

# Single Nucleotide Polymorphisms in the *PRDX3* and *RPS19* and Risk of HPV Persistence and Cervical Precancer/Cancer

Mahboobeh Safaeian<sup>1\*</sup>, Allan Hildesheim<sup>1</sup>, Paula Gonzalez<sup>2</sup>, Kai Yu<sup>1</sup>, Carolina Porras<sup>2</sup>, Qizhai Li<sup>3</sup>, Ana Cecilia Rodriguez<sup>2</sup>, Mark E. Sherman<sup>1</sup>, Mark Schiffman<sup>1</sup>, Sholom Wacholder<sup>1</sup>, Robert Burk<sup>4</sup>, Rolando Herrero<sup>2</sup>, Laurie Burdette<sup>5</sup>, Stephen J. Chanock<sup>5</sup>, Sophia S. Wang<sup>6</sup>

**1** Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, United States of America, **2** Proyecto Epidemiológico Guanacaste, Fundación INCIENSA, San José, Costa Rica, **3** Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, **4** Albert Einstein College of Medicine, Bronx, New York, United States of America, **5** Core Genotyping Facility, Division of Cancer Epidemiology and Genetics, Advanced Technology Program, SAIC Frederick Inc, NCI-Frederick, Frederick, Maryland, United States of America, **6** Division of Etiology, Department of Population Sciences, City of Hope, Duarte, California, United States of America

## Abstract

**Background:** Host genetic factors might affect the risk of progression from infection with carcinogenic human papillomavirus (HPV), the etiologic agent for cervical cancer, to persistent HPV infection, and hence to cervical precancer and cancer.

**Methodology/Principal Findings:** We assessed 18,310 tag single nucleotide polymorphisms (SNPs) from 1113 genes in 416 cervical intraepithelial neoplasia 3 (CIN3)/cancer cases, 356 women with persistent carcinogenic HPV infection (median persistence of 25 months) and 425 randomly selected women (non-cases and non-HPV persistent) from the 10,049 women from the Guanacaste, Costa Rica HPV natural history cohort. For gene and SNP associations, we computed age-adjusted odds ratio and p-trend. Three comparisons were made: 1) association with CIN3/cancer (compared CIN3/cancer cases to random controls), 2) association with persistence (compared HPV persistence to random controls), and 3) progression (compared CIN3/cancers with HPV-persistent group). Regions statistically significantly associated with CIN3/cancer included genes for peroxiredoxin 3 *PRDX3*, and ribosomal protein S19 *RPS19*. The single most significant SNPs from each gene associated with CIN3/cancer were *PRDX3* rs7082598 ( $P_{\text{trend}} < 0.0001$ ), and *RPS19* rs2305809 ( $P_{\text{trend}} = 0.0007$ ), respectively. Both SNPs were also associated with progression.

**Conclusions/Significance:** These data suggest involvement of two genes, *RPS19* and *PRDX3*, or other SNPs in linkage disequilibrium, with cervical cancer risk. Further investigation showed that they may be involved in both the persistence and progression transition stages. Our results require replication but, if true, suggest a role for ribosomal dysfunction, mitochondrial processes, and/or oxidative stress, or other unknown function of these genes in cervical carcinogenesis.

**Citation:** Safaeian M, Hildesheim A, Gonzalez P, Yu K, Porras C, et al. (2012) Single Nucleotide Polymorphisms in the *PRDX3* and *RPS19* and Risk of HPV Persistence and Cervical Precancer/Cancer. PLoS ONE 7(4): e33619. doi:10.1371/journal.pone.0033619

**Editor:** Javier S. Castresana, University of Navarra, Spain

**Received:** October 6, 2011; **Accepted:** February 14, 2012; **Published:** April 9, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** The study was funded by intramural NCI and the NIH Office of Research On Women's Health. The authors are grateful to Sabrina Chen from the Information Management Services, Inc. (IMS) for data management and programming support. They are also grateful to Amy Hutchinson and Belynda Hicks for their management of this genotyping effort at the NCI Core Genotyping Facility and for their scientific contributions to the authors' research efforts. Q. Li is partially supported by National Young Science Foundation of China No. 10901155. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Laurie Burdette and Stephen J. Chanock are employed by The Core Genotyping Facility (CGF) which is within the Division of Cancer Epidemiology and Genetics (DECG) of the National Cancer Institute (NCI) and supported by SAIC-Frederick Inc. Information Management Services, Inc. also supported this study, along with Amy Hutchinson and Belynda Hicks with their management of this genotyping effort at the NCI Core Genotyping Facility and for their scientific contributions to the authors' research efforts. This does not alter the authors' adherence to all PLoS ONE policies on sharing data and materials.

\* E-mail: safaeianm@mail.nih.gov

## Introduction

While it is well-known that carcinogenic human papillomaviruses (HPVs) are the causal agents of cervical cancer, HPV infections are extremely common relative to rare cancer incidence, indicating that many infections spontaneously resolve [1], or persist without progression. Host genetic factors may play a role in cervical carcinogenesis and are thought to influence who develops

persistent HPV infection and perhaps who further progresses to cancer [2–7].

The role of host genetic factors and other co-factors associated with cervical cancer are particularly interesting because the stepwise pathogenesis of the disease has been extensively studied. From its initiation through HPV infection at the cervical transformation zone, and subsequent steps related to viral persistence, progression to precancer, and invasion [1], the same

or different factors can be associated with each step towards pathogenesis. The role of non-genetic co-factors in persistence and progression has been well-studied, but there are fewer studies on the host genetics role on the pathogenesis of cervical cancer. Thus, genetic studies of cervical cancer supplement comparisons of cancer cases to non-cancer or uninfected controls, by investigating each intermediate causal step, namely persistent infection and progression to CIN3/cancer.

We used data from the well-characterized, longitudinal cohort study on HPV natural history (NHS) in Guanacaste, Costa Rica, and recently reported results from a panel of 7,140 candidate single nucleotide polymorphisms (SNPs). These polymorphisms were chosen to represent variation in 305 genes based on *a priori* hypotheses of association with HPV infection and cervical cancer (DNA repair, viral infection and cell entry pathways). That effort identified 8 potential genes associated with cervical cancer, including the immune genes 2',5' oligoadenylate synthetase gene 3 (*OAS3*) and sulfatase 1 (*SULT1*), and the epidermal dysplasia verruciformis (EV)-associated *EVER1* and *EVER2* genes, *TMC6* and *TMC8* [8]. We now report our findings regarding the remaining 18,310 SNPs (covering 1,113 genes) that were genotyped on the same iSelect chip. These additional SNPs were selected based on their *a priori* hypothesized relationship with a wide range of cancers, but not specifically with HPV persistence and cervical cancer. These genes were selected based on the collective effort of numerous cancer researchers and include genes in several immune, cytokine and inflammation response, DNA replication and differentiation, macrophage differentiation, toll-like receptor signaling, and T cell receptor signaling pathways, to name a few, and presumably have lower prior probabilities of association with cervical cancer than those studied in the report of Wang et al. [8].

## Results

### Gene region-based associations

Table 1 shows results for 14 genes/regions with  $p < 0.005$  for association with either disease (CIN3/cancer versus controls), disease progression (CIN3/cancer versus HPV persistence), or persistent infection (HPV persistence versus controls) (arranged by p-value for disease associations). This analysis identified 9 gene regions as statistically significantly associated with CIN3/cancer at a p-value of  $\leq 0.005$  (*PRDX3* p-value 0.00015; *RPS19* p-value 0.00045; *DDX1* p-value 0.0006; *TELO2* p-value 0.0009; *C1RL* p-value 0.00165; *ILDR1* p-value 0.00285; *THRAP4* p-value 0.0037; *GDF10* p-value 0.004; and *GDF2* p-value 0.004). Two gene regions were identified as statistically significantly associated with disease progression at p-value of  $\leq 0.005$  (*GC* p-value 0.0004; and *IL2RA* p-value 0.00115). In addition, 2 gene regions were identified as significantly associated with type-specific HPV persistence (*TYMS* p-value 0.0015; and *EVPL* p-value 0.0018). Of the genes identified to be associated with CIN3/cancer *RPS19* was also associated with progression to CIN3/cancer (p-value 0.006). Similarly, of the genes associated with CIN3/cancer, *C1RL*, *GDF10*, and *GDF2* were also associated with persistence (p-values of 0.00875, 0.0423, and 0.04080, respectively). Only two of the genes - *PRDX3* and *RPS19*, were notable with a  $FDR \leq 0.2$  comparing CIN3/cancers to random controls. All gene-based results are shown in Table S1.

### SNP-based associations

Six SNPs in 6 genes identified by the gene-based analysis were significantly associated with CIN3/cancer, disease progression, or persistent oncogenic HPV infection (Table 2) in SNP-based analysis. The *PRDX3* rs7082598 minor allele variants was

associated with statistically significant decreased risk of cervical cancer compared to controls ( $OR_{CT/TT} = 0.41$ , 95%CI 0.30–0.58,  $P_{trend} < 0.0001$ ), and this protective effect was observed for both progression (CIN3+ compared with HPV-persistence) and HPV-persistence (persisters compared to controls) ( $OR_{CT/TT} = 0.58$ , 95%CI 0.40–0.83  $P_{trend} = 0.008$ , and 0.71, 95%CI 0.52–0.97,  $P_{trend} = 0.02$  respectively). The *RPS19* rs2305809 T allele was also associated with statistically significant decreased risk of cervical cancer ( $OR_{CT} = 0.66$ , 95%CI 0.48–0.91;  $OR_{TT} = 0.43$ , 95%CI 0.28–0.66,  $P_{trend} = 0.00007$ ). Further examination by intermediate steps showed that this effect remained for progression from persistence to CIN3+ ( $OR_{CT} = 0.83$ , 95%CI 0.59–1.15;  $OR_{TT} = 0.51$ , 95%CI 0.33–0.78,  $P_{trend} = 0.003$ ). Lastly, the *IL2RA* rs2476491 T variant was also associated with significantly decreased risk of cervical cancer ( $OR_{AT/TT} = 0.69$ , 95%CI 0.51–0.92,  $P_{trend} = 0.02$ ). This association was strongest for progression from HPV-persistence to CIN3+ ( $OR_{AT/TT} = 0.53$ , 95%CI 0.39–0.71,  $P_{trend} = 0.00002$ ).

Variants alleles in *TELO2* rs4786772 ( $OR_{AG} = 1.54$ , 95%CI 1.14–2.09;  $OR_{GG} = 2.51$ , 95%CI 1.59–3.94,  $P_{trend} = 0.00002$ ), *C1RL* rs12227050 ( $OR_{AG/GG} = 2.85$ , 95%CI 1.71–4.74,  $P_{trend} = 0.0001$ ), and *TYMS* rs2342700 ( $OR_{CG} = 1.56$ , 95%CI 1.17–2.09;  $OR_{GG} = 2.18$ , 95%CI 1.30–3.64,  $P_{trend} = 0.0002$ ) were associated with significantly increased risk of cervical cancer (Table 2). The increased risk of variants for *TELO2* remained for progression from persistence to CIN3+, while the increased risk for *C1RL* and *TYMS* remained for persistence compared to random controls.

We also carried out a single statistical analysis including all SNPs, and found that the SNP RS7082598 (*PRDX3* gene) retained significance in this analysis when comparing CIN3+ vs. random controls (multiple comparison adjusted p-value 0.011).

## Discussion

In this analysis of genes and SNPs identified to be broadly relevant for cancer etiology, but not specifically for cervical carcinogenesis, we found 14 genes/regions that were significantly associated with CIN3/cancer at  $p < .005$ . Two of these genes – *RPS19* and *PRDX3* – were notable at a  $FDR \leq 0.2$ . Replication of these results is warranted to eliminate the role of chance finding.

Effects of putative genetic etiologic factors for cervical cancer may be mapped to specific transition states from HPV infection of the cervical transformation zone, progression of persistently infected cervical cells to precancer and invasive cancer. In this study we had the opportunity to investigate and identify genes which may influence some of these stages of cervical carcinogenesis. We note that the protective effects observed for *PRDX3* and *RPS19* SNPs remained for both disease progression (from HPV-persistent infection to CIN3+) and for HPV-persistence.

Mutations of the ribosomal gene *RPS19* have been associated with Diamond-Blackfan anemia (DBA), which is a constitutional erythroblastopenia characterized by absent or decreased erythroid precursors, in a subset of patients. Patients with DBA have increased risk of osteosarcoma. This association with DBA suggests a possible extra-ribosomal function for this gene in erythropoietic differentiation and proliferation, in addition to its ribosomal function. In some primary colon carcinomas, higher expression levels of this gene have been observed compared to matched normal colon tissues [9]. *PRDX3* is in the peroxiredoxin family which encodes a protein with antioxidant function, is localized in the mitochondrion, and may function to protect mitochondria from oxidative stress. Sequence comparisons with recently cloned mammalian homologues suggest that these genes consist of a

**Table 1.** Significance levels (p values) for gene regions with  $p \leq 0.005$ , for either an association with (i) cervical precancer/cancer, (ii) progression to cervical precancer/cancer, (iii) HPV persistence.

Gene	Chromosome	# SNPs	(i)	(ii)	(iii)
			CIN3+ (n = 415) vs. RC (n = 425)†	CIN3+ (n = 415) vs. HPV persistence (n = 356)	HPV persistence (n = 356) vs. RC (n = 425)
<i>PRDX3*</i>	10q25-q26	11	<b>0.00015</b>	0.0744	0.16314
<i>RPS19*</i>	19q13.2	4	<b>0.00045</b>	0.00585	0.45823
<i>DDX1</i>	2p24	13	<b>0.0006</b>	0.69087	0.21284
<i>TELO2</i>	16p13.3	20	<b>0.0009</b>	0.17849	0.65932
<i>C1RL</i>	12p13.31	20	<b>0.00165</b>	0.14029	0.00875
<i>ILDR1</i>	3q13.33	11	<b>0.00285</b>	0.05830	0.88461
<i>THRAP4</i>		6	<b>0.00370</b>	0.80776	0.05615
<i>GDF10</i>	10q11.22	20	<b>0.00400</b>	0.13639	0.04230
<i>GDF2</i>	10q11.22	19	<b>0.00400</b>	0.12639	0.04080
<i>TYMS</i>	18p11.32	20	0.0055	0.30233	<b>0.0015</b>
<i>EVPL</i>	17q25	10	0.02055	0.08985	<b>0.00180</b>
<i>GC</i>	4q12-q13	19	0.0343	<b>0.0004</b>	0.00965
<i>IL2RA</i>	10p15-p14	44	0.24394	<b>0.00115</b>	0.44318
<i>PIK3CA</i>	3q26.3	11	0.91060	0.05145	<b>0.00490</b>

\*FDR for both *PRDX3* and *RPS19* = 0.19.

†RC = random controls.

doi:10.1371/journal.pone.0033619.t001

family that is responsible for regulation of cellular proliferation, differentiation, and antioxidant functions [10,11]. Neither of these genes has an obvious relationship to the known carcinogenic processes that lead to cervical cancer.

There are several study limitations to be considered. We combined the CIN3+ cases from the NHS cohort-based study with the CIN3+ supplemental cases drawn from the community at the same period for increased analytic power. There was a higher proportion of CIN3 (a corollary of detection by screening in NHS rather than symptoms) in the NHS cases than supplemental cases; additionally supplemental cases were older, mainly because of the larger proportion of cancers compared to the NHS cohort-based cases where there were higher proportion of CIN3. Tests of associations between NHS and supplemental populations for 94.2% (17,149/18,208) of the SNPs were not significant at 5% level, importantly, all the SNPs associated with the two genes of interest (*PRDX3* and *RPS19*) were similarly distributed between NHS and supplemental cases, justifying combining the two cohorts. Additionally, this study was powered on a combined end-point of CIN3 and cancer; 95.5% (17,394/18,208) of the SNPs were not statistically different between CIN3 and cancer cases. Importantly, none of SNPs associated with the two genes of interest (*PRDX3* and *RPS19*) were statistically significantly different between CIN3 and cancers. Because SNPs were chosen as tagging markers for genetic regions rather than function, the observed associations with SNPs may be due to linkage disequilibrium with other causal unmeasured SNPs. We were underpowered to perform analyses restricted to carcinogenic HPV types due to small sample size as only women enrolled in the original NHS had HPV typing data. Future studies should therefore consider HPV genotypes in analysis. We were also unable to evaluate genetic factors associated with invasive cervical cancer separately. Future studies with large number of cases will be required to address whether certain genes are associated with transition from in situ

invasive cervical cancer. Although we performed haplotype-based analyses (defined by blocks of linkage disequilibrium); results were generally consistent with the gene region- and SNP-based findings. No new regions of interest were identified in haplotype analysis using the sliding window approach of 3 SNPs. The genes evaluated here were also not selected based on their previously reported associations with cervical cancer, but by agnostic analysis, encompassing a global effort to identify genes involved with a range of infection and non-infection-related cancers.

In summary, these data suggest involvement of two genes, *RSP19* and *PRDX3*, or other SNPs in linkage disequilibrium, with cervical cancer risk. Further investigation showed that they may be involved in both the persistence and progression transition stages. If replicated, these results may suggest a role for ribosomal dysfunction, mitochondrial processes, and/or oxidative stress, or as yet unknown function in cervical cancer pathogenesis.

## Materials and Methods

### Study population

Data are from the Guanacaste HPV Natural History Study (NHS), a population-based cohort study in Guanacaste, Costa Rica. Details of the cohort study methods [12,13] and of the sub-population selected for genetic analyses [14,15] have been reported elsewhere. Briefly, NHS is a population-based cohort of 10,049 women recruited over an 18-month period in 1993-4 and followed for seven years. The primary objective of NHS was to study the natural history of HPV infection and cervical intraepithelial neoplasia (CIN). Cervical cells were available for HPV DNA testing, and buffy coat specimens were available for host gene polymorphism studies.

A host genetic sub-study was nested within NHS, as described previously [14,15]. Briefly, individuals selected for the genetic sub-study included: (i) all women in the cohort histologically confirmed to have prevalent or incident CIN3 or cancer (CIN3+;

**Table 2.** Odds ratios and 95% confidence intervals for top-ranked SNPs ( $p \leq 0.0001$ ) with either an association with (i) cervical precancer/cancer, (ii) progression to cervical precancer/cancer, or (iii) HPV persistence: (all models adjusted for age).

Gene	rs #	Random control	HPV persistence	CIN3/cancer	CIN3/cancer	CIN3/cancer vs	HPV persistence
		N (%)	N (%)	N (%)	vs RC	HPV persistence	vs RC
				OR (95% CI)		OR (95% CI)	
<i>PRDX3</i>	<b>RS7082598</b>						
	CC	288 (68)	266 (75)	346 (83)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	CT	127 (30)	87 (24)	64 (15)	0.41 (0.29–0.58)	0.55 (0.38–0.80)	0.74 (0.53–1.01)
	TT	10 (2)	3 (1)	5 (1)	0.45 (0.15–1.35)	1.29 (0.29–5.65)	0.33 (0.09–1.20)
	CT/TT	137 (32)	90 (25)	69 (17)	0.41 (0.30–0.58)	0.58 (0.40–0.83)	0.71 (0.52–0.97)
				p-trend < 0.00001		p-trend = 0.008	
<i>RPS19</i>	<b>RS2305809</b>						
	CC	100 (24)	102 (29)	141 (34)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	CT	228 (54)	172 (48)	213 (51)	0.66 (0.48–0.91)	0.83 (0.59–1.15)	0.74 (0.53–1.05)
	TT	97 (23)	82 (23)	60 (14)	0.43 (0.28–0.66)	0.51 (0.33–0.78)	0.83 (0.56–1.25)
	CT/TT	325 (76)	254 (71)	273 (66)	0.59 (0.43–0.80)	0.72 (0.53–0.99)	0.77 (0.56–1.06)
				p-trend = 0.00007		p-trend = 0.003	
<i>IIZRA</i>	<b>RS2476491</b>						
	AA	262 (62)	196 (55)	289 (70)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	AT	147 (35)	135 (38)	114 (27)	0.69 (0.51–0.93)	0.56 (0.41–0.77)	1.22 (0.91–1.65)
	TT	15 (4)	25 (7)	12 (3)	0.68 (0.31–1.5)	0.34 (0.16–0.70)	2.24 (1.15–4.37)
	AT/TT	162 (38)	160 (45)	126 (30)	0.69 (0.51–0.92)	0.53 (0.39–0.71)	1.32 (0.99–1.76)
				p-trend = 0.02		p-trend = 0.00002	
<i>TELO2</i>	<b>RS4786772</b>						
	AA	185 (44)	151 (42)	134 (32)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	AG	197 (46)	147 (41)	211 (51)	1.54 (1.14–2.09)	1.65 (1.19–2.27)	0.92 (0.68–1.24)
	GG	42 (10)	58 (16)	69 (17)	2.51 (1.59–3.94)	1.40 (0.91–2.15)	1.68 (1.06–2.65)
	AG/GG	239 (56)	205 (58)	280 (68)	1.71 (1.28–2.28)	1.57 (1.16–2.13)	1.05 (0.79–1.39)
				p-trend = 0.00002		p-trend = 0.03	
<i>C1RL</i>	<b>RS12227050</b>						
	GG	402 (95)	310 (87)	357 (86)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	AG	22 (5)	45 (13)	57 (14)	2.91 (1.73–4.89)	1.06 (0.69–1.62)	2.73 (1.60–4.65)
	AA	1 (0.2)	1 (0.3)	1 (0.2)	1.37 (0.08–22.18)	0.77 (0.05–12.91)	1.36 (0.08–21.92)
	AG/AA	23 (5)	46 (13)	58 (14)	2.85 (1.71–4.74)	1.05 (0.69–1.61)	2.67 (1.58–4.51)
				p-trend = 0.0001		p-trend = 0.85	
<i>TYMS</i>	<b>RS2342700</b>						
	CC	227 (53)	145 (41)	169 (41)	1.0 (ref)	1.0 (ref)	1.0 (ref)
	CG	169 (40)	164 (46)	199 (48)	1.56 (1.17–2.09)	1.06 (0.78–1.44)	1.53 (1.13–2.06)
	GG	29 (7)	47 (13)	46 (11)	2.18 (1.30–3.64)	0.87 (0.54–1.40)	2.57 (1.54–4.27)
	CG/GG	198 (47)	211 (59)	245 (59)	1.65 (1.25–2.18)	1.02 (0.76–1.37)	1.68 (1.26–2.23)
				p-trend = 0.0002		p-trend = 0.78	

doi:10.1371/journal.pone.0033619.t002

CIN3 = 140, cancer = 45); (ii) all women in the cohort who at the time of selection into the study had evidence of HPV persistence, defined as women positive for the same HPV type (either carcinogenic or not) at two consecutive visits at least 12 months apart ( $n = 432$ ) (median length of observed persistence: 25 months); and (iii) a random selection of participants without CIN3, or cancer from the baseline cohort ( $n = 492$ ). To increase power for studies of host genetics, we conducted a supplemental substudy that captured all CIN3 ( $n = 240$ ) and cancer cases ( $n = 87$ ) who were not participants in NHS but were independently

diagnosed with CIN3 or cancer at Social Security clinics from the same study area and during the same period in which NHS was conducted [15]. The study was approved by both the US NCI and Costa Rica Institutional Review Boards and all subjects signed informed consent.

#### Laboratory Methods

**DNA extraction from blood.** DNA was extracted from buffy coats with PureGene purification kits/Autopure protocol (Gentra Systems) at SeraCare (Frederick, MD). For the supplemental cases

the DNA extraction was done at the University of Costa Rica using the same kit.

**HPV testing.** PCR-based HPV DNA testing was conducted using the L1 MY09/MY11 consensus primer methods [12,16,17] on cervical cells stored in specimen transport media (Qiagen, USA) from the natural history study only. Because cervical cells were not obtained from the supplemental cases, HPV results are restricted to women within the original cohort.

**Host genotyping.** A panel was designed as part of an effort of numerous investigators based on their expertise in specific cancers. Genotyping of tag SNPs from 990 candidate gene regions hypothesized to be involved in wide number of cancers such as colon, osteosarcoma, esophageal and stomach, biliary, Fanconi anemia, bladder, breast and other cancers, was conducted at the NCI Core Genotyping Facility (Advanced Technology Center, Gaithersburg, MD; <http://snp500cancer.nci.nih.gov>) [18] using a custom-designed iSelect Infinium assay (Illumina, [www.illumina.com](http://www.illumina.com)). The Infinium included a total of 27,904 tag SNPs. Tag SNPs of the genes were chosen from the designable set of common SNPs (minor allele frequency (MAF)>5%) genotyped using the all 3 HapMap populations for tagging dependent on the population of interest for the investigator suggesting the candidate gene (Data Release 20/Phase II, NCBI Build 36.1 assembly, dbSNPb126) using the software Tagzilla (<http://tagzilla.nci.nih.gov/>), which implements a tagging algorithm based on the pairwise binning method of Carlson *et al.* [19]. For each original target gene, SNPs within the region spanning 20 kb 5' of the start of transcription (exon 1) to 10 kb 3' of the end of the last exon were grouped using a binning threshold of  $r^2 > 0.8$  to define a gene region. When there were multiple transcripts available for genes, only the primary transcript was assessed.

**Quality control (QC).** Tag SNPs that failed manufacturing (ordered but failed assay development), failed validation (no amplification or clustering) and assays that had less than 80% completion or 80% concordance with the 270 HapMap samples used for validation were excluded ( $n = 269$ ). SNPs with low completion rate (<90% of samples) were further excluded ( $N = 482$ ). SNPs with QC discordance among our 100 QC duplicates and among HapMap samples <98% were excluded ( $n = 1,703$ ). We also excluded samples with a completion rate <90% ( $n = 7$ ). Hardy–Weinberg equilibrium was evaluated among controls, 49 SNPs showed evidence of deviation from Hardy–Weinberg proportions. Our QC data did not suggest any obvious genotyping error in these 49, and their results are therefore presented. Of the 20,764 SNPs, 18,310 SNPs from 1113 genes were included in our present analysis.

**Final analytic population.** We evaluated a total of 416 women diagnosed with CIN3 or cancer, 356 women with HPV persistent infection, and 425 random controls for whom validated genotyping results were obtained.

## References

- Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S (2007) Human papillomavirus and cervical cancer. *Lancet* 370: 890–907.
- Czene K, Lichtenstein P, Hemminki K (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 99: 260–266.
- Hemminki K, Chen B (2006) Familial risks for cervical tumors in full and half siblings: etiologic apportioning. *Cancer Epidemiol Biomarkers Prev* 15: 1413–1414.
- Hemminki K, Dong C, Vaitinen P (1999) Familial risks in cervical cancer: is there a hereditary component? *Int J Cancer* 82: 775–781.
- Hussain SK, Sundquist J, Hemminki K (2008) Familial clustering of cancer at human papillomavirus-associated sites according to the Swedish Family-Cancer Database. *Int J Cancer* 122: 1873–1878.
- Hildesheim A, Wang SS (2002) Host and viral genetics and risk of cervical cancer: a review. *Virus Res* 89: 229–240.

## Statistical Analysis

**Gene-based analyses.** We obtained a gene-level summary of association using the adaptive combination of p-values [20], which combines gene-level association evidence through adaptive rank truncated product method. We highlight results of the genes that were significant at  $p\text{-value} < 0.005$ . Because some of our results could be due to false-positive findings, we calculated the false discovery rate (FDR) among associations considered significant using the method of Benjamini and Hochberg [21] to the gene region-based tests<sup>19</sup>. We considered an FDR value of <0.2 as notable.

**SNP-based associations.** We calculated odds ratios (OR) and 95% confidence intervals (95% CI) for each genotype with each disease outcome (CIN3/cancers vs. random controls; HPV-persistence vs. CIN3/cancers; HPV-persistence vs. random controls), using the homozygous wild type (WT) genotype as the referent group. We first compared CIN3/cancer cases to random controls. We further evaluated their associations for HPV progression and/or persistence by comparing: the group of CIN3/cancer cases ( $n = 415$ ) to HPV persisters ( $n = 356$ ) for evaluation of SNPs relevant to progression and (ii) HPV persisters ( $n = 356$ ) to random controls ( $n = 425$ ) for evaluating SNPs relevant to persistence. We note that persistence does not always precede CIN3, as it may be a result of a CIN3 lesion.

We conducted both crude and age-adjusted (<30, 30–49, 50+ years) analyses. For each outcome, we calculated the  $P_{\text{trend}}$  based on the three-level ordinal variable (0, 1, and 2) of homozygote wildtype, heterozygote, and homozygote variant in a logistic regression model. Because of the small cell size examining some of the SNPs (less than 5%), we also show combined effect of heterozygous and homozygous variant genotypes; thus, results of the two-level models are discussed. All logistic regression models were unconditional and conducted using SAS version 9.1 (SAS Institute, Cary, NC).

## Supporting Information

**Table S1** Results for all gene based tests for either an association with (i) cervical precancer/cancer, (ii) progression to cervical precancer/cancer, or (iii) HPV persistence. (XLS)

## Author Contributions

Conceived and designed the experiments: M. Safaeian AH S. Wang S. Wacholder RH PG CP M. Schiffman ACR. Performed the experiments: RB LB SC. Analyzed the data: KY QL M. Safaeian S. Wacholder. Contributed reagents/materials/analysis tools: PG CP KY ACR. Wrote the paper: M. Safaeian AH PG CP KY QL ACR M. Sherman M. Schiffman S. Wacholder RB RH LB SJC S. Wang.

11. Kim K, Yu M, Han S, Oh I, Choi YJ, et al. (2009) Expression of human peroxiredoxin isoforms in response to cervical carcinogenesis. *Oncol Rep* 21: 1391–1396.
12. Herrero R, Hildesheim A, Bratti C, Sherman ME, Hutchinson M, et al. (2000) Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *J Natl Cancer Inst* 92: 464–474.
13. Herrero R, Schiffman MH, Bratti C, Hildesheim A, Balmaceda I, et al. (1997) Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste Project. *Rev Panam Salud Publica* 1: 362–375.
14. Wang SS, Hildesheim A, Gao X, Schiffman M, Herrero R, et al. (2002) Comprehensive analysis of human leukocyte antigen class I alleles and cervical neoplasia in 3 epidemiologic studies. *J Infect Dis* 186: 598–605.
15. Wang SS, Bratti MC, Rodriguez AC, Herrero R, Burk RD, et al. (2009) Common variants in immune and DNA repair genes and risk for human papillomavirus persistence and progression to cervical cancer. *J Infect Dis* 199: 20–30.
16. Schiffman M, Herrero R, Hildesheim A, Sherman ME, Bratti M, et al. (2000) HPV DNA testing in cervical cancer screening: results from women in a high-risk province of Costa Rica. *JAMA* 283: 87–93.
17. Herrero R, Castle PE, Schiffman M, Bratti MC, Hildesheim A, et al. (2005) Epidemiologic profile of type-specific human papillomavirus infection and cervical neoplasia in Guanacaste, Costa Rica. *J Infect Dis* 191: 1796–1807.
18. Packer BR, Yeager M, Burdett L, Welch R, Beeran M, et al. (2006) SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucl Acids Res* 34: D617–D621.
19. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, et al. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 74: 106–120.
20. Yu K, Li Q, Bergen AW, Pfeiffer RM, Rosenberg PS, et al. (2009) Pathway analysis by adaptive combination of P-values. *Genet Epidemiol*.
21. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B* 57: 289–300.