Polymorphisms in the 3'-untranslated region of the *CDH1* gene are a risk factor for primary gastric diffuse large B-cell lymphoma

Gunnar Jacobs,[1#] Stephan Hellmig,[2#‡] Klaus Huse,[3] Andrea Titz,[2] Andre Franke,[1] Ruta Kwiatkowski,[1] Stephan Ott,[2] Markus Kosmahl,[4] Wolfgang Fischbach,[5] Ralph Lucius,[6] Wolfram Klapper,[7] Ulrich R. Fölsch,[2] Jochen Hampe,[2] Stefan Schreiber,[1,2*] and Philip Rosenstiel[1*]

[1]Institute of Clinical Molecular Biology, Christian-Albrechts University, Kiel, Germany; [2]Department of General Internal Medicine, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; [3]Leibniz Institute for Age Research, Fritz Lipmann Institute, Jena, Germany; [4]Institute for Pathology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; [5]II. Medical Department, Hospital of Aschaffenburg, Aschaffenburg, Germany; [6]Institute of Anatomy, Christian-Albrechts University, Kiel, Germany, and [7]Institute for Pathology, Hematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

Citation: Jacobs G, Hellmig S, Huse K, Titz A, Franke A, Kwiatkowski R, Ott S, Kosmahl M, Fischbach W, Lucius R, Klapper W, Fölsch UR, Hampe J, Schreiber S, and Rosenstiel P. Polymorphisms in the 3'-untranslated region of the CDH1 *gene are a risk factor for primary gastric diffuse large B-cell lymphoma. Haematologica 2011;96(7):987-995. doi:10.3324/haematol.2010.033126*

Online Supplementary Table S1. Disease stages and histological grade of the 144 patients with primary gastric B-cell lymphoma from the European multicenter study. Low-grade: extranodal marginal zone lymphoma of MALT type; high-grade: primary gastric diffuse large B-cell lymphomas.

Patients with primary gastric B-cell lymphoma (n=144)			
Disease stage	N.	Low-grade	High-grade
E I	39	28	11
E I1	48	40	8
E I2	26	9	17
E II	3	1	2
E II1	22	7	15
E III	2	1	1
E IV	4	2	2

Online Supplementary Table S2. SNP database (dbSNP) identities, localization and context sequences of the SNP genotyped. Alleles are coded according to the plus strand (5'→3' orientation of *CDH1*).

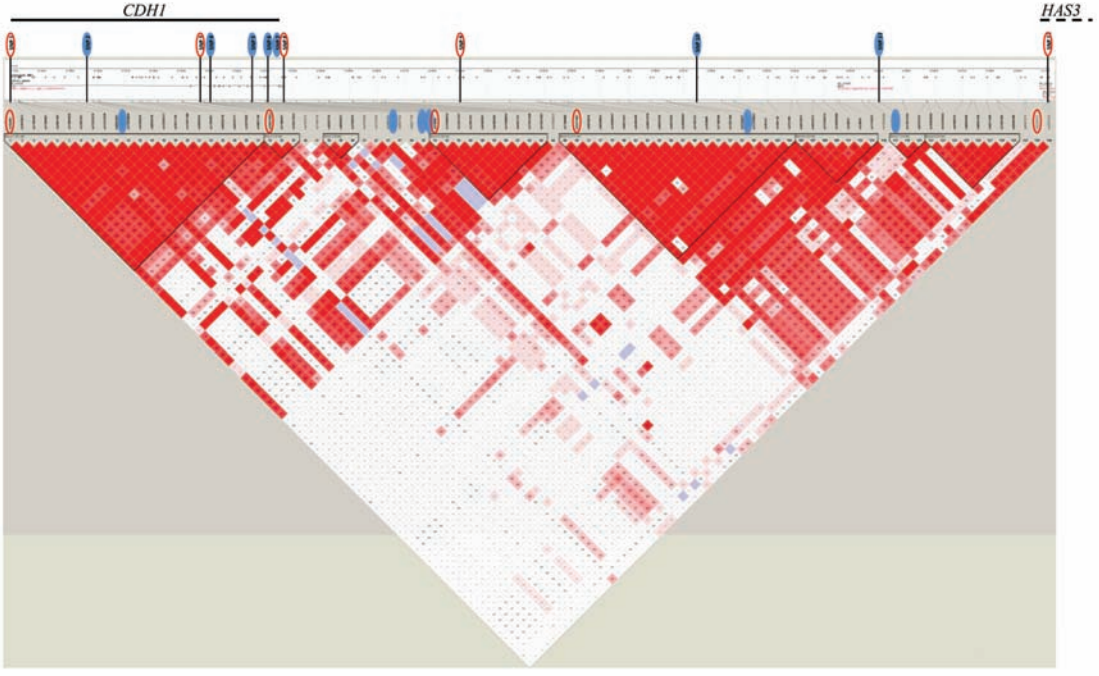
	SNP		Context sequence
1	*CDH1* rs16260	-160	AGGGTCA[A/C]CGCGTCTAT
2	*CDH1* rs1559366	intron 2	CTTTA[G/A]GTATTTCTAGACATG
3	*CDH1* rs10431923	intron 3	CTGAGGGATCACTTCCACGTGCAGT[G/T]TTTTCAGGCATGACTCCCTTTCATT
4	*CDH1* rs7188750	intron 5	CCTCTTGACCTGTTGCTAAGGAGAA[A/G]TGATGGGAGAACGTGGGACAGTTTG
5	*CDH1* rs1801552	exon 13 Ala 692 Ala	CCGC[T/C]GGCGTCTGTAG
6	*CDH1* rs33965115	intron 13	GCATGGCTCATCTCTAAGCTCAGGA[A/G]GAGTTGTGTCAAAAATGAGAAAAAG
7	*CDH1* rs1801026	3'-UTR	AAATCA[T/C]GTTGCTGG
8	*CDH1* rs13689	3'-UTR	GAAGTT[C/T]GTGTCTTT
9	intergenic rs1477407		AATTGGTGATATCATCATTAGGCCT[A/G]TACATGACTTCTGAACGCTGCTGTA
10	intergenic rs12935840		CTGTTTTTGTCAATTTCTAAGTCAT[A/G]TAAATTCCTGGGTTATTTGGTTGCT
11	intergenic rs8062856		GGGGACAGACCCAAGCTGAGGGAAG[C/T]TGATGAACAGCAAAAAGTCAAAGCC
12	*HAS3* rs2232228	exon 2 Ala 93 Ala	GCTCGGTGGCACTGTGCATTGCCGC[G/A]TACCAGGAGGACCCTGACTACTTGC

Online Supplementary Table S3. Genotype counts and frequencies of all markers in patients with extranodal marginal zone lymphoma of MALT type (MALT lymphoma) from the European multicenter study (EMCS) and healthy blood donors. Alleles are coded according to the plus strand (5'→3' orientation of *CDH1*). Differences between n per category and sum of genotype scores per category and SNP are due to missing genotypes. *P* values calculated for carriership of the rare allele. OR: odds ratio; 95% CI: 95% confidence interval. HWE: Hardy-Weinberg equilibrium (Pearson), CR: call rate.

SNP	Locus	Geno-type	Healthy controls (n=361)	MALT lymphoma (n=88)	CR control	CR cases	HWE control	*P* value	OR (95% CI)
1	*CDH1*	A/A	26 – 7.3%	6 – 7.0%	100%	98%	0.924	0.225	0.477
rs16260	-160	A/C	142 – 39.8%	37 – 43.0%					(0.1-1.6)
		C/C	189 – 52.9%	43 – 50.0%					
2	*CDH1*	A/A	64 – 18.0%	18 – 22.0%	98%	93%	0.774	0.406	1,284
rs1559366	intron 2	A/G	177 – 49.7%	40 – 48.8%					(0.7-2.3)
		G/G	115 – 32.3%	24 – 29.2%					
3	*CDH1*	G/G	87 – 24.6%	21 – 24.4%	100%	98%	0.405	0.976	0,991
rs10431923	intron 3	G/T	169 – 47.7%	40 – 46.5%					(0.6-1.7)
		T/T	98 – 27.7%	25 – 29.1%					
4	*CDH1*	A/A	7 – 2.0%	1 – 1.1%	100%	99%	0.523	0.614	0.585
rs7188750	intron 5	A/G	97 – 27.0%	26 – 29.9%					(0.1-4.8)
		G/G	255 – 71.0%	60 – 69.0%					
5	*CDH1*	T/T	65 – 18.5%	10 – 11.5%	100%	99%	0.919	0.119	0.571
rs1801552	exon 13	T/C	171 – 48.7%	47 – 54.0%					(0.3-1.2)
		C/C	115 – 32.8%	30 – 34.5%					
6	*CDH1*	A/A	4 – 1.1%	0 – 0.0%	100%	98%	0.056	0.323	0.453
rs33965115	intron 15	A/G	41 – 11.5%	12 – 14.0%					(0.0-8.5)
		G/G	311 – 87.4%	74 – 86.0%					
7	*CDH1*	T/T	7 – 2.0%	1 – 1.2%	100%	99%	0.500	0.606	0.578
rs1801026	3'UTR	T/C	97 – 27.3%	25 – 28.7%					(0.1-4.8)
		C/C	251 – 70.7%	61 – 70.1%					
8	*CDH1*	C/C	11 – 3.1%	3 – 3.5%	100%	99%	0.851	0.847	1,136
rs13689	3'UTR	C/T	107 – 29.6%	27 – 31.0%					(0.3-4.2)
		T/T	243 – 67.3%	57 – 65.5%					
9	intergenic	A/A	9 – 2.5%	1 – 1.2%	100%	97%	0.242	0.447	0,455
rs1477407		A/G	114 – 32.3%	30 – 35.3%					(0.1-3.6)
		G/G	230 – 65.2%	54 – 63.5%					
10	intergenic	G/G	5 – 1.4%	2 – 2.3%	100%	98%	0.288	0.544	1.661
rs12935840		G/A	58 – 16.4%	17 – 19.8%					(0.3-8.7)
		A/A	291 – 82.2%	67 – 77.9%					
11	intergenic	T/T	71 – 20.2%	22 – 25.9%	100%	97%	0.639	0.248	1.381
rs8062856		T/C	179 – 50.8%	40 – 47.1%					(0.8-2.4)
		C/C	102 – 29.0%	23 – 27.0%					
12	*HAS3*	G/G	64 – 18.4%	14 – 16.3%	100%	98%	0.890	0.648	0.863
rs2232228	exon 2	G/A	169 – 48.6%	45 – 52.3%					(0.5-1.6)
		A/A	115 – 33.0%	27 – 31.4%					

Online Supplementary Table S4. Chr. 16: 68,867,175-68,869,437 (assembly: Feb. 2009). The 3'-terminal region of the RefSeq of the *CDH1* gene (NM_004360.3) was aligned to genomic DNA. Sequences were obtained from Genome Browser (http://genome.ucsc.edu). The coding sequence is underlined, the stop-codon (***TAG***) is in bold and italics, the poly-adenylation signal (TTATTT) is in red and bold, rs1801026 [C/T] (SNP 7) and rs13689 [G/A] (SNP 8) is in red.

ttgtcccttcttctttagAATCTGAAAGCGGCTGATACTGACCCCACAGCCCCGCCTTATGATTCTCTGCTCGT
TTTGACTATGAAGGAAGCGGTTCCGAAGCTGCTAGTCTGAGCTCCCTGAACTCCTCAGAGTCAGACAAAGACCA
GACTATGACTACTTGAACGAATGGGGCAATCGCTTCAAGAAGCTGGCTGACATGTACGGAGGCGGCGAGGACGA
TAGGGGACTCGAGAGAGGCGGGCCCCAGACCCATGTGCTGGGAAATGCAGAAATCA[C/T]GTTGCTGGTGGTT
TTCAGCTCCCTTCCCTTGAGATGAGTTTCTGGGGAAAAAAAAGAGACTGGTTAGTGATGCAGTTAGTATAGCTT
ATACTCTCTCCACTTTATAGCTCTAATAAGTTTGTGTTAGAAAAGTTTCGAC**TTATTT**CTTAAAGCTTTTTTTT
TTTCCCATCACTCTTTACATGGTGGTGATGTCCAAAAGATACCCAAATTTTAATATTCCAGAAGAACAACTTTA
CATCAGAAGGTTCACCCAGCACCTTGCAGATTTTCTTAAGGAATTTTGTCTCACTTTTAAAAAGAAGGGGAGAA
TCAGCTACTCTAGTTCTGTTGTTTTGTGTATATAATTTTTTAAAAAAAATTTGTGTGCTTCTGCTCATTACTAC
CTGGTGTGTCCCTCTGCCTTTTTTTTTTTTTTAAGACAGGGTCTCATTCTATCGGCCAGGCTGGAGTGCAGTGG
GCAATCACAGCTCACTGCAGCCTTGTCCTCCCAGGCTCAAGCTATCCTTGCACCTCAGCCTCCCAAGTAGCTGG
ACCACAGGCATGCACCACTACGCATGACTAATTTTTTAAATATTTGAGACGGGGTCTCCCTGTGTTACCCAGGC
GGTCTCAAACTCCTGGGCTCAAGTGATCCTCCCATCTTGGCCTCCCAGAGTATTGGGATTACAGACATGAGCCA
TGCACCTGCCCAGCTCCCCAACTCCCTGCCATTTTTTAAGAGACAGTTTCGCTCCATCGCCCAGGCCTGGGATG
AGTGATGTGATCATAGCTCACTGTAACCTCAAACTCTGGGGCTCAAGCAGTTCTCCCACCAGCCTCCTTTTTAT
TTTTTGTACAGATGGGGTCTTGCTATGTTGCCCAAGCTGGTCTTAAACTCCTGGCCTCAAGCAATCCTTCTGCC
TGGCCCCCCAAAGTGCTGGGATTGTGGGCATGAGCTGCTGTGCCCAGCCTCCATGTTTTAATATCAACTCTCAC
CCTGAATTCAGTTGCTTTGCCCAAGATAGGAGTTCTCTGATGCAGAAATTATTGGGCTCTTTTAGGGTAAGAAG
T[G/A]GTGTCTTTGTCTGGCCACATCTTGACTAGGTATTGTCTACTCTGAAGACCTTTAATGGCTTCCCTCTT
CATCTCCTGAGTATGTAACTTGCAATGGGCAGCTATCCAGTGACTTGTTCTGAGTAAGTGTGTTCATTAATGTT
ATTTAGCTCTGAAGCAAGAGTGATATACTCCAGGACTTAGAATAGTGCCTAAAGTGCTGCAGCCAAAGACAGAG
GGAACTATGAAAAGTGGGCTTGGAGATGGCAGGAGAGCTTGTCATTGAGCCTGGCAATTTAGCAAACTGATGCT
AGGATGATTGAGGTGGGTCTACCTCATCTCTGAAAATTCTGGAAGGAATGGAGGAGTCTCAACATGTGTTTCTG
CACAAGATCCGTGGTTTGTACTCAAAGCCCAGAATCCCCAAGTGCCTGCTTTTGATGATGTCTACAGAAAATGC
GGCTGAGCTGAACACATTTGCCCAATTCCAGGTGTGCACAGAAAACCGAGAATATTCAAAATTCCAAATTTTTT
CTTAGGAGCAAGAAGAAAATGTGGCCCTAAAGGGGGTTAGTTGAGGGGTAGGGGGTAGTGAGGATCTTGATTTG
ATCTCTTTTTATTTAAATGTGAATTTCAACTTTTGACAATCAAAGAAAAGACTTTTGTTGAAATAGCTTTACTG
TTCTCAAGTGTTTTGGAGAAAAAAATCAACCCTGCAATCACTTTTTGGAATTGTCTTGATTTTTCGGCAGTTCA
GCTATATCGAATATAGTTCTGTGTAGAGAATGTCACTGTAGTTTTGAGTGTATACATGTGTGGGTGCTGATAAT
GTGTATTTTCTTTGGGGGTGGAAAAGGAAAACAATTCAAGCTGAGAAAAGTATTCTCAAAGATGCATTTTTATA
ATTTTATTAAACAATTTTGTTAAACCAT



Online Supplementary Figure S1. HapMap of the region chr. 16: 67,328,000-67,702,000 (assembly: Mar. 2006) based on the D' values and relative position of genotyped SNPs. Criteria for the inclusion of markers into the HapMap view were: CEU, HWE *P* value cutoff: 0.05, min. genotype: 95%, max. # Mendel errors: 1, minimum MAF: 0.05. SNPs genotyped in this study and included in the HapMap view are marked with red circles. In this study genotyped SNPs not listed in the HapMap view (blue oval) are positioned according to the relative physical position to other markers of the HapMap view.