Genomics 107 (2016) 76–82

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Ancestral-derived effects on the mutational landscape of laryngeal cancer

Meganathan P. Ramakodi [a,b,c,d], Rob J. Kulathinal [b,c,d], Yujin Chung [b,d], Ilya Serebriiskii [e,f], Jeffrey C. Liu [a,g], Camille C. Ragin [a,c,g,h,*]

[a] Cancer Prevention and Control Program, Fox Chase Cancer Center-Temple Health, Philadelphia, PA 19111, USA
[b] Department of Biology, Temple University, Philadelphia, PA 19122, USA
[c] African-Caribbean Cancer Consortium
[d] Center for Computational Genetics and Genomics, Temple University, Philadelphia, PA 19122, USA
[e] Developmental Therapeutics, Fox Chase Cancer Center- Temple Health, Philadelphia, PA 19111, USA
[f] Kazan Federal University, Kazan, Russia
[g] Department of Otolaryngology - Head and Neck Surgery, Temple University School of Medicine, Philadelphia, PA 19140, USA
[h] College of Public Health, Temple University, Philadelphia, PA 19122, USA

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form 26 November 2015
Accepted 21 December 2015
Available online 22 December 2015

Keywords:
Laryngeal cancer
Cancer genomics
African-Americans
European-Americans
Mutational landscapes
Context nucleotides

ABSTRACT

Laryngeal cancer disproportionately affects more African-Americans than European-Americans. Here, we analyze the genome-wide somatic point mutations from the tumors of 13 African-Americans and 57 European-Americans from TCGA to differentiate between environmental and ancestrally-inherited factors. The mean number of mutations was different between African-Americans (151.31) and European-Americans (277.63). Other differences in the overall mutational landscape between African-American and European-American were also found. The frequency of C > A, and C > G were significantly different between the two populations (p-value < 0.05). Context nucleotide signatures for some mutation types significantly differ between these two populations. Thus, the context nucleotide signatures along with other factors could be related to the observed mutational landscape differences between two races. Finally, we show that mutated genes associated with these mutational differences differ between the two populations. Thus, at the molecular level, race appears to be a factor in the progression of laryngeal cancer with ancestral genomic signatures best explaining these differences.

1. Introduction

Laryngeal cancer afflicts approximately 12,000 new individuals in the United States each year [1,2] with different incidence and survival rates across ethnic groups [1]. This particular cancer type affects more African-American (Afr-Amr) individuals than European-Americans (Eur-Amr) [1] and the five year survival rate for Afr-Amr with laryngeal cancer is consistently lower than that for Eur-Amr [1]. While socio-economic factors and life styles are associated with the higher incidence and lower survival rates among Afr-Amr [3], we have shown that the contribution of an individual's genetics cannot be ignored [4].

The major risk factors for laryngeal cancer are tobacco smoke and alcohol consumption [5,6]. Pro-carcinogens found in tobacco smoke are absorbed by cells, metabolized to form active carcinogens, and subsequently excreted from the body following detoxification [7]. If the active carcinogens are not excreted from the cell, the carcinogenic compounds may bind to and ultimately damage DNA [7]. The effect of alcohol with tobacco is synergistic; it is hypothesized that alcohol accelerates the absorption and action of tobacco-based carcinogens [8]. Defects in the enzyme activity or metabolic pathway of tobacco metabolism may lead to the accumulation of tobacco carcinogens in the body and increase the risk of tumor progression. Higher levels of nicotine and cotinine (the major nicotine-based metabolite that contributes to cancer development) have been reported in Afr-Amr compared to individuals of European descent, irrespective of smoking levels [9–12]. In addition, reduced metabolic clearance of nicotine to cotinine and decreased excretions of nicotine and cotinine have been observed in Afr-Amr, relative to Caucasians, for similar cigarette consumption [11,12]. Genetic studies have identified gene variants associated with reduced rates of nicotine metabolism in populations with significant African descent [13–15]. African-ancestry related genetic variants associated with susceptibility to cancer chemotherapeutic agents have also been demonstrated [16]. In addition, genetic variants associated with increased risk for head and neck cancers in patients of African descent have also been revealed by meta-analysis [17]. These evidences suggest the possible role of genetic ancestry, together with other non-genetic factors, in increased laryngeal cancer risk and poor survival rate among Afr-Amr. Nevertheless, genome-wide analysis to address the disparity issues in laryngeal cancer has

* Corresponding author at: Cancer Prevention and Control Program, Fox Chase Cancer Center- Temple Health, 333 Cottman Avenue, Philadelphia, PA 19111, USA.
E-mail address: Camille.Ragin@fccc.edu (C.C. Ragin).

http://dx.doi.org/10.1016/j.ygeno.2015.12.004
0888-7543/

not been conducted and genome level analyses are warranted to understand the molecular basis of cancer disparity.

The recent advancements in sequencing technologies has enabled researchers to analyze the whole genome/exomes of tumor and matched normal samples of several cancer types including laryngeal cancer [18–21]. Nonetheless, the potential baseline effect of population (racial/ethnic) level genetic variation in laryngeal cancer has not been examined. This study compares the distribution of *de novo* point mutations that have developed in laryngeal cancers among Afr-Amr and Eur-Amr patients to gain insight into the genetic basis of racial disparities in laryngeal cancers.

2. Materials and methods

2.1. Data source

Level-2 mutation calls based on whole exome sequence data analyses for Head and Neck Squamous Cell Carcinoma were obtained from The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/). These publically available mutation data were manually curated by the TCGA experts.

2.2. Laryngeal cancer data

Dataset-1: We stratified clinical patient data along with their unique IDs for laryngeal cancer from the TCGA portal based on race: Afr-Amr (n = 18) and Eur-Amr (n = 91). Of these, only 13 Afr-Amr and 57 Eur-Amr had data in the publicly available TCGA database and we included all 13 Afr-Amr and 57 Eur-Amr patients in Dataset-1.

Dataset-2: Afr-Amr patients possessed a number of common characteristics: current smoker or current reformed smoker; clinical stage III or stage IV; and all were less than 70 years of age. We matched 36 Eur-Amr with similar characteristics to the Afr-Amr patients: all 13 Afr-Amr and 36 Eur-Amr patients represent Dataset-2.

We retrieved the somatic mutations specific for each individual from TCGA data using custom perl/shell scripts. The clinical data contained patient ID and other metadata. We used patient IDs to retrieve corresponding mutations from the TCGA dataset.

2.3. Statistical analyses

On Dataset-1 and Dataset-2, respectively, we conducted a Mann–Whitney U test to compare the differences in each potential factor such as age and pack years between Afr-Amr and Eur-Amr patients. The effects of race, age, smoking status, pack years, and the number of years smoked on the number of mutations of 13 Afr-Amr and 36 Eur-Amr patients (Dataset-2) were studied using multiple linear regression models. We sequentially removed non-significant covariates at a 5% significance level and the final fitted regression model with significant covariates against mutational load (i.e., the number of mutations) was plotted using R.

The following analyses were carried out for all two datasets. We compared the distribution of individual somatic mutations from each individual's tumor for each population. Each mutation was classified as transitions (Ti) (substitution of purine to purine or pyrimidine to pyrimidine) and transversions (Tv) (substitution of purine to pyrimidine or vice-versa), their frequencies were estimated for each individual, and distributions were plotted for each group. Ti and Tv were further classified into all six possible mutational changes, C > T, C > A, C > G, T > A, T > G and T > C, and transitional frequencies were estimated for each individual. A Mann–Whitney U test at a 5% significance level was employed to compare the differences in the: (a) number of mutations, and (b) frequency of mutations for each mutation type between Afr-Amr and Eur-Amr patients. We used custom perl/ shell/R scripts for mutation estimates and STATA 10.0 (Stata Corp, College Station, TX) was used for the Mann–Whitney U test.

2.4. Context nucleotide signatures

We studied the context nucleotide signatures for somatic point mutations in Dataset-2 as this dataset contains matched samples for potential risk factors associated with laryngeal cancer. The development of point mutations highly depends on the localized, or contextual, neighborhood sequence that they are located in [22,23]. Context nucleotides for a mutation are adjacent nucleotides of that mutation (i.e., nucleotides that exist 3′ and 5′ adjacent to the point mutation) and studying the context nucleotide signatures may explain the genomics factor associated with observed mutational landscape differences. We used the Bioconductor package, SomaticSignatures [24] to analyze the context nucleotide signatures in Afr-Amr and Eur-Amr patients. In total, we analyzed 96 context nucleotide signatures (as there are 16 possible combinations of the four nucleotides (A, T, G, C) at the 5′ and 3′ end of each of 6 possible mutation types.). Differences in the frequency of each context nucleotide signature between these two ethnic groups were assessed statistically using a Mann–Whitney U test.

2.5. Significantly differently mutated genes

We created a list of genes mutated in one or more Afr-Amr or Eur-Amr patients (Dataset-2) and the differences in frequency of patients with mutations between Afr-Amr and Eur-Amr groups were studied using chi-square test in R. In addition, we obtained a list of 44 cancer driver genes for Head and Neck Squamous Cell Carcinoma (HNSCC) from the Broad GDAC Firehose (gdac.broadinstitute.org). The Broad Institute has identified these cancer driver genes using HNSCC dataset of TCGA. We analyzed the frequency of patients with mutations in these driver genes in Afr-Amr and Eur-Amr groups and the differences between the two populations were examined by chi-square test for homogeneity with continuity correction in R.

3. Results

3.1. Laryngeal cancer samples

Summary statistics of age and pack years for Afr-Amr and Eur-Amr patients for each dataset are given in Table 1. The age and number of pack years were not significantly different between Afr-Amr and Eur-Amr patients in Dataset-1 and Dataset-2. Other clinical characteristics of patients in Dataset-2 are given in Supplementary Table 1.

3.2. Dataset-1: somatic point mutations and mutational landscapes

The distributions of the number of somatic point mutations per sample for Afr-Amr and Eur-Amr were different (Fig. 1A). Specifically, Eur-Amr possessed more point mutations compared to Afr-Amr. The number of mutations ranged from 46 to 1026 with a mean of 277.63 and a median of 186 for Eur-Amr whereas the number of mutations varied from 29 to 313 with a mean of 151.31 and a median of 150 for Afr-Amr patients. At a significance level of 5%, the medians of the number of mutations in Afr-Amr and Eur-Amr were not significantly different (Table 1; P = 0.063).

We classified the somatic point mutations into transitions (Ti) and transversions (Tv) and the medians of Ti and Tv frequencies from Afr-Amr and Eur-Amr patients were found to be significantly different between these two racial groups (Fig. 1C; P = 0.0454). In particular, Afr-Amr patients had a higher proportion of Ti (median = 53.11; Q1 = 47.18; Q3 = 67.01) than Tv (median = 46.89; Q1 = 32.99; Q3 = 52.82). In contrast, Eur-Amr patients had higher Tv proportions (median = 50.6; Q1 = 42.52; Q3 = 57.92) compared to Ti (median = 49.4; Q1 = 42.08; Q3 = 57.48).