBL	12	174	10p13	DNMT2		15478					17263165	no	SINE
10	PBL	12	37	13q21.33	DACH1			257087	3			69982244	no	no
10	PBL	12	27	16q22.1	TERF2		602				yes	69196663	no	no
10	PBL	12	70	16q22.2	DHODH		561				yes	71818378	no	SINE
10	PBL	12	206	17q21.2	STAT3			68661	18			40845017	no	no
10	PBL	12	40	19p13.2	LYL1					2012	yes	13068836	yes	no
10	PBL	12	53	1p31.2	PDE4B			508633	6			66137414	no	no
10	PBL	12	80	20q13.2	ZNF217		72855					52957913	no	no
10	PBL	12	24	3p12.3	ROBO1		58138					79618100	no	DNA

Supplementary Table 1

Pat.	Source	Time	Length	Chromo- some	RefSeq Gene	Identity	5' of TSS [bp]	3' of TSS [bp]	3' of RefSeq Intron Gene [bp]	CpG Island (1kb)	UCSC Locus	CGP	Repetitive Elements
10	PBL	12	87	3p13	FOXP1			484067	6		71029018	no	no
10	PBL	12	265	4q23	DAPP1			40034	3		101236418	no	LTR
10	PBL	12	82	5p13.1	RICTOR		988			yes	39120990	no	no

Supplementary Table 2

CIS	No. of CIS	Mean Value	Median	Minimum	Maximum	Standard Deviation	p-Value
2nd Order	102	3.2	3	0	12	1.79	0
3rd Order	81	0	0	0	0	0	0
4th Order	147	0	0	0	0	0	0
5th Order	132	nd	nd	nd	nd	nd	nd
6th Order	57	nd	nd	nd	nd	nd	nd
7th Order	28	nd	nd	nd	nd	nd	nd
8th Order	9	nd	nd	nd	nd	nd	nd
9th Order	1	nd	nd	nd	nd	nd	nd

Supplementary Table 3

CIS	IS	Locus	Next RefSeq Gene	Location	Patient	UCSC locus	A 1	A 2	A 3
1	9	12p13.32	CCND2: Cyclin D2	1834 bp up	P5	4251301	8	8	8
				1910 bp up	P5	4251225			
				2116 bp up	P1	4251083			
				2138 bp up	P4	4251016			
				2273 bp up	P7	4250926			
				2494 bp up	P4	4250705			
				3170 bp up	P1	4249921			
				57826 bp up	P2	4195373			
				27642 bp up	P4	4225557			
2	8	20q13.2	ZNF217: Zinc Finger Protein 217	169521 bp up	P8	53054579	7	7	6
				156635 bp up	P5	53041693			
				156089 bp up	P4	53041147			
				122993 bp up	P7	53008051			
				72855 bp up	P10	52957913			
				71903 bp up	P8	52956961			
				36936 bp up	P8	52921994			
				9295 bp up	P4	52894353			
3	5	11p13	LMO2: LIM Domain Only 2	Intron 1	P4	33876126	9	9	9
				1260 bp up	P1	33879405			
				2965 bp up	P5	33881110			
				35753 bp up	P8	33913896			
				39519 bp up	P4	33917663			
4	4	14q24.3	C14orf4: Chromosome 14 Open Reading Frame 4	1961 bp up	P2	75487036	7	7	8
				10156 bp up	P5	75495231			
				22474 bp down	P5	75458455			
				68330 bp down	P4	75412599			
5	3	1p11.2	Notch 2: Neurogenic Locus Notch Homolog Protein 2 Precursor	Intron 2	P1	119898329	8	8	8
				Intron 2	P1	119902435			
				Intron 2	P2	119905094			
6	3	1p31.2	PDE4B: CAMP specific 3',5'-cyclic Phosphodiesterase 4B	Intron 6	P1	66166317	4	3	3
				Intron 7	P10	66137414			
				Intron 7	P5	66186848			
7	3	1p36.11	RUNX3: Runt-related Transcription Factor 3	103021 bp up	P4	24710557	8	8	7
				103050 bp up	P4	24710586			
				105496 bp up	P8	24713032			
8	3	1q21.3	TSRC1: Thrombospondin Repeat Containing 1	1553 bp down	P4	147751543	5	5	4
				2851 bp down	P6	147752841			
				6793 bp down	P7	147756783			
9	3	1q32.1	PTPRC: Protein Tyrosine Phosphatase ReceptorType C	Intron 1	P4	195912681	9	9	9
				Intron 1	P5	195911021			

Supplementary Table 3

CIS	IS	Locus	Next RefSeq Gene	Location	Patient	UCSC locus	A 1	A 2	A 3
10	3	1q42.3	TOMM20: Translocase of Outer Mitochondrial Membrane 20	Intron 1	P7	195911225	9	9	9
				148444 bp down	P1	232165365			
				132916 bp down	P7	232180893			
				128939 bp down	P8	232184870			
11	3	2p14	AFTIPHILIN: Hypothetical Protein	174388 bp down	P4	64969058	8	8	8
				174611 bp down	P4	64969281			
				155688 bp down	P7	64950358			
				2028 bp up	P2	46672924			
12	3	20q13.12	PRKCBP1:Protein Kinase C Binding Protein 1 Isoform c	Intron 1	P8	46667689	8	9	9
				3428 bp up	P8	46674324			
				Intron 1	P1	35321210			
				Intron 1	P2	35303949			
13	3	21q22.12	RUNX1: Runt-related Transcription Factor 1	Intron 1	P2	35318998	9	9	9
				46404 bp up	P2	12470630			
				46611 bp up	P2	12470837			
				42068 bp up	P8	12466294			
14	3	Xp22.22	FAM9C: Family with Sequence Similarity 9 Member C	84033 bp up	P4	108305456	0	1	1
				86003 bp up	P5	107948215			
				Intron 1	P4	39937757			
				1542 bp up	P7	39918180			
15	2	1p13.3	VAV3: VAV 3 Oncogene	17201 bp up	P2	31923629	9	9	9
				23527 bp up	P2	31929955			
				17292 bp down	P7	70334261			
				7838 bp down	P7	70343715			
16	2	1p34.2	CAP1: Adenylyl Cyclase-associated Protein	29310 bp down	P6	43395931	9	9	9
				7638 bp down	P7	43417603			
				215460 bp down	P4	103116226			
				231828 bp down	P4	103132594			
17	2	1p35.2	PTP4A2: Protein Tyrosine Phosphatase Type IVA Member 2	1724 bp up	P2	109263948	8	7	7
				455 bp up	P7	109262604			
				Intron 2	P4	170378699			
				Intron 2	P5	170392640			
18	2	2p13.3	FLJ20558: Hypothetical Protein	988 bp up	P10	39120990	7	7	7
				Intron 2	P8	39116640			
				868 bp up	P1	139055579			
				Intron 1	P4	139080163			
19	2	2p21	ZFP36L2: Butyrate Response Factor 2	31416 bp up	P1	42033699	5	5	5
				33936 bp up	P2	42031179			
				Intron 2	P4	16805898			
				Intron 2	P4	16809085			
20	2	3q12.3	NFKBIZ: Nuclear Factor of Kappa Light Polypeptide Gene	Intron 2	P4	16809085	n.p.	n.p.	n.p.
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
21	2	3q13.13	ESRRBL1: Estrogen-related Receptor Beta Like 1	Intron 2	P4	16809085	n.p.	n.p.	n.p.
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
22	2	3q26.2	MDS1: Myelodysplasia Syndrome Protein 1	Intron 2	P4	16809085	n.p.	n.p.	n.p.
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
23	2	5p13.1	RICTOR	Intron 2	P4	16809085	n.p.	n.p.	n.p.
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
24	2	5q31.2	CXXC5: CXXC finger 5	Intron 2	P4	16809085	n.p.	n.p.	n.p.
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
25	2	6p21.1	TBN: Taube Nuss	Intron 2	P4	16809085	n.p.	n.p.	n.p.
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
26	2	6p22.3	SCA1: Ataxin 1	Intron 2	P4	16809085	n.p.	n.p.	n.p.
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			
				Intron 2	P4	16809085			

Supplementary Table 3

CIS	IS	Locus	Next RefSeq Gene	Location	Patient	UCSC locus	A 1	A 2	A 3
27	2	7p22.1	FBXL18: F-box and Leucine-rich Repeat Protein 18	34771 bp down	P4	5180533	5	5	5
				5947 bp down	P4	5209257			
28	2	8q24.22	SIAT4A: Sialyltransferase 4A	Intron 2	P4	134488740	2	4	3
				Intron 3	P6	134466361			
29	2	9q32	AKNA: AT-hook Transcription Factor	Intron 1	P7	112526173	7	6	6
				Intron 1	P8	112521792			
30	2	9q33.2	GSN: Gelsolin Isoform b	Intron 1	P1	119409501	7	7	8
				5570 bp up	P7	119400882			
31	2	9q33.3	NEK6: Putative Serine-threonine Protein Kinase	Intron 1	P5	122424625	9	8	8
				Intron 1	P8	122429648			
32	2	9q34.11	CDK9: Cyclin-dependant Kinase 9	2155 bp up	P2	125922227	4	4	5
				1728 bp up	P2	125922654			
33	2	10q22.1	PALD: Paladin	Intron 1	P4	71587730	n.p.	n.p.	n.p.
				Intron 1	P4	71602160			
34	2	11q12.1	Hypothetical Protein MGC15937	53626 bp down	P4	58553502	n.p.	n.p.	n.p.
				54091 bp down	P4	58553967			
35	2	11q13.5	PAK1: p21-activated Kinase 1	Intron 1	P7	76857337	7	7	7
				Intron 1	P8	76874635			
36	2	12q22	SOCS2: Suppressor of Cytokine Signaling-2	23700 bp down	P1	92496146	9	9	9
				19207 bp down	P5	92491653			
37	2	14q13.2	PSMA6: Proteasome Alpha 6 Subunit	1551 bp up	P4	33749966	9	9	9
				4319 bp up	P5	33747269			
38	2	16p12.3	THUMPD1: THUMP Domain Containing 1	5738 bp up	P2	20725409	8	8	8
				3018 bp up	P4	20722689			
39	2	17q21.31	FMNL1: Formin-like 1	Intron 1	P4	43779824	6	6	6
				Intron 1	P7	43776896			
40	2	17q25.3	ACTG1: Actin Gamma 1 Propeptide	56718 bp down	P7	80120400	9	9	9
				56689 bp down	P8	80120429			
41	2	18q11.2	CABLES1: Cdk5 and Abl enzyme substrate	Intron 4	P4	19060018	7	7	6
				Intron 4	P7	19061022			
42	2	19p13.3	EMR1: EGF-like Module Containing Mucin-like Hormone	Intron 1	P7	6840680	4	3	3
				Intron 1	P8	6841318			
43	2	19q13.2	R29124_1: Carcinoembryonic Antigen-related Cell Adhesion	22202 bp up	P2	46752225	n.p.	n.p.	n.p.
				30013 bp up	P7	46744414			
44	2	20q11.21	BCL2L1: Apoptose Regulator BCL2-like1	Intron 1	P1	31055248	6	7	7
				252 bp up	P8	31055908			
45	2	22q12.2	MGC26710: Hypothetical Protein	3934 bp down	P2	28901548	n.p.	n.p.	n.p.
				4544 bp down	P2	28902158			
46	2	Xp22.11	EIF2S3: Eukaryotic Translation Initiation Factor 2	34015 bp up	P2	23400476	9	9	9
				28932 bp up	P8	23405559			

Supplementary Table 3

CIS	IS	Locus	Next RefSeq Gene	Location	Patient	UCSC locus	A 1	A 2	A 3
47	2	Xq24	SEPT6: Septin 6 Isoform A	Intron 1	P8	117569952	n.p.	n.p.	n.p.
				Intron 2	P9	117567502			
48	2	1q32.1	NUCKS: Nuclear Ubiquitous Casein Kinase	14645 bp down	P4 Pre	202849203	9	9	9
				14577 bp down	P4 Pre	202849271			
49	2	13q13.1	13 CDNA 73: Hypothetical Protein	Intron 3	P4 Pre	30482648	n.p.	n.p.	n.p.
				Intron 3	P4 Pre	30487006			
50	2	16q24.2	ZFPM1: Friend of GATA-1	21986 bp up	P4 Pre	88244681	7	7	8
				Intron 1	P4 Pre	88270693			
51	2	17p13.1	GAS7: Growth Arrest-specific 7 Isoform c	Intron 1	P4 Pre	10258973	6	5	5
				Intron 1	P4 Pre	10258969			

Legends

Supplementary Table 1 online Proviral integration site sequences detected by LAM-PCR in SCID-X1 patients. LAM-PCR amplicons were shotgun cloned and sequenced. The genomic part of each sequenced integration site was aligned to the human genome database using University of California, Santa Cruz BLAT search tools (<http://genome.ucsc.edu/>). Integrations that could not be mapped to the human genome (too short, no definite or multiple hits) are not shown. 'Pat.' means patient; 'Source' denotes the surface marker of the cells on which the LAM-PCR was performed; 'PBL', peripheral blood leukocytes; 'PBMC', peripheral blood mononuclear cells; 'CD15', purified granulocytes; 'CD 14', purified monocytes; 'CD19', B lymphocytes; 'CD3', early T-cells; 'CD34', hematopoietic stem-/progenitor cells; 'Tgd', T-cell receptor gamma delta negative cells; 'Time' means months after transplantation; 'Length' denotes the size of the LAM-PCR amplicon without linker- and LTR-sequences; 'Identity', shows the cell type in which exactly the same RIS was found; TSS, transcription start site; bp, Base pairs; 'CpG' island, the RIS site is situated in or not more than 1 kb apart from a CpG island; CGP, the gene is listed in the Cancer genome project (<http://www.sanger.ac.uk/genetics/CGP/>); 'Repetitive elements', if the RIS is located in a repetitive element, the element is given.

Supplementary Table 2 online Statistical analysis of common integration sites (CIS) of different orders detected in the engrafted patient cells. Statistical analysis was performed on CIS which are formed of 2 (second order) or more (third and higher orders, respectively) integration sites. CIS formation detected in patient cells was significantly non-random as determined by computer simulations. P-value from simulations = proportion of the 10,000 runs, in which the number of CIS was at least

as high as the number observed in the trial. 'No. of CIS', number of retrieved CIS. 'nd', the expected number of CIS of 5th and higher order is lower than that of 3rd and 4th order. Therefore, no simulations were carried out for order $n > 4$.

Supplementary Table 3 online Summary of all common integration sites (CIS) detected among the 9 patients. Shown are all CIS which are formed of 2 or more integration sites. The RefSeq gene hit is shown with the UCSC integration locus. Comparative analysis of CIS formation in correlation with gene expression of target CD34+ cells is shown in the last three columns. Expression level (bins 0 – 9; **Fig. 3a, 3b**) of the genes associated with CIS formation is shown for each of the three individual microarray chips A 1, A 2 and A 3. 'pre', pre-transplant; 'n.p.', not present on microarray chip; 'bp', base pairs; 'up', upstream; 'down', downstream.