Supporting Information

Atwal et al. 10.1073/pnas.0901298106

SI Methods

Haplotype and Selection Analysis. ***Populations and genotyping.*** The 10 SNPs analyzed in the cohort of 84 previously used (1), lymphoblastoid cell lines from the Coriell Diversity Cell line panel, were genotyped by using TaqMan Allelic Discrimination methods, Assays by Design, Assays on Demand, and standard conditions (Applied Biosystems). The SNPs were chosen to be tagSNPs that captured as much of the haplotype diversity of the MDM4 gene as possible. The clustering of genotype calls was performed by the SDS 2.1 software (Applied Biosystems). The cohort of 299 females of AJ ancestry were chosen to be a control group in a breast cancer study at Memorial Sloan-Kettering Cancer Center (MSKCC), New York, and had participated in cancer screening at MSKCC, were cancer free, and had no first-degree family history of breast cancer. The 13 SNPs genotyped in this cohort represent SNPs present on the EA Affymetrix 500K SNP array, within 90 kb of the MDM4 gene that fit several criteria. First, all 13 SNPs were successfully genotyped as estimated by using Hardy–Weinberg Equilibrium tests and a >96% call rate. Second, these SNPs were preferentially selected to capture as much of the haplotype diversity of this locus in both ancestries ($r^2 > 0.8$) as possible. The genotypes were phased into haplotypes by using PHASE (2), a Bayesian program that utilizes a Markov Chain Monte Carlo algorithm to infer the most probable set of haplotypes by using a coalescent-based prior.

Selection test. The entropy selection test was previously described in detail (1). Briefly, the notion of multiallelic correlation can be quantified by the entropy of the haplotype distribution centered on a particular SNP, and this is then compared with the expected entropy of a set of neutral haplotypes as determined by various Monte Carlo simulations. The coalescent-based simulations (3) are parameterized with similar mutation rates and recombination rates as in the original datasets, and in addition the models can be constrained to incorporate known population demographics of humans (e.g., the simulation for the non-African datasets modeled a rapidly expanding population from a small effective population, mimicking the bottleneck that occurred during the migration out of Africa). The results of selection were seemingly robust to differing models of population dynamics. The classical frequency-based selection tests were carried out by using DnaSP v.4 (4).

AJ Breast Cancer Cases and Controls. ***Populations.*** The case and control populations studied in the GWAS portion of this study were described previously in detail (5). Briefly, for the GWAS familial breast cancer ascertainment, 250 AJ Caucasian women that presented with breast cancer, a family history of 3 or more breast cancers in a single lineage, and tested negative for the 3 AJ BRCA founder mutations were enrolled through the collaborative efforts of cancer clinics in North America and Israel. For the GWAS sporadic breast cancer ascertainment, 243 AJ women were enrolled, who presented at MSKCC with breast cancer, no first-degree family history of disease, and tested negative for the 3 AJ BRCA founder mutations. As controls, the study enrolled 299 healthy AJ women who were participating in cancer screening at MSKCC, were cancer free, and had no first-degree family history of breast cancer.

For the replication analysis, we have included a cohort of 654 AJ sporadic breast cancer cases that were ascertained through anonymized protocols at MSKCC. Patients were enrolled who presented at MSKCC with breast cancer, no first-degree family history of disease, and tested negative for the 3 AJ BRCA founder mutations. A group of 1,085 AJ healthy females, who participated in the New York Cancer Project (NYCP) sponsored by the Academic Medicine Development Company (AMDeC) Foundation, were used as the control population. The NYCP is a cohort study, involving consent for biospecimen collection and follow up, of 8,000 healthy volunteers in the same geographical region as the cases used in this study (6).

Genotyping. The 250 familial breast cancer cases, 243 sporadic breast cancer cases, and 299 controls were genotyped by using the EA Affymetrix 500K SNP array and Affymetrix Commercial Version 500K Genotyping Chips, as a part of a collaborative genome-wide association study for breast cancer that was recently published (5). Genotyping of rs2369244 in the cohort for the replication analysis was performed by TaqMan Allelic Discrimination methods using an Assay by Design and standard conditions (Applied Biosystems). To avoid potential bias by inclusion of data from samples previously genotyped by other methods (Affymetrix 500K and Illumina GoldenGate assays), we have regenotyped all published cohorts by conventional TaqMan allelic discrimination. All genotypes showed 100% concordance. The clustering of genotype calls was performed by the SDS 2.1 software (Applied Biosystems).

Statistics. The analytical pipeline for the GWAS analysis was previously described in detail (5). Deviations from the genotype frequencies in the controls, from those expected under Hardy–Weinberg equilibrium, were evaluated by a χ^2 test (1° of freedom). Linkage disequilibrium (LD) and haplotype analyses were performed by using the combination of HaploView and FastPHASE software under default parameters as previously described. Breast cancer risk associated with rs2369244 was estimated as odds ratios (OR) for the genotype model and per-allele (each copy of rare allele) with the common homozygote as a reference category. Odds ratios were calculated by using conditional logistic regression. All models were adjusted for exact age at diagnosis (cases) or at the time of inclusion in the study (controls) and ethnicity.

Familial and Sporadic Ovarian Cancer.

Sporadic ovarian cancer population. This patient population was previously described in great detail (7). Briefly, paraffin-embedded tissue samples, from patients who were diagnosed with invasive ovarian carcinomas at the Institute of Pathology, Martin-Luther-University Halle-Wittenberg, Germany, between 1997 and 2005, were selected based on the availability of tissue. The study was approved by the local ethical committee. All histological slides were reevaluated by 2 pathologists (E.G. and S.H.) at a multihead microscope. Histology was classified according to the World Health Organization, and grading was assessed according to Silverberg (8). Data retrieved from clinical files included the patient's age, amount of residual tumor, International Federation of Gynecology and Obstetrics (FIGO) stage, adjuvant chemotherapy, and follow-up (Table S1).

1. Atwal GS, et al. (2007) Haplotype structure and selection of the MDM2 oncogene in humans. *Proc Natl Acad Sci USA* 104:4524–4529.

2. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for halotype reconstruction from population data. *Am J Hum Genet* 68:978–989.

3. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
4. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
5. Gold B, et al. (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci USA* 105:4340–4345.
6. Mitchell MK, Gregersen PK, Johnson S, Parsons R, Vlahov D (2004) The New York Cancer Project: Rationale, organization, design, and baseline characteristics. *J Urban Health* 81:301–310.
7. Bartel F, et al. (2008) Both germ line and somatic genetics of the p53 pathway affect ovarian cancer incidence and survival. *Clin Cancer Res* 14:89–96.
8. Silverberg SG (2000) Histopathologic grading of ovarian carcinoma: A review and proposal. *Int J Gynecol Pathol* 19:7–15.

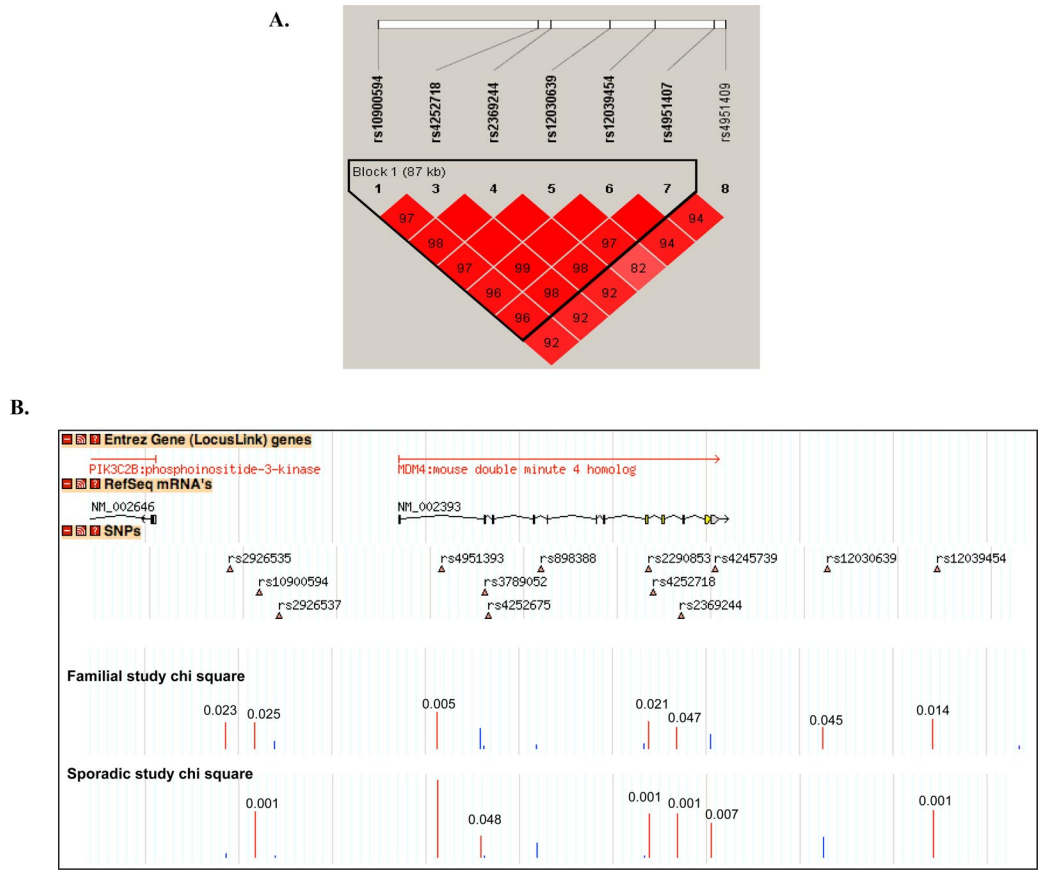

Fig. S1. The SNPs that differentiate the nonneutral MDM4-haplotype from the neutral haplotypes demonstrate significant allelic differences in AJ familial and sporadic breast cancer risk. (*A*) A triangle plot illustrating the LD structure of the MDM4 locus using the associated SNPs in both the case cohort AJ ascertainment ($n = 495$) and the 299 AJ controls. D′ values were calculated by using Haploview. (*B*) An excerpt from a published GWAS analysis focused on the MDM4 locus in the ascertainments of 2 independent breast cancer case cohorts: 250 familial breast cancer cases (a familial study) and 245 sporadic breast cancer cases (a sporadic study) as compared with 299 AJ female controls genotyped on the Affy 500K platform. Statistically significant associations with $P < 0.05$ (red peaks) proportionally reflect the level of significance of association (peak height).

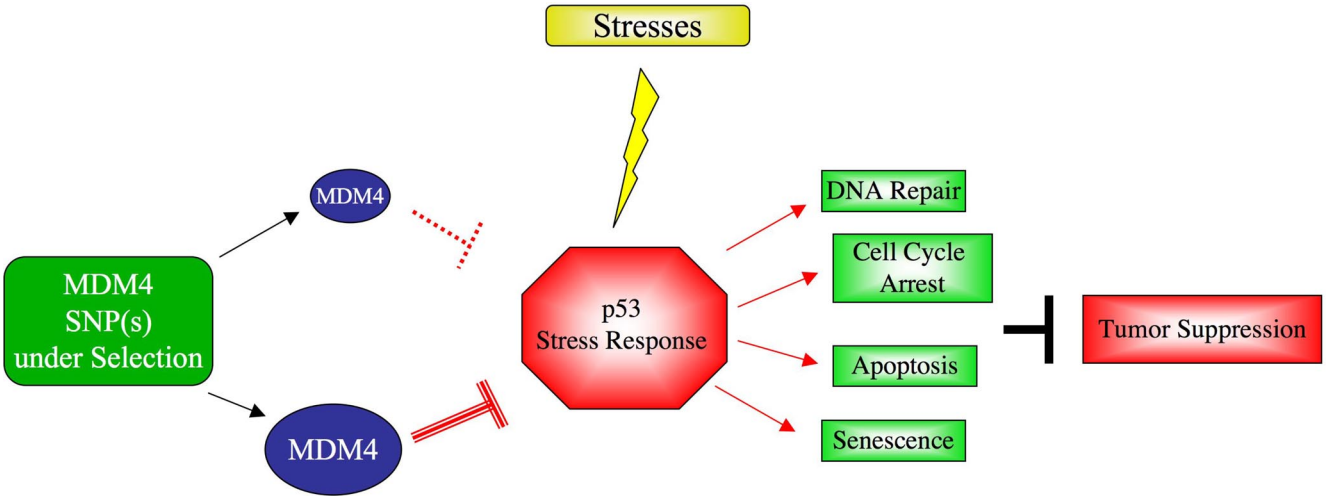

Fig. S2. The data presented in this report support the hypothesis that the haploinsufficient MDM4, like the haploinsufficient MDM2, harbors a genetic variant that affects human cancer. With what is known of MDM4 function to date, it is tempting to speculate that a SNP(s) associated with this haplotype result(s) in allelic differences in MDM4 activity levels. In turn, the differences in MDM4 activity levels would result in differences in the p53 stress response resulting in a higher mutation rate, poorer DNA repair processes, reduced cell cycle arrest, apoptosis, and senescence leading to faster and more frequent tumor formation, which is depicted in the presented model.

Table S1. Significant enrichments of the minor allele homozygotes suggest increased familial and sporadic breast cancer risk

			Number		Percent		
Study	SNP	Genotype	Controls	Cases	Controls	Cases	p-value*
Familial breast cancer, GWAS	rs10900594	GG	134	94	45	39	
		CG	140	108	47	45	
		CC	23	36	8	15	0.0065
	rs2369244	CC	130	95	44	39	
		CG	144	114	48	47	
		GG	24	36	8	15	0.01408
	rs12039454	CC	133	95	44	38	
		CT	143	114	48	46	
		TT	23	39	8	16	0.00294
Sporadic breast cancer, GWAS	rs10900594	GG	126	83	47	35	
		CG	119	119	45	50	
		CC	21	35	8	15	0.00231
	rs2369244	CC	121	80	45	34	
		CG	123	117	46	50	
		GG	22	37	8	16	0.00187
	rs12039454	CC	124	82	46	34	
		CT	122	119	46	50	
		TT	21	37	8	16	0.00118
Sporadic breast cancer	rs2369244	CC	436	229	40	35	
		CG	508	318	47	48	
		GG	141	107	13	16	0.0168

*Two-sided Fisher's Exact Test

Table S2. Association data from Genetic Markers of Suscseptility Study (CGEM) on MDM4 locus based on genome-wide analysis of 500K SNPs (Illumina) in 1,145 breast cancer cases and 1,145 controls

SNP number (Fig. 2)	SNP	Position	Gene	A1	A2	TEST	Cased	Controls	*P* value
5	rs4252675	202,761,189	MDM4	C	T	Per-allele	136/2154	105/2179	0.04228
6	rs898388	202,766,870	MDM4	A	G	Per-allele	482/1808	455/1829	0.3451

Table S3. Summary of clinicopathological data of patients with ovarian cancer

Characteristics	Patients, 121	
	No.	%
Tumor cell type		
serous	67	55.4
endometroid	19	15.7
mixed	14	11.6
clear cell	10	8.3
TCC	1	0.8
UC	8	6.6
MC	2	1.7
Tumor stage		
FIGO stage I	36	29.8
FIGO stage II	11	9.1
FIGO stage III	67	55.4
FIGO stage IV	7	5.8
Patient age, years		
Mean	64.1	
Median	64	
SD	10.7	
Type of Therapy		
Cisplatin + Taxol	69	57
platinum-based chemo w/o Taxol	21	18
other	2	1.7
none (FIGO Ia)	9	7.6
refused/dead	13	11
missing	4	3.3
Residual tumor		
none	47	44
$<$ 1 cm	24	23
$>$ 1 cm	35	33