Supplementary Material

<u>No specific short haplotype was found in the vicinity of SNP rs1801552</u>

Material and Methods: To address whether a shorter haplotype could be identified in the vicinity of rs1801552, we chose five SNPs (rs9940010, rs13330170, rs9935563, rs9925080, rs9925161) with MAFs between 0.008 and 0.438 that were genotyped in six probands displaying high AI. Primer sequences and amplicon sizes are listed in the Supplementary Table 6 (Haplotype Analysis 2).

Results and discussion: In order to address whether a shorter haplotype could be found in the vicinity of SNP rs1801552, we chose five SNPs (rs9940010, rs13330170, rs9935563, rs9925080, rs9925161) with MAF between 0.008 and 0.438 that were genotyped in 6 probands displaying high *CDH1* AI. We registered no differences between our series and the HapMap control's haplotypes, and therefore we did not proceed with further analysis.

<u>SNPs variation at miRNA target sites in the CDH1 3'UTR do not explain AI</u>

Material and Methods: To assess whether genetic variation within or in the vicinity of microRNAs binding sites, at the 3'UTR of the *CDH1* gene, could modulate the binding of specific microRNAs, we genotyped a polymorphism rich region encompassing six SNPs overlapping several miRNA binding sites predicted by the following bioinformatic tools: PicTar (http://pictar.mdc-berlin.de/), DIANA-MicroT (http://diana.cslab.ece.ntua.gr/), MicroInspector (http://mirna.imbb.forth.gr/microinspector/) and TargetScan (http://www.targetscan.org/ and the Sanger miRNA database http://microrna.sanger.ac.uk/). Searching the Patrocles database (http://www.patrocles.org/) we were able to list two putative microRNA-target interactions near

polymorphic sites in the 3'-UTR of *CDH1* (rs33956791 and rs34182664). We genotyped these SNPs in our control and proband's cohorts, but these two SNPs were not represented in these individuals. Primers used and amplicon size are listed in Supplementary Table 7.

Results and discussion: Although rs1801552_C allele was apparently not in phase with SNPs downstream from the *CDH1* 3'UTR region, we still tested whether genetic variation occurring at miRNA binding sites could exist resulting in modulation of *CDH1* ASE. To accomplish this task, we have applied a number of algorithms and databases to identify putative microRNA binding sites at the 3'UTR of the gene (See Material and Methods), at the same time that we spotted, using HapMap, all SNPs that localized at selected microRNA binding sites. Searching the Patrocles database, we were able to list two putative microRNA-target interactions near polymorphic sites at the *CDH1* 3'UTR (rs33956791 and rs34182664), nevertheless genotyping of these SNPs, both in control and probands' cohorts, demonstrated high degree of homozygosity in all cases excluding the possibility of an allele specific effect. One possible explanation for high *CDH1* AI arises from recent studies showing that the affinity of microRNAs (miRNAs) for a target mRNA sequence is potentially modulated by the presence of a SNP, as some SNPs result in alteration of the minimum free energy hybridization between the miRNA and mRNA. Nevertheless and although SNPs at predicted miRNAs binding sites and adjacent regions were identified, at the *CDH1* 3'UTR sequence, the high degree of homozygosity, both in controls and probands, hampered this possibility.

Supplementary Tables

Supplementary table 1: Genotyping of Cancer-Free Individuals and Patients.

SNP	DNA primers	Primer sequence	PCR product size
rs1801552	F	TGGCCTTAGAGGTGGGTGAC	262 bp
	R	CCAGGAAATAAACCTCCTCCAT	
rs33964119	F	CTCTCAACACTTGCTCTGTC	206 bp
	R	AGAGATCACCACTGAGCTAC	
rs1801026	F	AGACCCATGTGCTGGGAAAT	289 bp
	R	CTGGGTGAACCTTCTGATGC	

Supplementary table 2: ASE Analysis.

SNP	Primers	Primer sequence	PCR product size
RNA rs1801552	F	GGATAACCAGAATAAAGACCA	178 bp
	R	CAGCAAGAGCAGCAGAATCA	
	SBE1 (F)	TGACTGTGAAGGGGCCGC	NA
	SBE2 (R)	CTGTGCCTTCCTACAGACGCC	NA
RNA rs33964119	F	TGCTGCTCTTGCTGTTTCTTC	144 bp
	R	GCAGCTGGCTCAAGTCAAAG	
	SBE	GAGGATGACACCCGGGACAA	NA
RNA rs1801026	F	CTACTTGAACGAATGGGGCAAT	174 bp
	R	TTCCCCAGAAACTCATCTCAAG	
	SBE	GTGCTGGGAAATGCAGAAATCA	NA
gDNA rs1801552	F	TGGCCTTAGAGGTGGGTGAC	262 bp
	R	CCAGGAAATAAACCTCCTCCAT	
	SBE	CTGTGCCTTCCTACAGACGCC	NA

Note: NA, Not applicable.

Supplementary table 3: *CDH1* Germline Promoter Methylation Analysis.

CDH1 **CpG flanking primers**	**Primer sequence**	**PCR product size**
F	GGTAGGTGAATTTTTAGTTAATTAG	226 bp
R	ACTCCAAAAACCCATAACTAACC	

Supplementary table 4: Arms PCR.

CDH1 **ARMS primers**	**Primer sequence**	**PCR product size**
F	AGATTTTAGTAATTTTAGGTTAGAGGGTTA	339 bp
R	AAAACTACAACTCCAAAAACCCATA	
ARMS_A	AGTAATTTTAGGTTAGAGGGTTA**A**	332 bp
ARMS_C	AGTAATTTTAGGTTAGAGGGTTA**T**	332 bp

Supplementary table 5: Haplotype Analysis 1.

SNP	**Primer**	**Primer sequence**	**PCR product size**
rs11641611	F	AATCGACTCCTTGCCCTCAG	100 bp
	R	CAGCACTGGGGAATTTTTGC	
	SBE	CAGTGTACACACGTACACAAA	NA
rs2862778	F	GGGTCGGGTTTCAGAGGAT	75 bp
	R	ACGGTTCCTTCCCCCTAAAA	
	SBE	CACTGAAATTTTTGGGAA	NA
rs2296409	F	GGAGTTGGAACTGGGAGGAA	114 bp
	R	GCAACCACCCCATTGTAGGT	
	SBE	GTGTAGATGGCATCATCCTCATC	NA
rs4783665	F	AAGGGTGGAGGGATTCAACC	129 bp
	R	TCCAAGAATGTCACCATCCCTA	
	SBE	CTTTATCTCCAACTTTCACGC	NA
rs17715450	F	GGAGGTGGTTCTCCGCAAT	140 bp
	R	ATAAAGGGTGGCCCTCGACT	

	SBE	GACACCCATGTACCGTCCT	NA
rs16260	F	AACCGTGCAGGTCCCATAAC	252 bp
	R	CAAGCTCACAGGTGCTTTGC	
	SBE	GGCCTCGCGTAGACGCG	NA
rs12930371	F	AGGCATCCTCATGCCACTCT	215 bp
	R	AACAGACCCACCCTCAAGGA	
	SBE	AGCTCACATCTATACCCTTGTCA	NA
rs1125557	F	CCACTTTGGATTCGGTTGGA	232 bp
	R	TTTCTTTGGAGAGCAACATGC	
	SBE	CTTCCCCTATAATAAAGCATAGCTG	NA
rs10431923	F	TCCTGACCTCAAGCAGTCCA	188 bp
	R	TGAACACTGTGCTGGCAAAA	
	SBE	AAGGGAGTCATGCCTGAAAA	NA
rs2276330	F	TCCTCCCCTGGTCTCATCAT	171 bp
	R	CACAGTCACACACGCTGACC	
	SBE	CTTGGGCTGGAGAAAGCA	NA
rs1801552	F	TGGCCTTAGAGGTGGGTGAC	262 bp
	R	CCAGGAAATAAACCTCCTCCAT	
	SBE	CTGTGCCTTCCTACAGACGCC	NA
rs7203904	F	GCCACAGAGTGAGGCACAAT	200 bp
	R	CCCCACGGTACATGAGAAAGA	
	SBE	TGAGCAATAGAGCAACCAATC	NA

Note: NA, Not applicable.

Supplementary table 6: Haplotype Analysis 2.

SNP	RNA primers	Primer sequence	PCR product size
rs9940010, rs13330170	F	TGGGATTACAGGTGTGAGC	434 bp
	R	TCAACAAGAACACTGAATTGGC	
rs9935563, rs9925080, rs9925161	F	TTTGGGCTAACCTAGGAGCC	384 bp
	R	CACGTGACACTCACATCCCT	

Supplementary table 7: SNPs in MiRNA Binding Sites at *CDH1* 3'-UTR.

***CDH1* MiRNA binding region flanking sites**	**Primer sequence**	**PCR product size**
F	TTTGTGTGCTTCTGCTCATT	303 bp
R	TACTCTGGGAGGCCAAGATG	

Supplementary Figure Legends

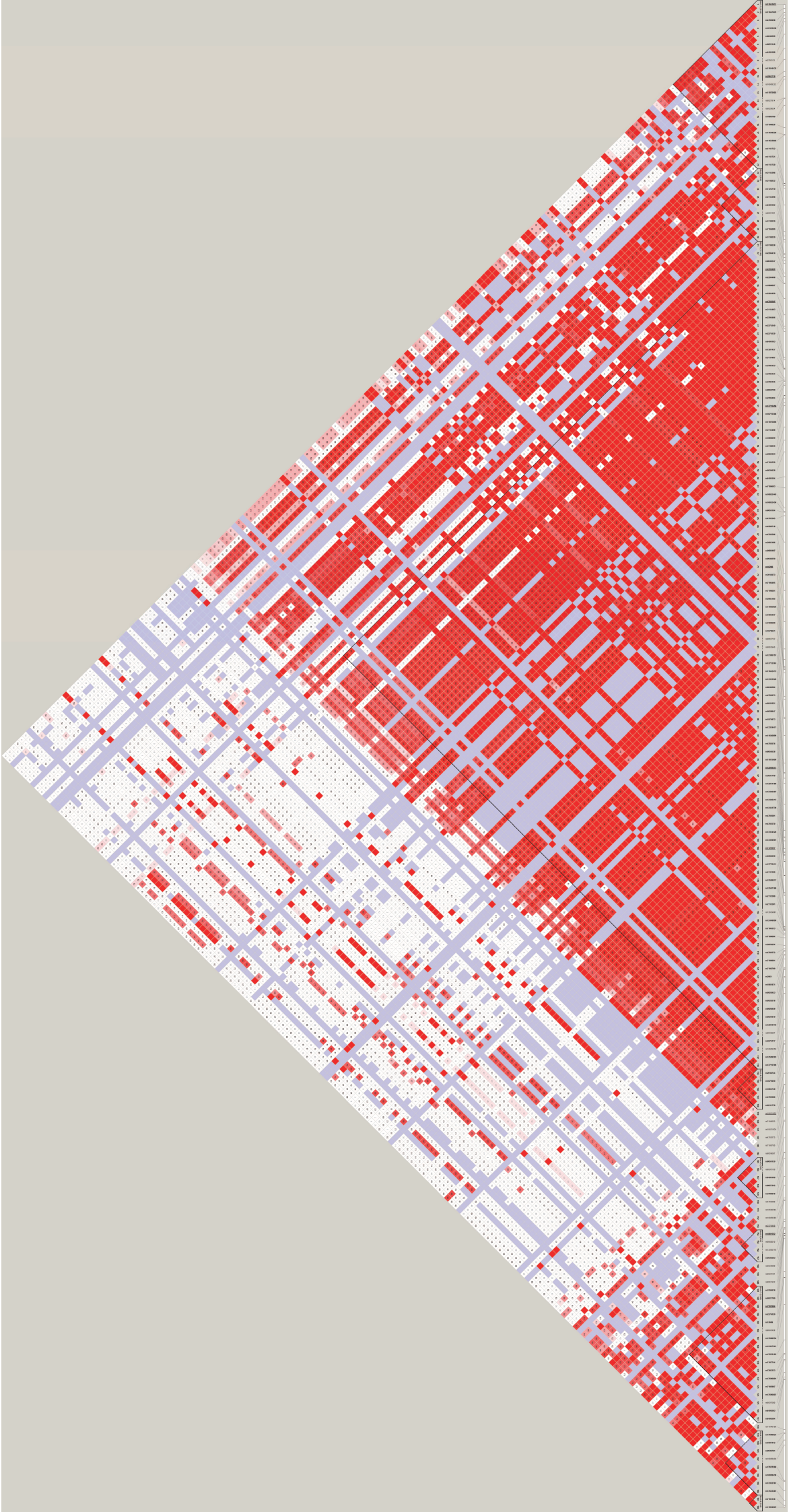
Supplementary Figure 1. Linkage Disequilibrium (LD) plot of HapMap phase II CEU data centered on *CDH1* and *CDH3* region of chromosome 16. The image was built using Haploview 4.1 software,[*] and encompasses the region between positions 67.227.572 and 67.420.442 that includes the 12 SNPs chosen for the Haplotype Analysis. The triangular units designate haplotype blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| =1 bright red; |D'| >1 shades of pink/red). The positions corresponding to the analyzed SNPs are: rs11641611-1; rs2862778-10, rs2296409-34; rs4783665-38; rs17715450-51; rs16260-71; rs12930371-96; rs1125557-106; rs10431923-139; rs2276330-153; rs1801552-154; rs7203904-163 and are underlined in the figure.

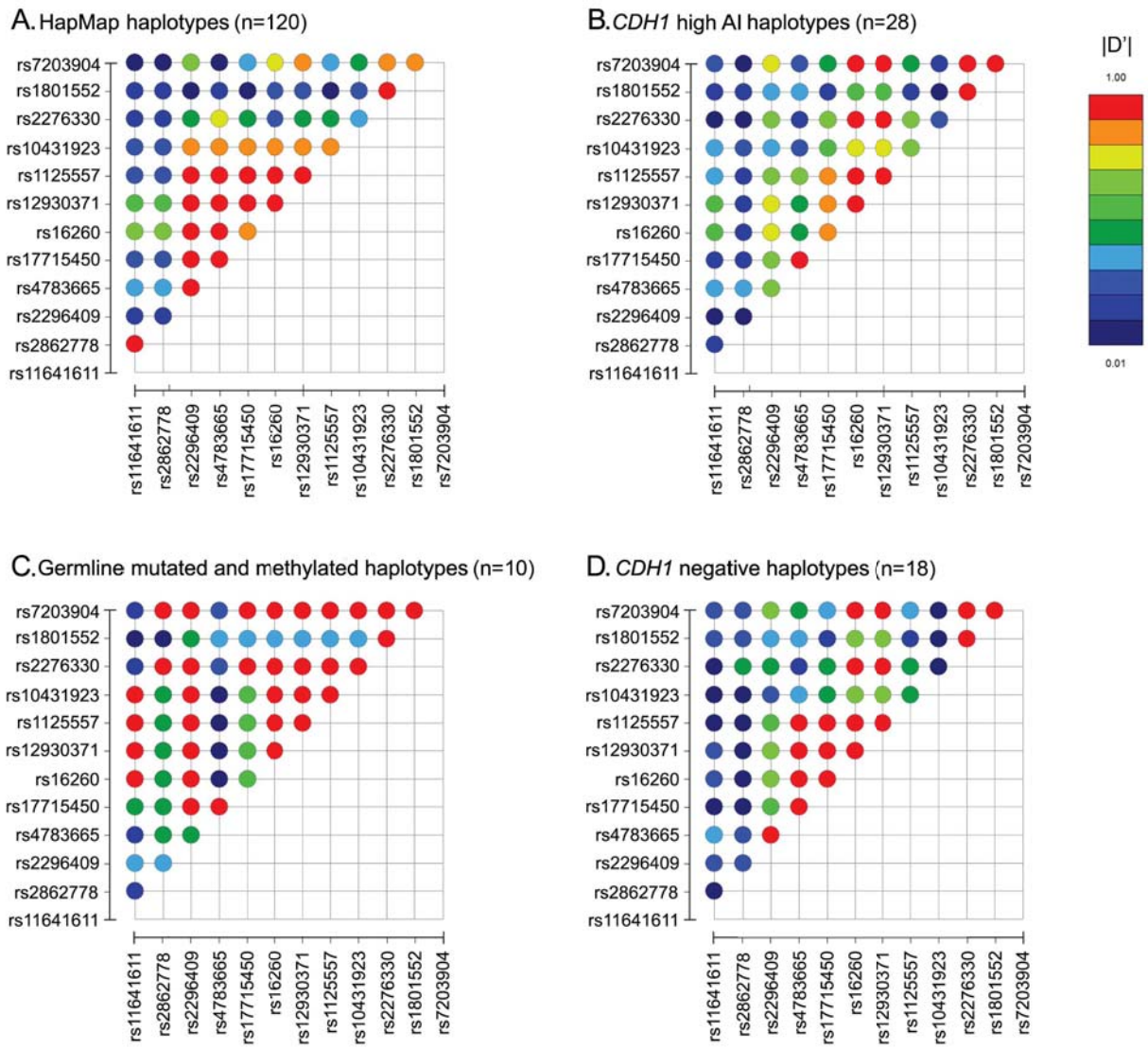
Supplementary Figure 2. Graphical overview of linkage disequilibrium regarding the 12 SNPs analyzed using the GOLD application[24]. **A**, haplotypes from the HapMap database. **B**, haplotypes from *CDH1* high AI probands; **C**, haplotypes from germline mutant and methylated probands; **D**, haplotypes from *CDH1* negative probands. Color code indicates values of the linkage

disequilibrium for |D’| statistic between pairs of two SNPs from dark blue (0-0.10) to bright red (0.90-1.00).

Supplementary Reference

*Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-265.




A. HapMap haplotypes (n=120)
B. CDH1 high AI haplotypes (n=28)
C. Germline mutated and methylated haplotypes (n=10)
D. CDH1 negative haplotypes (n=18)
|D'|
1.00
0.01
rs7203904
rs1801552
rs2276330
rs10431923
rs1125557
rs12930371
rs16260
rs17715450
rs4783665
rs2296409
rs2862778
rs11641611
rs11641611
rs2862778
rs2296409
rs4783665
rs17715450
rs16260
rs12930371
rs1125557
rs10431923
rs2276330
rs1801552
rs7203904