SUPPLEMENTARY APPENDIX

Analysis of retrotransposon subfamily DNA methylation reveals novel early epigenetic changes in chronic lymphocytic leukemia

Timothy M. Barrow,[1] Nicole Wong Doo,[2,3] Roger L. Milne,[2,4,5] Graham G. Giles,[2,4,5] Elaine Willmore,[6] Gordon Strathdee[7] and Hyang-Min Byun[7]

[1]Faculty of Health Sciences and Wellbeing, University of Sunderland, Sunderland, UK; [2]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia; [3]Concord Hospital, University of Sydney, Sydney, Australia; [4]Center for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia; [5]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Australia; [6]CR UK Drug Discovery Unit, Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK and [7]Newcastle University Center for Cancer, Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK

 doi:10.3324/haematol.2019.228478

Received: June 11, 2019.

Accepted: January 7, 2020.

Pre-published: January 9, 2020.

Correspondence: *TIMOTHY BARROW* - timothy.barrow@sunderland.ac.uk

Supplementary methods

Validation cohort

Results from the discovery cohort patients from the study by Kulis *et al*[1] were validated in a cohort of 24 CLL patients attending clinic at the Freeman Hospital (Newcastle upon Tyne, UK), Queen Elizabeth Hospital (Gateshead, UK) and Sunderland Royal Hospital (Sunderland, UK). The patient characteristics are provided in Supplementary Table 2. Mononuclear cells were isolated from peripheral blood samples taken from patients with leukocyte counts of >30 x 10^9/L, using density centrifugation with Lymphoprep media (Stem Cell Technologies) according to the manufacturer's instructions. Genomic DNA was extracted using the Qiagen Blood and Tissue kit according to the manufacturer's protocol, and analysed by HM450K array at the Edinburgh Clinical Research Facility, University of Edinburgh (United Kingdom). Ethical approval for this study was granted by the North East - Newcastle & North Tyneside 1 Research Ethics Committee (REC reference number 17/NE/0361). All donors provided written informed consent.

Identification of microarray probes mapping to retrotransposon subfamilies

Probes on the HM450K microarray mapping to CpG sites within retrotransposons in the human genome were identified using the Data Integrator function of the UCSC Genome Browser[2] with annotation from RepeatMasker[3] for the human genome build GRCh37/hg19. Those mapping to L1, *Alu*, half-L1 (HAL1), fossil *Alu* monomer (FAM), free right *Alu* monomer (FRAM) and free left *Alu* monomer (FLAM) elements were extracted and taken forward for analysis. Probes with single nucleotide polymorphisms (SNPs) reported at or within 10 base pairs of the target CpG site were included to facilitate the identification of variant-associated differential methylation. L1 elements were categorised into oldest (L1M, mammalian-wide), intermediate (L1P, primate-specific) and youngest (L1HS, human- specific and L1PA, primate-amplified)

subfamilies. L1PAs are active in retrotransposition and have a high sequence similarity to L1HS[4], thus these subfamilies were categorised together. *Alu* elements were categorised into oldest (*Alu*J), intermediate (*Alu*S) and youngest (*Alu*Y) subfamilies[5]. HAL1, FAM, FLAM and FRAM elements were excluded from analysis by evolutionary age. The subfamilies and number of probes mapping to each are displayed in Figure 1 and Supplementary Table 3.

Differential methylation of retrotransposon subfamilies

Mean differences in methylation (β) between CLL patients and healthy individuals in the discovery cohort were calculated for each of the loci mapping to retrotransposons, from which the mean methylation change ($\Delta\beta$) for each of the 117 L1 and 37 *Alu* subfamilies was calculated. Analysis of methylation change by subfamily evolutionary age was performed by linear regression using estimates of time since sequence amplification (millions of years ago, MYA) from the studies of Kapitonov *et al*[6] and Khan *et al*[7].

Statistical analyses

Methylation of L1 and *Alu* elements by subfamily and genomic feature were analysed in samples of normal $CD19^+$ B-cells from 14 individuals by ANOVA, using data on genomic features extracted from the Illumina HM450K annotation file. Loci that are differentially methylated in CLL were identified by t-test, with significance defined as $p<0.05$ following correction for false discovery rate (FDR) by the Benjamini-Hochberg method. Enrichment of differentially methylated loci by genomic region was examined by Fisher's exact test, using features defined in the Illumina H450K annotation file. Correlations with gene expression were examined by Spearman's rank correlation test. The impact upon patient prognosis was assessed in the Discovery cohort by Kaplan-Meier plot and logrank test, for which patients were dichotomised into high and low expression

groupings by ROC curve analysis, with the optimal threshold determined by the Youden index. Comparison of locus-specific methylation in prospective CLL cases and age-matched controls was made by Mann-Whitney U test, with correlations in DNA methylation between different loci and with time to diagnosis assessed by Spearman's rank correlation test. Where appropriate, p-values were adjusted for multiple hypothesis testing by the Benjamini, Krieger and Yekutieli method (q=0.05), with significance defined as $P_{FDR}<0.05$. All statistical analyses were performed in R (version 3.4.0) using the ggplot2, heatmap2, qqman and corrplot packages, and in GraphPad Prism (version 7.0b).

Supplementary references

1. Kulis M, Heath S, Bibikova M *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet. 2012;44(11):1236-1242.
2. Kent WJ, Sugnet CW, Furey TS *et al.* The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.
3. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013
4. Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N. Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics. 2014;15:583.
5. Mighell AJ, Markham AF, Robinson PA. *Alu* sequences. FEBS Lett. 1997;417(1):1-5.
6. Kapitonov V, Jurka J. The age of *Alu* subfamilies. J Mol Evol. 1996;42(1):59-65.
7. Khan H, Smit A, Boissinot S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 2006;16(1):78-87.

Data type	Study	Access code	Disease / cell type	n
DNA methylation	Kulis *et al*	EGAS00001000272	Normal CD19+ B-cells	14
			CLL	139
	Hannum *et al*	GSE40279	Healthy	656
	EPIC	GSE51057	Healthy	329
	Young Finns Study	GSE69270	Healthy	184
	Nordlund *et al*	GSE47051	ALL	797
	Ferreira *et al*	GSE62298	AML	68
	Maupetit-Mehouas *et al*	GSE106600	CML	12
	Matsunaga *et al*	GSE42372	Lymphoma	31
	Reinius *et al*	GSE35069	Whole blood	6
			PBMC	6
			CD16+ neutrophils	6
			Eosinophils	6
			CD14+ monocytes	6
			CD19+ B-cells	6
			CD4+ T-cells	6
			CD8+ T-cells	6
			CD56+ Natural killer cells	6
	Lee *et al*	GSE45461	Multipotent progenitors	6
			Pre-B-I cells	6
			Pre-B-II cells	6
			Immature B-cells	4
	MCCS	N/A	Control	82
			CLL case	82
Gene expression	Kulis *et al*	EGAS00001000272	CLL	139
	Landau *et al*	GSE50006	Normal CD19+ B-cells	32
			CLL	188

Supplementary Table 1: Utilised DNA methylation and gene expression microarray datasets. Details of the studies from which datasets originated are provided with their Gene Expression Omnibus (GSE) and European Genome-phenome Archive (EGA) accession codes where appropriate.

Characteristic		***n***
Patients	Total	24
Gender	Male	16
	Female	8
Age	Median	66
Binet stage	A	8
	B	5
	C	10
	Unknown	1
IGHV	Mutated	13
	Unmutated	10
	Equivalent	1

Supplementary Table 2: Characteristics of CLL patients within the validation cohort.

Category	Subfamily	Probes
*Alu*J (old)	AluJb	1689
	AluJo	832
	AluJr	888
	AluJr4	154
*Alu*S (intermediate)	AluSc	377
	AluSc5	75
	AluSc8	225
	AluSg	480
	AluSg4	88
	AluSg7	103
	AluSp	591
	AluSq	240
	AluSq10	41
	AluSq2	606
	AluSq4	18
	AluSx	1541
	AluSx1	1240
	AluSx3	387
	AluSx4	63
	AluSz	1077
	AluSz6	474
*Alu*Y (young)	AluY	1256
	AluYa5	33
	AluYa8	2
	AluYb8	23
	AluYc	94
	AluYc3	6
	AluYd8	1
	AluYf4	15
	AluYg6	7
	AluYh9	2
	AluYk11	3
	AluYk4	26
Alu FAM/FLAM/FRAM	FAM	73
	FLAM_A	109
	FLAM_C	201
	FRAM	90
L1M (old)	L1M	4
	L1M1	97

	L1M2	56
	L1M2a	9
	L1M2a1	3
	L1M2b	3
	L1M2c	3
	L1M3	48
	L1M3a	5
	L1M3b	5
	L1M3c	9
	L1M3d	8
	L1M3de	7
	L1M3e	9
	L1M3f	19
	L1M4	176
	L1M4b	60
	L1M4c	40
	L1M5	483
	L1M6	40
	L1M7	30
	L1MA1	18
	L1MA10	25
	L1MA2	27
	L1MA3	71
	L1MA4	48
	L1MA4A	48
	L1MA5	28
	L1MA5A	19
	L1MA6	32
	L1MA7	39
	L1MA8	58
	L1MA9	130
	L1MB1	32
	L1MB2	87
	L1MB3	434
	L1MB4	110
	L1MB5	259
	L1MB7	366
	L1MB8	212
	L1MC	110
	L1MC1	166
	L1MC2	93
	L1MC3	216

	L1MC4	407
	L1MC4a	357
	L1MC5	202
	L1MCa	48
	L1MCb	20
	L1MCc	36
	L1MD	70
	L1MD1	119
	L1MD2	178
	L1MD3	53
	L1MDa	75
	L1MDb	30
	L1ME1	388
	L1ME2	135
	L1ME2z	79
	L1ME3	136
	L1ME3A	212
	L1ME3B	113
	L1ME3C	139
	L1ME3D	34
	L1ME3E	69
	L1ME3F	39
	L1ME4a	271
	L1ME5	47
	L1MEa	3
	L1MEb	16
	L1MEc	184
	L1MEd	76
	L1MEe	54
	L1MEf	114
	L1MEg	129
	L1MEg1	1
	L1MEg2	5
L1P (intermediate)	L1P1	25
	L1P2	5
	L1P3	10
	L1P3b	2
	L1P4	8
	L1P4a	12
	L1P4c	4
	L1P4d	6
	L1P5	1

	L1PB	4
	L1PB1	58
	L1PB2	18
	L1PB3	35
	L1PB4	47
	L1PBa	35
	L1PBa1	17
	L1PBb	18
	L1PREC2	59
L1HS + L1PA (young)	L1HS	62
	L1PA10	40
	L1PA11	34
	L1PA12	22
	L1PA13	27
	L1PA14	17
	L1PA15	48
	L1PA15-16	11
	L1PA16	93
	L1PA17	49
	L1PA2	443
	L1PA3	429
	L1PA4	64
	L1PA5	38
	L1PA6	35
	L1PA7	39
	L1PA8	19
	L1PA8A	27
L1 HAL1	HAL1	313
	HAL1-2a_MD	12
	HAL1-3A_ME	2
	HAL1b	28

Supplementary Table 3: Distribution of HM450K probes by L1 and *Alu* subfamily.

ProbeID	Genomic location	SNPs	MAF
cg20820557	6:53453216	rs76774835	0.002
cg04258086	16:85966544	rs556243593, rs535919808	0.000, 0.000
cg05922591	19:55174624	rs554876917	0.000
cg17505852	22:30003602	rs35821749	NFD
cg11665613	1:12268883	rs888649456	NFD
cg00981250	6:144330345		
cg16564946	6:32304275	rs555256321	0.000
cg10795552	5:138605898	rs527638242	0.000
cg22894805	15:41983772	rs747065280	<0.001
cg17342709	2:235319934		
cg27020349	6:144643174	rs1052158873, rs6911171	NFD, 0.007
cg08641155	16:70469236	rs545836907	0.000
cg18406010	14:51137625	rs140554169	0.000
cg02575319	16:81602666	rs11644375	0.470
cg10474404	3:129148046	rs964623436, rs765349980	NFD, NFD
cg25995870	19:38571162	rs1031011006, rs956770938	NFD, NFD
cg04312999	21:45399399	rs575491202	0.000
cg09149541	22:19898335		
cg18286080	6:30291349	rs761784554	NFD
cg12115081	4:151038390	rs771959741	NFD
cg14266770	9:136721182	rs778084658	NFD
cg19679081	12:14514672	rs915031575	NFD
cg22813097	19:18629910	rs3888232	0.000
cg16386046	16:89394863		
cg17084653	22:26822031	rs200075284, rs142721270	NFD, 0.000
cg23985408	5:171410890	rs55805925, rs575232522	0.009, 0.000
cg07134930	2:240176050	rs1034335966	NFD
cg02948444	17:80741558	rs187516783, rs750796759	0.000, NFD

cg25947773	2:223771010	rs920385023	NFD
cg16273734	1:181087919		
cg16348358	1:32731477	rs575444419	0.000
cg10664272	2:85638053	rs998718100	NFD
cg17713912	7:103449892	rs34778492, rs147083966, rs752127891	0.210, 0.001, NFD
cg07293188	3:38209834	rs575956128, rs552411201	NFD, 0.001
cg04211501	16:85966435	rs754878910, rs567410565	NFD, 0.000
cg13988440	11:69240805	rs770280243	NFD
cg10180165	17:40810558	rs1027402131	NFD
cg23840797	6:144330232		
cg14985591	10:126274649	rs190435833, rs561989868	0.000, 0.000
cg11848483	8:144543485	rs543377530, rs376594263	0.000, NFD
cg20959920	17:75238948	rs766102651	NFD
cg27447753	2:162047157	rs370794613	NFD
cg26924822	4:26332762	rs115428818, rs80320396	0.000, NFD
cg21518709	18:3063385		
cg10163122	6:34633132	rs370048158	NFD
cg23177739	10:71087363	rs563555918, rs115198540	NFD, 0.007
cg12079885	12:125328475	rs922267132, rs543984432	NFD, 0.001
cg15129876	11:82865068	rs979181925, rs187227188	NFD, 0.000
cg23876355	17:47723651	rs143055540	0.001
cg09153458	4:979307	rs76774835	0.002

Supplementary Table 4: SNPs reported at the 50 leading retrotransposon loci.

SNPs mapping to the CpG dinucleotides (C or G) interrogated by the HM450K probes are provided for the 50 most significantly differentially methylated loci. The HM450K probe ID, genomic position (chromosome and base coordinate) to which they map and SNP are provided, with the minor allele frequency ('MAF') reported in European populations from the 1000 Genomes Project (NFD = no frequency distribution data available).

				Discovery (n=139)		Validation (n=24)	
ProbeID	**Element**	**chr**	**Gene**	**Δβ**	**P_{FDR}**	**Δβ**	**P_{FDR}**
cg20820557	AluSx	6:53453216		0.70	1.88×10^{-104}	-	-
cg04258086	L1ME3B	16:85966544		0.58	3.83×10^{-87}	0.52	4.35×10^{-12}
cg05922591	AluY	19:55174624	*LILRB4*	0.59	1.44×10^{-78}	0.58	4.63×10^{-14}
cg17505852	L1M4b	22:30003602	*NF2*	0.69	3.81×10^{-78}	0.66	2.09×10^{-15}
cg11665613	FRAM	1:12268883	*TNFRSF1B*	0.58	6.35×10^{-73}	0.60	1.52×10^{-12}
cg00981250	HAL1b	6:144330345	*PLAGL1*	0.56	2.09×10^{-72}	0.55	1.29×10^{-10}
cg16564946	L1PA16	6:32304275	*C6orf10*	0.55	2.97×10^{-70}	0.55	1.55×10^{-14}
cg10795552	L1MB7	5:138605898		0.60	1.89×10^{-69}	0.65	7.67×10^{-14}
cg22894805	L1MB8	15:41983772	*MGA; MIR626*	0.54	3.28×10^{-68}	0.61	5.38×10^{-15}
cg17342709	AluSg	2:235319934		0.65	2.48×10^{-66}	0.65	1.44×10^{-15}
cg27020349	AluSx1	6:144643174	*UTRN*	0.68	6.10×10^{-66}	-	-
cg08641155	AluSp	16:70469236	*ST3GAL2*	0.48	2.48×10^{-64}	0.50	2.16×10^{-12}
cg18406010	AluSx1	14:51137625		0.46	4.56×10^{-63}	0.56	1.24×10^{-11}
cg02575319	L1M5	16:81602666	*CMIP*	0.58	9.92×10^{-63}	-	-
cg10474404	L1MB7	3:129148046	*C3orf25*	0.49	2.85×10^{-62}	0.55	1.83×10^{-12}
cg25995870	L1MC5	19:38571162	*SIPA1L3*	0.46	1.07×10^{-61}	0.47	6.14×10^{-10}
cg04312999	L1MEd	21:45399399	*AGPAT3*	0.54	2.75×10^{-61}	0.67	6.60×10^{-15}
cg09149541	AluSx1	22:19898335	*TXNRD2*	0.49	3.12×10^{-61}	0.52	2.11×10^{-10}
cg18286080	L1MC4a	6:30291349	*HCG18*	0.52	1.68×10^{-59}	0.53	2.41×10^{-11}
cg12115081	L1PB3	4:151038390	*DCLK2*	0.56	4.25×10^{-59}	0.64	3.39×10^{-16}
cg14266770	L1MB5	9:136721182	*VAV2*	0.51	4.90×10^{-58}	0.57	1.02×10^{-12}
cg19679081	AluSq2	12:14514672		0.50	2.08×10^{-57}	0.49	2.82×10^{-10}
cg22813097	AluY	19:18629910	*ELL*	0.42	5.92×10^{-57}	-	-
cg16386046	L1MD2	16:89394863	*ANKRD11*	0.47	2.17×10^{-56}	0.42	1.88×10^{-8}
cg17084653	AluSx	22:26822031		0.43	1.14×10^{-55}	0.48	2.40×10^{-12}
cg23985408	L1ME3	5:171410890	*FBXW11*	0.51	1.69×10^{-55}	0.59	1.78×10^{-11}
cg07134930	L1MEe	2:240176050	*HDAC4*	0.54	3.00×10^{-55}	-	-
cg02948444	L1MB7	17:80741558	*TBCD*	0.48	3.33×10^{-54}	0.53	7.45×10^{-11}

cg25947773	L1M4b	2:223771010	*ACSL3*	0.48	3.45×10^{-53}	0.51	7.51×10^{-11}
cg16273734	AluSq2	1:181087919		0.40	3.01×10^{-52}	0.47	6.06×10^{-13}
cg16348358	FLAM_C	1:32731477	*LCK*	0.43	7.64×10^{-52}	0.51	3.48×10^{-12}
cg10664272	AluJo	2:85638053	*CAPG*	0.42	3.92×10^{-51}	0.46	3.70×10^{-8}
cg17713912	HAL1	7:103449892	*RELN*	0.50	7.45×10^{-51}	0.53	1.88×10^{-9}
cg07293188	L1MB7	3:38209834	*OXSR1*	0.45	3.20×10^{-50}	0.49	5.23×10^{-10}
cg04211501	L1ME3B	16:85966435		0.79	3.99×10^{-50}	0.79	4.09×10^{-21}
cg13988440	L1MC2	11:69240805		0.61	9.69×10^{-50}	0.67	6.99×10^{-16}
cg10180165	L1MB3	17:40810558	*TUBG2*	0.29	1.13×10^{-49}	-	-
cg23840797	HAL1b	6:144330232	*PLAGL1*	0.51	1.14×10^{-48}	0.53	3.61×10^{-14}
cg14985591	AluSz	10:126274649	*LHPP*	0.41	1.29×10^{-48}	0.49	8.71×10^{-10}
cg11848483	L1MB5	8:144543485	*ZC3H3*	0.52	1.63×10^{-48}	0.65	6.16×10^{-10}
cg20959920	L1MC4	17:75238948		0.41	3.75×10^{-47}	0.45	2.28×10^{-9}
cg27447753	L1MB3	2:162047157	*TANK*	0.34	1.08×10^{-46}	0.43	2.29×10^{-12}
cg26924822	AluJb	4:26332762	*RBPJ*	0.39	2.62×10^{-46}	-	-
cg21518709	L1MEd	18:3063385		0.41	2.75×10^{-46}	0.51	2.22×10^{-11}
cg10163122	AluJo	6:34633132	*C6orf106*	0.44	3.80×10^{-46}	0.52	2.40×10^{-11}
cg23177739	L1ME3C	10:71087363	*HK1*	0.38	6.27×10^{-46}	0.41	6.22×10^{-10}
cg12079885	L1ME5	12:125328475	*SCARB1*	0.28	1.51×10^{-45}	-	-
cg15129876	AluSz	11:82865068		0.36	2.09×10^{-45}	0.45	1.08×10^{-10}
cg23876355	AluSz	17:47723651	*SPOP*	0.38	8.21×10^{-45}	-	-
cg09153458	L1MEg	4:979307	*SLC26A1; IDUA*	0.50	4.56×10^{-44}	0.49	4.80×10^{-9}

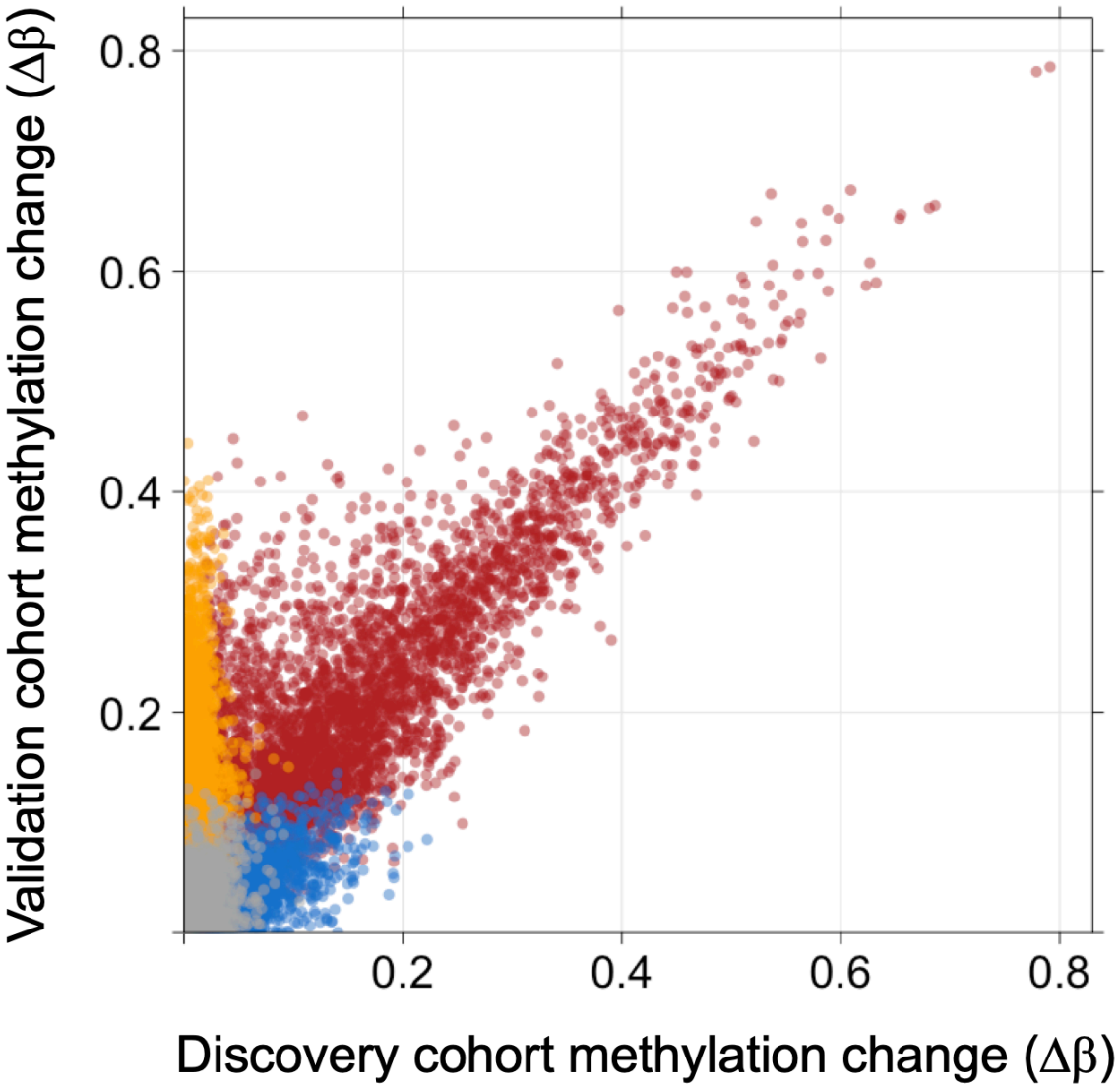
Supplementary Table 5: Validation of differentially methylated retrotransposon loci in a second CLL cohort. Results for the 50 most significantly differentially methylated loci in the discovery cohort are presented. The mean difference in methylation ($\Delta\beta$) between CLL patients and healthy controls (n=14) and the FDR-adjusted p-values are presented for the discovery (n=139) and validation (n=24) cohorts. The Illumina probe ID, retrotransposon type ('Element'), genomic location ('chr') and gene from the Illumina annotation file are provided for each locus. Data was missing for nine of the loci within the validation cohort dataset and could not be examined.

Probe ID	Element	Gene	Genomic feature	Enhancer	Mean methylation (β)			Meth/exp correlation		Expression in CLL		OS
					B-cell	Disc.	Val.	rho	P_{FDR}	Change	P_{FDR}	P_{FDR}
cg11665613	FRAM	*TNFRSF1B*	3'UTR	TRUE	0.86	0.28	0.26	-0.38	<0.0001	Increased	0.0002	0.0088
cg09149541	AluSx1	*TXNRD2*	Body		0.88	0.39	0.36	-0.27	0.0142	Increased	<0.0001	0.1880
cg12115081	L1PB3	*DCLK2*	Body	TRUE	0.85	0.28	0.20	-0.59	<0.0001	Increased	<0.0001	0.0011
cg02948444	L1MB7	*TBCD*	Body		0.88	0.40	0.35	0.25	0.0286	Increased	0.0008	0.0268
cg27447753	L1MB3	*TANK*	Body		0.82	0.49	0.40	-0.25	0.0286	Decreased	<0.0001	0.0007
cg10163122	AluJo	*C6orf106*	Body	TRUE	0.83	0.38	0.30	-0.28	0.0142	Increased	<0.0001	0.0086
cg23177739	L1ME3C	*HK1*	Body	TRUE	0.85	0.47	0.44	-0.34	<0.0001	Increased	<0.0001	0.0201

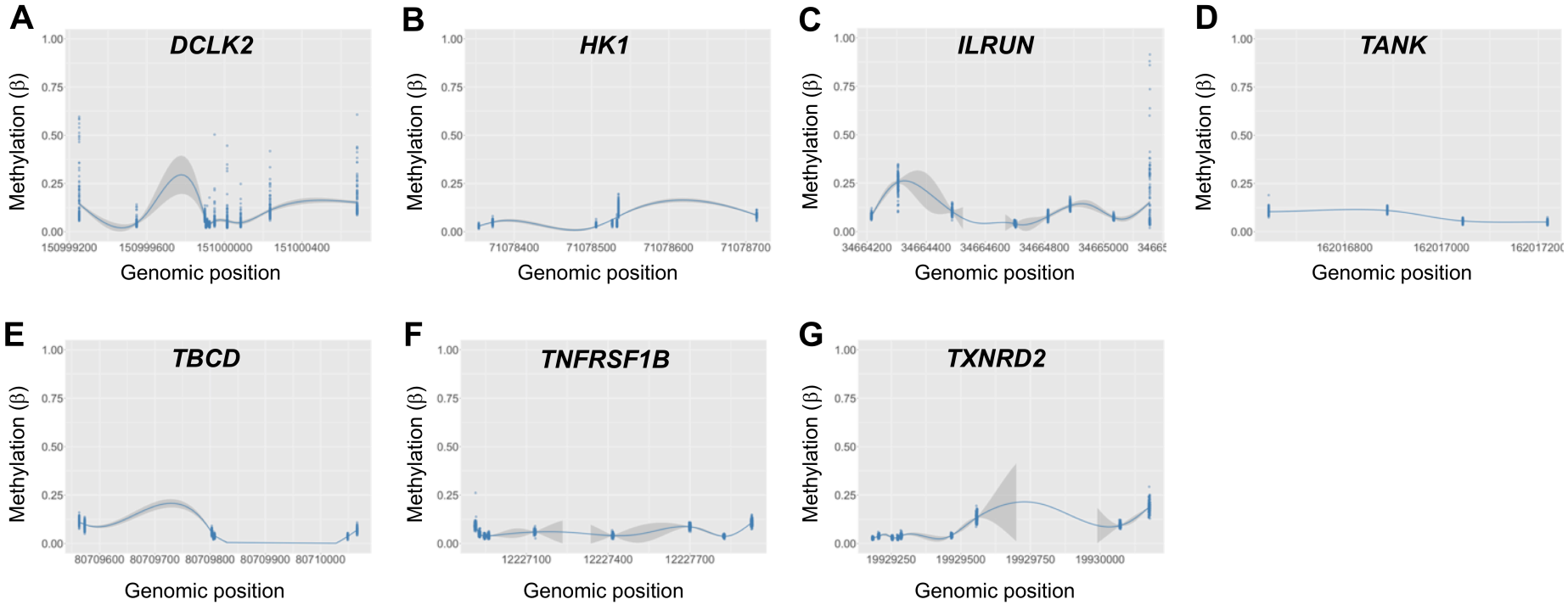
Supplementary Table 6: Effect of locus-specific hypomethylation on proximal gene expression. Methylation (β) at seven loci in CLL patients within the discovery ('Disc.') and validation ('Val.') cohorts, and in normal $CD19^+$ B-cells isolated from healthy individuals ('B-cell'). The correlation between methylation (β) and expression (log2) within the discovery cohort was calculated by Spearman's rank correlation test ('Meth/Exp correlation'). Differential expression between normal $CD19^+$ B-cells and CLL was determined by Mann-Whitney U test ('Expression in CLL'). Associations with patient overall survival (OS) were calculated by log-rank test.

Disease	cg20820557	cg04258086	cg05922591	cg17505852	cg11665613	cg00981250	cg16564946	cg10795552	cg22894805	cg17342709
Hannum	0.89 ±0.02	0.86 ±0.03	0.87 ±0.02	0.88 ±0.03	0.89 ±0.02	0.93 ±0.02	0.74 ±0.11	0.88 ±0.02	0.38 ±0.08	0.91 ±0.02
YFS	0.94 ±0.01	-	0.90 ±0.01	0.92 ±0.02	0.93 ±0.01	0.95 ±0.01	0.79 ±0.07	0.90 ±0.01	0.48 ±0.07	0.94 ±0.01
EPIC	0.91 ±0.02	0.87 ±0.02	0.89 ±0.02	0.92 ±0.03	0.93 ±0.02	0.96 ±0.02	0.78 ±0.09	0.90 ±0.02	0.42 ±0.07	0.93 ±0.02
CLL	0.16 ±0.12	0.26 ±0.14	0.22 ±0.15	0.18 ±0.09	0.28 ±0.17	0.32 ±0.17	0.29 ±0.15	0.26 ±0.18	0.25 ±0.14	0.18 ±0.11
ALL	0.95 ±0.04	0.91 ±0.05	0.88 ±0.12	0.94 ±0.10	0.90 ±0.14	0.97 ±0.05	0.94 ±0.09	0.87 ±0.16	0.72 ±0.19	0.96 ±0.02
AML	0.90 ±0.02	0.85 ±0.04	0.82 ±0.14	0.89 ±0.04	0.90 ±0.03	0.93 ±0.02	0.87 ±0.06	0.89 ±0.03	0.40 ±0.21	0.89 ±0.03
CML	0.90 ±0.01	0.85 ±0.02	0.86 ±0.02	0.85 ±0.02	0.90 ±0.02	0.92 ±0.02	0.74 ±0.22	0.86 ±0.04	0.16 ±0.03	0.88 ±0.01
Lymphoma	0.67 ±0.24	0.69 ±0.23	0.72 ±0.21	0.65 ±0.20	0.67 ±0.25	0.73 ±0.13	0.65 ±0.18	0.61 ±0.24	0.56 ±0.21	0.70 ±0.21

Supplementary Table 7: Methylation of the leading 10 retrotransposon loci in healthy individuals and haematological malignancies. Mean methylation levels ($\beta \pm$ SD) for ten leading retrotransposon loci in three studies of healthy individuals (Hannum *et al*, n=656; EPIC, n=329; the Young Finns Study (YFS), n=184), CLL patients within the discovery cohort (n=139), and patients with ALL (n=797), AML (n=68), CML (n=12), and diffuse large B-cell and Burtkitt's lymphomas (Lymphoma, n=31).



Supplementary Figure 1: Validation of differentially methylated retrotransposon loci in a second CLL cohort. Correlation between the absolute difference in methylation (Δβ) between CLL patients and healthy individuals observed in the discovery cohort (x-axis; n=139) and the validation cohort (y-axis; n=24). Each of the interrogated retrotransposon loci is represented by a dot that is coloured according to statistical significance (P_{FDR}<0.05) in both the discovery and validation cohorts (red), only the discovery cohort (blue), only the validation cohort (yellow), or neither (grey).



Supplementary Figure 2: Promoter methylation of seven differentially expressed proximal genes. DNA methylation (β) at CpG sites mapping to the promoters of the *DCLK2* (A), *HK1* (B), *ILRUN* (C), *TANK* (D), *TBCD* (E), *TNFRSF1B* (F) and *TXNRD2* (G) genes among 139 chronic lymphocytic patients within the discovery cohort. Promoter-associated loci were identified through the Illumina HumanMethylation 450 microarray annotation file. Genomic position is according to the human genome build GRCh37/hg19.