The race associated allele of Semaphorin 3B (SEMA3B) T415I and its role in lung cancer in African-Americans and Latino-Americans

Carmen J. Marsit, John K. Wiencke¹, Mei Liu and Karl T. Kelsey*

Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, MA 02115, USA and ¹Department of Neurological Surgery, University of California at San Francisco, San Francisco, CA 94143-0560, USA

*To whom correspondence should be addressed. Tel: +1 617 432 3313; Fax: +1 617 432 0107; Email: kelsey@hsph.harvard.edu

SEMA3B has been implicated as important for neuronal development and as a tumor suppressor in lung cancer. A single nucleotide alteration of this gene leads to the amino acid substitution T415I, and functionally, this variant protein has a reduced ability to act as a tumor suppressor. The prevalence of this variant in populations is unclear and its role in inherited lung cancer susceptibility has not been tested. Utilizing case-control studies of head and neck squamous cell carcinoma in a Caucasian population and of lung cancer in African-American and Latino-American populations, we determined both the prevalence of this polymorphic variant and its association with the case status of these patients. The variant Ile allele occurs at an allele frequency of 0.18 in African-American and 0.39 in Latino-American control subjects but not in Caucasian subjects. In analyses controlling for ethnicity and known lung cancer risk factors, a significant association was observed between case status and possession of the variant allele (OR 0.71, 95% CI 0.51–0.99). In stratified analysis, both Latino-Americans (OR 0.56, 95% CI 0.32–1.01) and African-Americans (OR 0.75, 95% CI 0.50–1.13) also showed a reduced risk of disease associated with the variant Ile allele. Possessing either the heterozygous or homozygous variant genotype confers a >40% reduced relative risk of lung cancer in Latino-Americans controlling for other lung cancer risk factors. This study points to the need for further examination of this gene and its variant in lung cancer and other diseases.

Introduction

The semaphorin (SEMA) family consists, in humans, of 19 genes encoding both secreted and membrane-bound proteins, all containing the signature SEMA domain at their N-terminal end. These proteins, both the membrane-spanning as well as the secreted, are thought to localize near the plasma membrane of the cells in which they were produced, and are likely, therefore, to participate in localized, contact-mediated events (1). SEMA3B is located in the LUCA region of chromosome 3p21.3, which is frequently deleted in human lung cancer, and is considered a putative tumor suppressor gene (2–4). Its role as a tumor suppressor has been supported by studies re-introducing this gene into cancer cell lines lacking its expression and observing reversal of the tumorigenic phenotypes of these cells (5,6). Additionally, studies have suggested a role for hypermethylation in the promoter region of SEMA3B as a mechanism of inactivation of this gene in lung cancer (6–8).

The biochemical function(s) which characterizes SEMA3B’s tumor suppressive properties are unclear. Re-expression of SEMA3B in lung cancer cell lines lacking its expression led to decreases in cell growth as well as increases in apoptotic cell fractions (6). Protein homology between the SEMA domain of SEMA3B and the MET and RON oncogenes also suggests the possibility of disruption of these cell growth signals as a mode of tumor suppression (2). Recent models in lung and breast cancer cell lines proposed that SEMA3B can antagonize the VEGF pathway by competing for binding of VEGF165 to NRP1, thereby blocking the autocrine tumor survival effect of VEGF and possibly disrupting the VEGF induction of tumor angiogenesis (9).

Sekido et al. (4) first reported a rare nucleotide substitution at position 1479, leading to an amino acid substitution of Thr—Ile at codon 415 of SEMA3B in lung cancer cell line NCI-H1648, which was also detected in the constitutional DNA of the patient. This highlighted the need to further explore this substitution to determine if it was polymorphism or an inherited mutation pre-disposing its carriers to cancer. Phenotypic analysis of this substitution revealed less effective suppression of colony formation and induction of apoptosis compared with expression of the wild-type protein, and in a single tumor sample heterozygous at this locus, the wild-type allele was preferentially silenced (6). This amino acid substitution occurs within the SEMA domain of SEMA3B and may alter its ability to dimerize and thereby reduce its biological activity.

Noting that lung cancer is the leading cause of cancer mortality in the US, and age adjusted incidence rates are higher among African-Americans than among Caucasians or Latino-Americans (10), we have systematically explored possible genetic factors responsible for this important difference. Considerable effort has been directed toward this end, but these differences mostly have not been understood. Some of this variation in cancer risk between groups may be explained by differences in the pattern of cigarette smoking or occupational exposures as well as by ethnic and genetic differences in cancer susceptibility (11–13). As SEMA3B may play an important role in the development of lung cancer, and the presence of a putative variant allele could alter the risk for lung cancer as well as potential therapies directed at this pathway, we sought to determine the effect of this variant allele on susceptibility for lung cancer in minority populations in the US, where we have found it occurs as a polymorphic allele. We have also explored the prevalence of this polymorphism in three populations, Caucasians, African-Americans and Latino-Americans.
Subjects and methods

Study populations
All subjects involved in these studies provided written informed consent under a protocol approved by the appropriate Institutional Review Boards. To determine the prevalence of the T415I variant in a Caucasian population, 286 control subjects from a larger case-control study on head and neck squamous cell cancer in the Boston metropolitan area were genotyped. As we hypothesized that the variant may occur too rarely in Caucasians to perform an analysis on the association of this variant with disease, we turned to our population-based case-control study of lung cancer in United States African-American and Latino-American populations in the San Francisco Bay area, to identify associations between genotypes and lung cancer risk. The study population has been described previously (14). Cases were ascertained from those reported to the Northern California Cancer Center, a population-based SEER registry, and rapid case ascertainment methods were employed to maximize the chances of contacting cases before death. Eligibility criteria for cases included: (1) self-identified Latino or African-American ethnicity; (2) residence in Alameda, San Francisco, Contra Costa, Santa Clara or San Mateo counties; (3) aged 21 years or older; and (4) a presumptive diagnosis of primary lung cancer. Epidemiologic data were collected through direct patient interviews and biological samples (buccal smears and blood samples) were collected at a site selected by the participant. Controls from the same geographic area were frequency matched on age and gender to lung cancer cases (with a ratio of 2:1) and were ascertained using both population-based sampling (random digit dialing and HCFA beneficiary records) and community-based recruiting methods (targeting churches, health fairs, senior centers, university employees and patients of primary care physicians serving African-Americans and Latinos). Control interviews were conducted in a manner similar to cases and biological samples were collected. Not including those who were deceased prior to contact, the participation rate for African-Americans was 72% and for Latino-Americans 68%. In controls, the participation was ~58% for African-Americans and 51% for Latino-Americans (14). A total of 312 cases and 808 controls with biological samples were genotyped for the SEMA3B polymorphism.

Sample processing and SEMA3B genotyping
Biological samples were coded and shipped to the laboratory in Boston where they were analyzed without knowledge of the subjects’ ethnicity or case status. DNA was isolated from whole blood or exfoliated buccal cells using the QIAamp DNA extraction system (Qiagen, Valencia, CA) following the manufacturer’s protocol. SEMA3B genotype was determined using an allelic discrimination 5′-nuclease assay (TaqMan) on the ABI PRISM 7000 Sequence Detection System (Applied Biosystems, Foster City, CA), in 96-well format. TaqMan primers and probes were designed using the Primer Express Oligo Design Software v2.0 (Applied Biosystems). Primers employed for the amplification of the SEMA3B region were SEMA3BF 5′-ACCCCTCTACACTGTTACAACCTCTGT and SEMA3BR 5′-TGTCCTCAATGTGTTGACAAGAG. Probes used for the detection of SEMA3B SNP (T415I) were 5′-VIC-CCCCACATGGGGAC-NFQ and 5′-FAM-CCCCCAATGGGGCGAC-NFQ. Appropriate positive and negative controls were included in each batch of samples.

Statistical analysis
Allele frequencies of the variant Ile allele were determined by examining its prevalence among control subjects in the Caucasian, Latino-American and African-American groups and deviation from Hardy-Weinberg equilibrium was calculated using the χ²-test. In the lung cancer study, to test for significant associations between SEMA3B genotypes and case-control status, univariate odds ratios (ORs) were calculated as a measure of the relative risk. Pack-years were calculated based on the self-reported number of years smoking. Former smokers were persons who had quit smoking >1 year prior to the interview. The SEMA3B region was genotyped for the homozygous variant allele, we combined heterozygotes with homozygous variant genotypes for statistical analysis. Since a single variant allele, previously characterized as a mutant, was shown to predispose tumors to loss of the wild-type allele, we combined heterozygotes with homozygous variant individuals in these analyses. Table I presents the adjusted ORs examining the association between the combined I/I and I/T genotype with lung cancer controlling for age, gender, education level, family history of lung cancer, smoking and in the model for all ethnicities and race. Among all cases, 33% carried at least one variant allele of SEMA3B compared with 40% in controls, hence an adjusted OR of 0.71 (95% CI 0.51–0.99).

Results

Allele frequencies of SEMA3B T415I
Genotyping of 286 Caucasian control subjects enrolled in a population-based case-control study of head and neck squamous cell cancer, revealed only the homozygous wild-type (T/T) genotype. In the African-American and Latino-American populations, among controls there was a significant ethnic difference in the distribution of SEMA3B genotypes. In African-American controls, the distribution of genotypes was 66% (361/547) for T/T, 31% (171/547) for T/I, and 3% (15/547) for I/I, whereas in Latino-American controls the distribution was 68.7% (224/328) for T/T, 29% (96/328) for T/I, and 3% (9/328) for I/I. In the African Americans, the frequency of the variant Ile allele was 0.18 while in the Latino-Americans it was 0.39. Tests for deviation from Hardy-Weinberg equilibrium demonstrated no significant deviation in the African-American controls (P ≤ 0.32). Looking overall in Latino-American controls we did observe some deviation from Hardy-Weinberg equilibrium (P ≤ 0.03), which is expected due to the heterogeneity of the population. To assess if in individual subgroups of Latino-Americans there was still deviation from Hardy-Weinberg equilibrium, we examined the allele frequencies that were examined by ethnic sub-groups (Mexican, P ≤ 0.2; Central American, P ≤ 0.3; South American, P ≤ 0.1; Caribbean, P ≤ 0.6 and US Latin, P ≤ 1) and no deviation was observed in the subgroups.

Demographics and smoking history of the population
Demographic information and smoking histories for study subjects used in this analysis of SEMA3B polymorphisms are presented in Table I. As might be expected, cases had marked differences in the number of education years, with some having less than a high school level of education and also having a slightly larger mean household size compared with controls. There was also the expected higher proportion of cases reporting occupational asbestos exposure, as well as a family history of lung cancer compared with controls. Cases were also significantly more often single or current smokers, and among smokers, the mean pack-years of smoking was 38.3 in cases compared with 23.3 in controls. Amongst former smokers, cases demonstrated a mean of years quit smoking of 8.8 years, compared with 18.0 years in controls.

Associations between SEMA3B T415I variant and lung cancer
The primary aim of this study was to compare the prevalence of wild-type and variant SEMA3B alleles in cases and controls. Since a single variant allele, previously characterized as a muta, was shown to predispose tumors to loss of the wild-type allele, we combined heterozygotes with homozygous variant individuals in these analyses. Table II presents the adjusted ORs examining the association between the combined I/I and I/T genotype with lung cancer controlling for age, gender, education level, family history of lung cancer, smoking and in the model for all ethnicities and race. Among all cases, 33% carried at least one variant allele of SEMA3B compared with 40% in controls, hence an adjusted OR of 0.71 (95% CI 0.51–0.99).

Due to the overall difference in lung cancer risk between Latino-Americans and African-Americans, as well as the potential differential effect of confounders on lung cancer risk in these two groups, we carried out a series of multivariate analyses stratified by ethnicity and controlled for potential confounders in order to appropriately assess the lung cancer risk attributable...
to the biological effect of this allele (Table II). The variant genotypes were prevalent in 39% of Latino-American cases, compared with 51% of controls, leading to an adjusted OR of 0.56 (95% CI 0.32–0.99)—lower than the observed adjusted OR in African-Americans of 0.75 (95% CI 0.50–1.13).

Discussion

SEMA3B has been implicated as an important tumor suppressor in the development of lung cancer due to its chromosomal location in the 3p21.3 LUCA region, its now well-defined biochemical phenotype demonstrating both enhanced apoptosis and reduced cell growth upon re-expression in lung cancer cell lines and its antagonism of the VEGF autocrine growth signal of lung cancer. The T415I substitution of SEMA3B, within its SEMA domain, appeared to functionally alter the biochemical properties of SEMA3B, reducing its ability to promote apoptosis and slow cell growth, as well as predisposing tumor cells to lose the wild-type allele (6,9).

Studies in cancer cell lines and primary lung tumors identified the T415I nucleotide substitution of SEMA3B only rarely, and we believe this may be attributable to the ethnicity of the patients from whom these tumors or cell lines are derived. We did not observe this variant allele in a Caucasian population, but did identify this variant at allelic frequencies of 0.18 in African-Americans and 0.39 in Latino-Americans. With the well-documented functional difference of the variant in mind, we hypothesized that, in populations where this variant allele was more prevalent, increased risk of lung cancer incidence may be related to the presence of the allele. The population differences in allele frequency observed here are marked; however, many ancestry associated markers have been mapped in the genome, although we are unaware of any that co-localize to the immediate vicinity of SEMA3B (16).

Interestingly, in the analysis of both African-American and Latino-American populations, we observed the variant allele to be more prevalent amongst controls. Controlling for additional factors related to lung cancer risk, in Latino-Americans, we observed a >40% reduced risk for lung cancer among people carrying the variant allele in either heterozygous or homozygous state. African-Americans, also demonstrated a protective effect of the variant allele, although not as great as that seen in the Latino-Americans.

This result is similar to reports associating polymorphic variants in members of the nucleotide excision repair family with protective effects in human cancer. The variant A—G SNP in the 5′-UTR of the XPA gene was associated with a lower risk of lung cancer (17), and coding polymorphisms in the XPD gene were associated with reduced risk of lung cancer (18) and basal cell carcinomas (19,20). Hence, constitutional variants in tumor suppressor genes that are phenotypically important have been previously associated with protection from lung cancer. Our results suggest that the variant allele of SEMA3B is phenotypically active in a remarkably complex biologic pathway that has not been previously appreciated.

It is unknown how the Thr to Ile substitution would alter the structure of the protein, but this structural change occurring within the SEMA domain is a dramatic alteration from the polar-uncharged threonine to non-polar isoleucine. Such alterations could lead to significant alterations in the protein conformation, thereby affecting the ability of this protein to interact with others having SEMA domains, including other SEMA family members, extracellular receptors such as plexins, or the MET and RON oncogenes (21). We speculate that the variant allele’s protective effect may be the result of

### Table II. Association of SEMA3B T415I polymorphism and lung cancer risk in African-Americans and Latino-Americans

| SEMA3B genotype | Cases n (%) | Controls n (%) | Adjusted OR (95% CI) | P
|-----------------|-------------|----------------|----------------------|-----
| All ethnicities |             |                |                      |     
| T/T             | 171 (67.3)  | 435 (60.1)     | 1.0 (ref.)           |     
| T/I + I/I       | 83 (32.7)   | 289 (39.9)     | 0.71 (0.51–0.99)     | 0.04 |
| African Americans|            |                |                      |     
| T/T             | 124 (70.4)  | 319 (65.5)     | 1.0 (ref.)           |     
| T/I + I/I       | 52 (29.6)   | 168 (34.5)     | 0.75 (0.50–1.13)     | 0.17 |
| Latino Americans |            |                |                      |     
| T/T             | 47 (60.3)   | 116 (49.0)     | 1.0 (ref.)           |     
| T/I + I/I       | 31 (39.7)   | 121 (51.0)     | 0.56 (0.32–1.01)     | 0.05 |

*ORs are adjusted for age by decade, gender, education level dichotomized at the mean value of the controls, family history of lung cancer, and smoking (never, pack-years below the mean in controls and pack-years above the mean in controls) and models are limited to subjects with data for all variables.

STUDENT'S t-TEST WAS USED. ALL TESTS ARE TWO-SIDED AND CONSIDERED STATISTICALLY SIGNIFICANT AT P < 0.05.

SD, STANDARD DEVIATION.
alterations in binding between the variant protein and its receptors, which include NRP1 and NRP2 (neuropilin 2). Altered binding between proteins in this pathway might allow SEMA3B to outcompete autocrine growth factors, such as VEGF165, for the same receptors, thus preventing the actions of these important growth signals. Additional experiments are necessary to clarify how this amino acid substitution might affect the binding of SEMA3B to its various extracellular receptors.

Other secreted members of the semaphorin family, such as SEMA3C are over-expressed in lung cancers, particularly in metastatic disease and are thus considered oncogenes (22,23). To provide its protective effect, it is also possible that the polymorphic protein could interact with these additional semaphorin family members, rendering them incapable of interacting with their appropriate receptors, and therefore unable to transmit their oncogenic growth signals.

The presence of the polymorphic variant allele in a single lung tumor predisposed this tumor to the loss of the wild-type allele (6). As the 3p21.3 region is prone to deletion events in human lung cancer, as well as widespread epigenetic silencing, the observation of wild-type allele loss in a heterozygous individual whose tumor retained the variant allele provides strong evidence that functionally important mechanisms associated with the variant play an important role in the control of the expression of this gene. A larger sample of tumors harboring this polymorphism may help to clarify if the wild-type allele truly exhibits a pre-disposition for silencing or deletion and if the variant allele is expressed in lung tumors.

These results demonstrate the necessity of examining variation in human population and its relation to disease, as well as raise important questions related to the functional significance of SEMA3B variation in human lung cancer. We do observe a protective effect of the variant allele, particularly in Latino-Americans, but it is imperative that this study be duplicated in other larger studies of genetic susceptibility to disease in order to confirm this result. It is uncertain if this difference in carrier prevalence of the SEMA3B variant allele is related to the lower incidence of lung cancer in Latino-Americans compared with African-Americans. As a secreted factor inactivated in lung cancers, SEMA3B is a promising target for therapeutic applications. Additional in vitro and population-based studies will be necessary to better understand the complex regulation of this system, but will undoubtedly prove crucial for the understanding of both ethnic-specific lung cancer and susceptibility in general as well as perhaps elucidate potential targets for treatment of the disease.

Acknowledgements

This work was supported by NIH grants R01 ES006717, ES00002, R01 CA100679 and T32 ES007155 (C.M.). This work was supported by NIH grants R01 ES006717, ES00002, R01 CA100679 and T32 ES007155 (C.M.).

References


Received February 15, 2005; revised March 18, 2005; accepted April 10, 2005

SEMA3B genotype and lung cancer risk in minority populations

1449