Methylation profiling of mediastinal gray zone lymphoma reveals a distinctive signature with elements shared by classical Hodgkin's lymphoma and primary mediastinal large B-cell lymphoma

Franziska C. Eberle,[1] Jaime Rodriguez-Canales,[1] Lai Wei,[2] Jeffrey C. Hanson,[1] J. Keith Killian,[3] Hong-Wei Sun,[4] Lisa G. Adams,[3] Stephen M. Hewitt,[1] Wyndham H. Wilson,[5] Stefania Pittaluga,[1] Paul S. Meltzer,[3] Louis M. Staudt,[5] Michael R. Emmert-Buck,[1] and Elaine S. Jaffe[1]

[1]Laboratory of Pathology, [3]Genetics Branch, and [5]Metabolism Branch, National Cancer Institute, Center for Cancer Research, [2]Laboratory of Immunology, National Eye Institute, [4]Biodata Mining and Discovery Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA

Citation: Eberle FC, Rodriguez-Canales J, Wei L, Hanson JC, Killian JK, Sun H-W, Adams LG, Hewitt SM, Wilson WH, Pittaluga S, Meltzer PS, Staudt LM, Emmert-Buck MR, and Jaffe ES. Methylation profiling of mediastinal gray zone lymphoma reveals a distinctive signature with elements shared by classical Hodgkin's lymphoma and primary mediastinal large B-cell lymphoma. Haematologica 2011;96(4):558-566. doi:10.3324/haematol.2010.033167

Online Supplementary Methods

Identification of a gene set for class prediction

Linear discriminant analysis (LDA), a statistical and machine learning method used to identify the linear combination of variables that best separates two or more classes, was used to build multiple predictive models (as implemented in Partek Genomics Suite 6.5). Model genes were selected with forward selection, a data-driven model building approach in which the best orthogonal variables were sequentially added to a model until the error rate becomes zero on the model-building data set. The total number of genes allowed in a model was limited to the number of samples in the model-building data set to prevent over-fitting. Genes already selected to build a model were removed from the gene pool and the process was repeated to create multiple predictive models. Each model was validated with a real leave-one-out cross validation approach, in which a model was completely rebuilt without the removed sample. This whole model building process allowed identification of multiple predictive gene sets. Results of the specific leave-one-out cross validation also allowed the identification of most frequently selected genes for building the models. The predictive models were finally validated by applying them to an independent data set using a simple majority voting scheme.

Online Supplementary Table S1. Occurence of *de novo* hypo- or hypermethylation in MGZL, CHLNS, PMLBCL, and DLBCL. Data show absolute and relative numbers of CpG in tumor cells isolated from the indicated lymphoma groups with *de novo* hypomethylation ($\beta_{RTB}-\beta_{tumor} \geq 0.30$) or *de novo* hypermethylation ($\beta_{tumor}-\beta_{RTB} \geq 0.30$) and their location outside CpG islands (non-CGI) or within CpG islands (CGI).

	De novo hypomethylation				*De novo* hypermethylation			
	non-CGI		CGI		non-CGI		CGI	
MGZL	11	58%	8	42%	1	1%	107	99%
CHLNS	0	0%	0	0%	12	14%	72	86%
PMLBCL	28	68%	13	32%	0	0%	223	100%
DLBCL	69	81%	16	19%	0	0%	244	100%

Online Supplementary Table S2. Pyrosequencing confirms the DNA methylation data generated by the Illumina GoldenGate Methylation Array. DNA from five lymphoma cell lines (Farage, K1106, L428, L1236, U2940) and three tissue samples of MGZL was used for pyrosequencing of four CpG sites selected from three different genes. Results were compared to the methylation data (β-values) generated with the Illumina GoldenGate Methylation Array. Methylation status of CpG sites as determined by pyrosequencing is given as percent of methylation (1.00 = 100%).

CpG site	Sample	Methylation Array	Pyro-sequencing	Pearson's correlation coefficient
CDH1_P52_R				
	Farage	0.99	0.91	
	K1106	0.99	0.94	
	L428	0.99	0.94	
	L1236	0.99	0.93	
	U2940	0.99	0.96	
	MGZL (12)	0.93	0.63	
	MGZL (13)	0.95	0.71	
	MGZL (16)	0.81	0.65	0.80
CDH1_P45_F				
	Farage	0.99	0.84	
	K1106	0.98	0.94	
	L428	0.98	0.98	
	L1236	0.98	0.91	
	U2940	0.98	0.93	
	MGZL (12)	0.61	0.59	
	MGZL (13)	0.59	0.59	
	MGZL (16)	0.43	0.39	0.98
FAT_P279_R				
	Farage	0.96	1.00	
	K1106	0.96	0.98	
	L428	0.93	1.00	
	L1236	0.94	0.90	
	U2940	0.98	1.00	
	MGZL (12)	0.60	0.75	
	MGZL (13)	0.79	0.69	
	MGZL (16)	0.51	0.75	0.82
SLIT2_P208_F				
	Farage	0.99	0.97	
	K1106	0.99	0.52	
	L428	0.99	0.99	
	L1236	0.99	0.91	
	U2940	0.99	0.97	
	MGZL (12)	0.48	0.66	
	MGZL (13)	0.70	0.49	
	MGZL (16)	0.35	0.52	0.65

Online Supplementary Table S3. Identification of differentially methylated CpG sites in MGZL compared to CHLNS or PMLBCL. Data show mean methylation levels (β-values) of CpG targets with differential methylation ($\Delta\beta \geq 0.30$) in MGZL compared to CHLNS (A) or PMLBCL (B). For completeness, mean β-values of PMLBCL (A) and CHLNS (B) were added.

A

Target ID*	MGZL	CHLNS	PMLBCL
NBL1_P24_F	0.44	0.90	0.44
CRIP1_P874_R	0.28	0.69	0.21
MMP9_P189_F	0.47	0.86	0.46
SPI1_P48_F	0.48	0.85	0.47
AIM2_P624_F	0.42	0.78	0.49
GFI1_P208_R	0.37	0.73	0.21
IL12B_P392_R	0.32	0.67	0.47
ERCC1_P440_R	0.19	0.55	0.16
CCL3_E53_R	0.35	0.70	0.28
KRT13_P676_F	0.53	0.88	0.47
AOC3_P890_R	0.43	0.77	0.37
HOXA5_P1324_F	0.23	0.55	0.48
PLA2G2A_E268_F	0.57	0.89	0.48
DDIT3_P1313_R	0.49	0.81	0.31
GRB7_P160_R	0.54	0.85	0.51
IFNG_P459_R	0.56	0.86	0.52
PLA2G2A_P528_F	0.53	0.83	0.45
CPA4_E20_F	0.51	0.81	0.37
FAS_P322_R	0.09	0.39	0.12
PLXDC2_P914_R	0.67	0.35	0.70

B

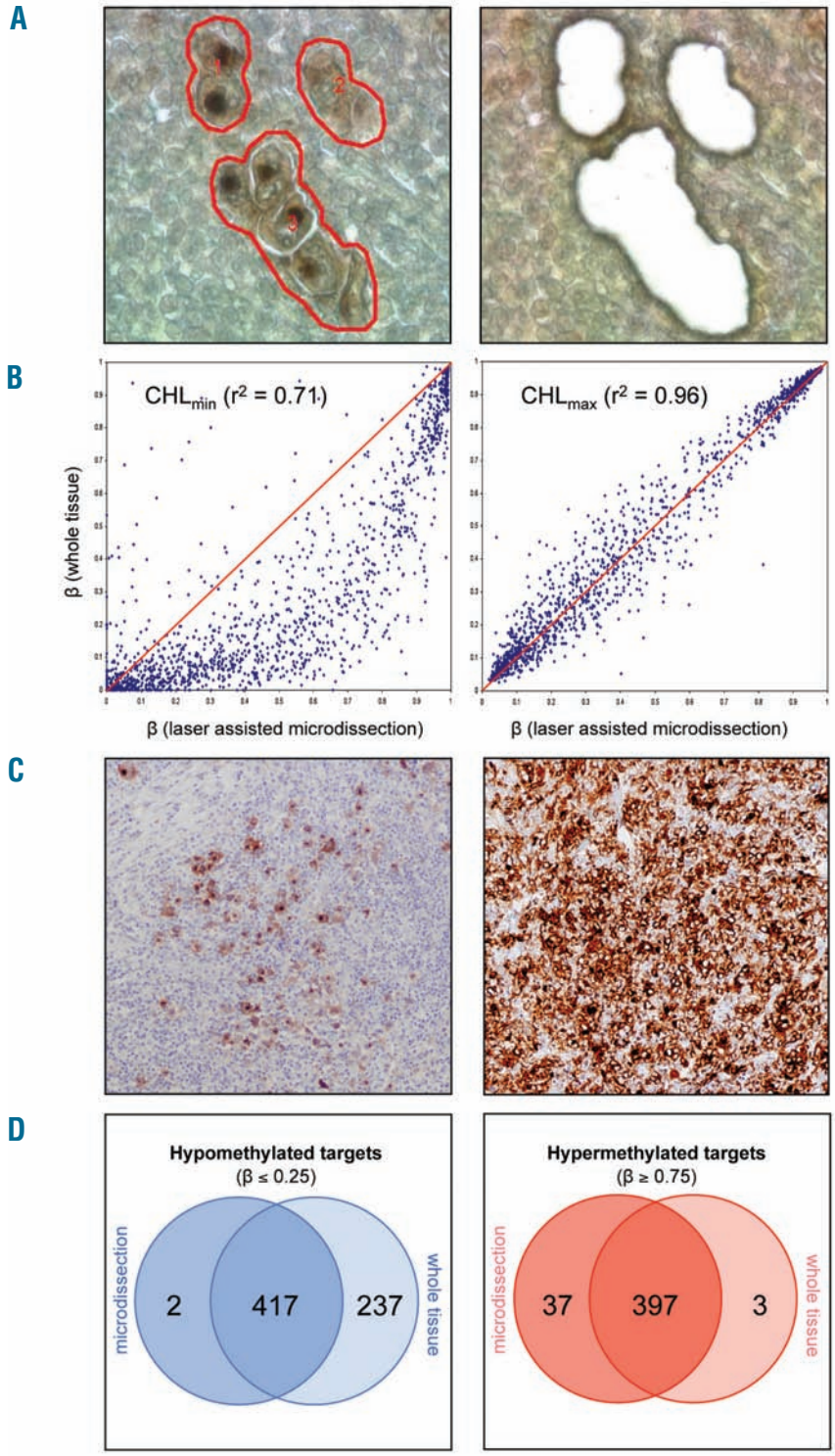
Target ID*	MGZL	PMLBCL	CHLNS
EPHA7_E6_F	0.29	0.74	0.25
CDH11_P203_R	0.30	0.74	0.23
EPHA7_P205_R	0.32	0.74	0.28
ASCL1_E24_F	0.23	0.65	0.23
DAPK1_P10_F	0.46	0.87	0.35
DBC1_P351_R	0.35	0.74	0.30
DAPK1_E46_R	0.31	0.69	0.36
WNT2_E109_R	0.23	0.61	0.17
RARB_P60_F	0.32	0.68	0.27
RBP1_P150_F	0.25	0.61	0.17
LMO1_E265_R	0.33	0.69	0.37
DIO3_P674_F	0.31	0.67	0.35
SEMA3A_P343_F	0.31	0.67	0.40
HIC2_P498_F	0.27	0.63	0.20
PTGS2_P308_F	0.24	0.59	0.29
EGFR_E295_R	0.37	0.71	0.25
ERBB4_P255_F	0.45	0.78	0.38
APBA1_P644_F	0.17	0.50	0.19
PRKCDBP_E206_F	0.32	0.65	0.22
MYOD1_P50_F	0.47	0.79	0.44
IGF2_P36_R	0.37	0.70	0.44
DAB2_P35_F	0.47	0.79	0.39
GABRB3_E42_F	0.36	0.68	0.23
CDH3_E100_R	0.48	0.79	0.49
RAB32_P493_R	0.33	0.64	0.23
NRG1_P558_R	0.47	0.77	0.35
CTGF_E156_F	0.26	0.56	0.20
SFRP1_E398_R	0.45	0.76	0.33
ERBB4_P541_F	0.50	0.81	0.42
MYOD1_E156_F	0.48	0.78	0.40
CDH1_P45_F	0.40	0.70	0.38
ABO_E110_F	0.40	0.70	0.30
TWIST1_P355_R	0.31	0.61	0.25
EGFR_P260_R	0.52	0.81	0.49
SIN3B_P607_F	0.77	0.47	0.92

** Gene symbols are contained within the Target ID before the first underscore.*

Online Supplementary Table S4. Validation of predictive models. Applying the created 13 predictive models to eight independent samples reveals an accuracy of the final combined prediction of 100%.

Sample	class	Model													h	g	p	vote
		1	2	3	4	5	6	7	8	9	10	11	12	13				
22	p	p	g	p	p	g	g	p	p	g	p	p	p	p	0	4	9	p
29	p	p	g	p	p	p	g	p	p	p	g	p	p	p	0	3	10	p
15	g	g	g	g	g	g	g	g	p	p	h	g	g	p	1	9	3	g
20	g	g	g	p	p	g	g	g	g	g	g	g	p	g	0	10	3	g
19P	g	g	p	g	p	g	g	g	p	p	g	g	p	p	0	7	6	g
19H	g	g	g	g	h	g	p	g	g	g	g	h	g	g	2	10	1	g
9	h	g	h	g	g	g	h	h	h	h	g	h	h	h	8	5	0	h
10	h	h	g	h	g	p	p	h	h	h	h	h	h	h	9	2	2	h
% correct	----	88%	50%	75%	38%	63%	50%	100%	75%	63%	63%	88%	75%	75%				100%

P, composite lymphoma: PMLBCL component; H, composite lymphoma: CHLNS component; p, PMLBCL; g, MGZL; h, CHLNS.



Online Supplementary Figure S1. Comparison of methylation levels of 1421 CpG from CHLNS tissue samples isolated by laser-assisted microdissection (LAM) or whole tissue scrape technique. The methylation pattern of tumor tissue from CHLNS (n=10) isolated by LAM or by whole tissue scrape was analyzed. (A) $CD30^+$ HRS cells in a CHLNS before microdissection (left panel), and after microdissection (right panel) using Leica LMD 6000 (CD30 staining on polyethylene naphthalate membrane slide for LAM without hematoxylin counterstaining and without cover slide; original magnification: x63). (B) Correlation of DNA methylation in microdissected tumor cells or whole tissue samples. Representative plots of individual cases show minimum of correlation (left plot) or maximum of correlation (right plot). (C) Immunohistochemistry of the corresponding cases from (B) demonstrates the presence of few tumor cells (left panel, low correlation) and a dense tumor infiltrate (right panel, high correlation) (CD30 staining of HRS cells, original magnification: x10). (D) Venn diagrams show the overlap of hypomethylated (left diagram) or hypermethylated (right diagram) CpG when comparing the methylation levels (β-values) of CHLNS tissue samples after LAM or whole tissue scrape. Data show mean β-values (n=10). LAM was essential to avoid a great number (237) of non tumor-specific hypomethylated targets as detected by whole tissue analysis (scrape technique) and to increase the identification of tumor-specific hypermethylated targets (37).