Oral Oncology 74 (2017) 8–14

Contents lists available at ScienceDirect

Oral Oncology

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

Multi-modality analysis supports APOBEC as a major source of mutations in head and neck squamous cell carcinoma

Daniel L. Faden[a,*], Sean Thomas[b], Paul G. Cantalupo[c], Nishant Agrawal[d], Jeffrey Myers[e], Joseph DeRisi[f,g]

[a] Department of Otolaryngology-Head and Neck Surgery, University of California, San Francisco, San Francisco, CA 94158, USA
[b] Gladstone Institute of Cardiovascular Disease, and Roddenberry Center for Stem Cell Biology and Medicine, Gladstone Institute, San Francisco, CA 94158 USA
[c] Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260 USA
[d] Division of Otolaryngology-Head and Neck Surgery, University of Chicago, Chicago, IL 60637 USA
[e] Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77230, USA
[f] Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA 94158, USA
[g] Howard Hughes Medical Institute, Chevy Chase, MD, 20815, USA

ARTICLE INFO

Keywords:
APOBEC
Head and neck squamous cell carcinoma
Human papilloma virus
Mutational signature

ABSTRACT

Objectives: The mutagenic processes underlying head and neck squamous cell carcinoma (HNSCC) are poorly understood. Pan-cancer mutational signature analyses have identified a signature for APOBEC, a cytosine deaminase, in a subset of cancers, including HNSCC. The role of APOBEC activity in HNSCC remains poorly understood. Therefore, we sought to determine the role of APOBEC in HNSCC pathogenesis.

Material and methods: Utilizing bioinformatic approaches we explored the role of APOBEC mediated mutations in tumor exomes, transcriptomes and germline exomes from 511 HNSCC patients in the TCGA.

Results: 58% of HNSCC were statistically enriched for the APOBEC signature. APOBEC3A expression had the highest correlation coefficient with APOBEC mutation rate. Gene specific motif analysis revealed a slight predominance of APOBEC3A mutations. Canonical pathway analysis demonstrated immune pathway upregulation in APOBEC mutation rich samples. Overall mutational burden was positively correlated with APOBEC enrichment.

Conclusions: APOBEC mediated mutations are highly prevalent in HNSCC. APOBEC3A is the most likely gene to be active in HPV+ HNSCC. APOBEC activity correlates with upregulation of immune signaling pathways, supporting the hypothesis that APOBEC activity could be activated as part of the innate immune response.

Introduction

Cancers arise as a result of accumulated genomic alterations, including somatic mutations. Numerous intrinsic and extrinsic processes can contribute to the development of mutations: for example, exposure to tobacco smoke and UV light (extrinsic) or defective DNA mismatch repair and error-prone polymerases (intrinsic). Our understanding of which precise sources cause mutations in many cancers remains limited. Specific mutagenic processes have mutational preferences, called signatures. Initially, signatures for only a few select processes were known, for example CC:GG > TT:AA double nucleotide transitions occur from UV light exposure experimentally and match mutations found in skin cancer genes [1]. With the advent of next generation sequencing (NGS) approaches, thousands of cancer genomes and exomes have been sequenced, producing catalogs of millions of somatic mutations. Some of these mutations are driver mutations, positively selected for due to their survival advantages, while the vast majority of mutations are passengers. Although the translational oncologic community has largely focused on driver mutations, passenger mutations, which have not been altered by clonal selection, are a historical record of the tumor, reflecting the various mutagenic processes active in the tumor since its inception. Capitalizing on this concept, Alexandrov, Nik-Zainal and Stratton, among others, developed mathematical algorithms based on non-negative matrix factorization and demonstrated the ability to extract mutational signatures from catalogues of somatic mutations [2–4]. This approach revealed both known and novel signatures and allowed not only the detection of which mutagenic processes are active in a given cancer, but the relative contribution of each,

* Corresponding author at: 203 Lothrop Street. Suite 500. Eye and Ear Institute, Department of Otolaryngology-Head and Neck Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA 15213, USA.
E-mail address: DLFaden@gmail.com (D.L. Faden).

http://dx.doi.org/10.1016/j.oraloncology.2017.09.002
Received 16 July 2017; Received in revised form 16 August 2017; Accepted 2 September 2017

as all cancers arise from an amalgam of mutagenic processes. As a result of large scale analyses across 40 types of human cancer from The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), 30 unique signatures have now been identified [2–4]. Approximately half of these signatures have known underlying biologic correlates (http://cancer.sanger.ac.uk/cosmic/signatures).

Head and Neck Squamous Cell Carcinoma (HNSCC) is a heterogeneous group of epithelial tumors. The major risk factors for HNSCC are tobacco, alcohol, betel nut and human papilloma virus (HPV) Infection. Pan-cancer mutational signature analyses, as described above, revealed the presence of six mutational signatures in HNSCC: Signature 1 exhibits spontaneous deamination of 5-methylcytosine and is associated with age and seen in all cancers. Signature 4 exhibits transcriptional strand bias for C > A mutations and is associated with tobacco exposure. Signature 5 exhibits transcriptional strand bias for T > C mutations at ApTpN contexts and has an unknown etiology. Signature 7 exhibits large numbers of CC > TT dinucleotide mutations at dipyrimidines and is associated with ultraviolet light exposure. Signatures 2 and 13 are discussed in detail below and are associated with activity of enzymes from the APOBEC (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like) family [2]. A subsequent analysis identified a unique signature in gingiva-buccal SCC related to tobacco chewing (Signature 29), bringing the total to seven [5].

APOBECs make up a family of eleven cytidine deaminases with diverse biologic functions that convert cytosine to uracil. The most well studied of these, Activation Induced Cytidine Deaminase (AID), functions in antibody diversification through somatic hypermutation and class-switch recombination of immunoglobulin genes in B cells [6]. The APOBEC3 subclass (A, B, C, D, F, G, H) have been shown to function in innate immunity against viral infection by (1) inhibiting viral replication through deaminating cDNA intermediates and (2) clearing foreign DNA through deamination-dependent mechanisms [7–11]. APOBEC3s appear to act on single stranded DNA (ssDNA) [12]. In essence, deamination of a C to U leads to C:G becoming a U:G mismatch. DNA base excision repair mechanisms then lead to transversions (G:C) and transitions (T:A) (see recent review by Burns et al for a more detailed description [13]) [6].

The fact that endogenous enzymes can mutate ssDNA within human cells lead to the hypothesis that APOBEC could play a role in the acquisition of somatic mutations. This line of thought is now supported by numerous in vitro and mouse model studies, demonstrating that APOBEC proteins are capable of endogenous mutagenesis [6,14–17]. Subsequent studies in human cancers have demonstrated that the APOBEC3 family appears to be an active mutagenic source in breast cancer [18]. Importantly, APOBEC3 have a well-defined context specificity for 5′TCA and 5′TCG where C is the mutated base (referred to as TCW). These mutations often occur in clusters, called kataegis, in close relationship to rearrangement break points and preferentially appear to affect the lagging strand template [18–22]. Roberts, Gordenin and colleagues capitalized on these findings, developing a statistical method for identifying and quantitating APOBEC mutations without needing mathematical decomposition techniques, which are bioinformatically more demanding, making studies of APOBEC activity in cancer exomes more feasible [23]. Specific studies investigating the prevalence of TCW mutations in human cancers have demonstrated that these mutations appear to be most prevalent in bladder, cervical, breast, lung cancer and HNSCC [23,24]. Recently, Henderson et al examined 250 HNSCC from TCGA finding that APOBEC activity was more prevalent in HPV + HNSCC and that APOBEC mutations were overrepresented in PIK3CA helical domain hotspots, which are known activating mutations [23,25]. This cohort included only 40 HPV positive tumors. The specifics of APOBEC activity in HNSCC, including why APOBEC activity is increased and which APOBEC gene is active, remains poorly understood. Here we investigate APOBEC mediated mutagenesis in 511 HNSCC.

Methods

Datasets

Somatic exome, germline, transcriptome, and paired clinical datasets were originally obtained from the TCGA data portal, for 511 HNSCC. These files now exist in the Genomic Data Commons (https://gdc.cancer.gov). 277 of these samples had a manually curated list of somatic mutations (MAF files) while the remaining 234 were uncurated automated mutation calls (VCF files). These files are referred to as "curated" and "uncurated", respectively, from here on). Germline polymorphisms were obtained from the Germline VCF files directly.

APOBEC enrichment scores

Per sample APOBEC enrichment (E) scores were calculated according to the methods described by Roberts [23]. Briefly, this is the fractions of all C/G mutations (CGm) that were TCW/WCA in nature (TCWm), normalized by the fraction of all C/G sites in the local 50 bp context (CGc) that were TCW motifs (TCWc).

$\varepsilon = \frac{TCW_m}{CG_m} \div \frac{TCW_c}{CG_c}$. Statistical evaluation of the over-representation of APOBEC signature mutations in each sample was performed using a one-sided Fisher's exact test comparing the ratio of C to T/G substitutions and G to A/C that occur within or outside the TCWm in a sample fraction of the genome [23]. P values were corrected using the Benjamini-Hochberg method. Q values < 0.05 were considered significant or as in Roberts [23].

Kataegis

Cluster analysis was performed according to the methods put forth by Roberts [23]. Briefly, TCW motifs separated by less than 10 kb were grouped into mutational groups, and significance of these groups as clusters of mutations was ascertained using the negative binomial distribution at $p < 1e-4$.

APOBEC gene analysis

Gene analysis was performed according to the methods put forth by Chan [26]. Briefly, the enrichment of YTCA and RTCA (where Y is a pyrimidine and R is a purine) mutations were independently calculated by calculating the expected number of such mutations given the total number of TCA mutations and the fraction of all TCA sites that were observed to be either YTCA or RTCA.

Expression analysis

Expression analysis was performed according to the methods put forth by Roberts et al. [23]. Briefly, RNA-seq derived mRNA levels of each APOBEC family member were standardized relative to TATA-binding protein. Linear regressions were then performed relating the APOBEC enrichment values in each exome to the observed expression of APOBEC genes. The slopes of the regression lines and the correlation values indicate the degree to which the expression of any particular APOBEC gene might be related to the observed frequency of APOBEC mutations.

Pathway analysis

Canonical pathway analysis was performed using QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City). For somatic mutation overrepresentation analysis, we first calculated an over-representation ratio for each gene in the genome: (number of samples with a TCW mutation in gene x /the number of samples with any mutation in gene x). We then selected genes with a ratio greater than or