Profile of esophageal squamous cell carcinoma mutations in Brazilian patients

Fernanda Franco Munari1, Wellington dos Santos1, Adriane Feijó Evangelista1, Ana Carolina Carvalho1, Paula Aguiar Pastrez2, Diego Bugatti2, Durval R. Wohnrath2, Cristovam Scapulatempo-Neto1,3, Denise Peixoto Guimarães1,4, Adhemar Longatto-Filho1,5,6,7 & Rui Manuel Reis1,6,7*

Esophageal cancer is an aggressive tumor that has a high rate of incidence and mortality worldwide. It is the 10th most frequent type in Brazil, being squamous cell carcinoma (ESCC) the predominant subtype. There is currently an incessant search to identify the frequently altered genes associated with esophageal squamous cell carcinoma biology that could be druggable. This study aimed to analyze the somatic mutation profile of a large panel of cancer-related genes in Brazilian ESCC. In a series of 46 ESCC diagnoses at Barretos Cancer Hospital, DNA isolated from paired fresh-frozen and blood tissue, a panel of 150 cancer-related genes was analyzed by next-generation sequencing. The genes with the highest frequency of mutations were TP53 (39/46, 84.8%), followed by NOTCH1 (7/46, 15.2%), NFE2L2 (5/46, 10.8%), RB1 (3/46, 6.5%), PTEN (3/46, 6.5%), CDKN2A (3/46, 6.5%), PTCH1 (2/46, 4.3%) and PIK3CA (2/46, 4.3%). There was no significant association between molecular and patients’ clinicopathological features. Applying an evolutionary action score of p53 (EAp53), we observed that 14 (35.9%) TP53 mutations were classified as high-risk, yet no association with overall survival was observed. Concluding, this is the largest mutation profile of Brazilian ESCC patients, which helps in the elucidation of the major cancer-related genes in this population.

Esophageal cancer is an aggressive tumor with high incidence and mortality rates1,2. Worldwide estimates show that esophageal cancer is the seventh most common type, with around 572,000 esophageal cancer cases in 20181,2. In Brazil, esophageal cancer is among the ten most common tumor types, ranking 6th among men and 15th among women3. According to the National Cancer Institute (INCA), for the year 2020–2021, it is estimated that about 11,390 new cases, 8,690 in men and 2,700 in women, also presenting mortality rates close to the incidence rates3. Esophageal cancer’s poor prognosis is due to a lack of specific symptoms in the early stages of the disease and late diagnosis, with less than 20% of cases showing an overall five-year survival4,5. However, when diagnosed early, the five-year survival rate increases to 80–90%5.

Histologically, 90% of cases are esophageal squamous cell carcinoma (ESCC) and 12% of adenocarcinoma (EADC)5. The highest incidence of ESCC occurs in northern Iran to central north China and in developing countries such as Brazil, whereas EADC occurs more frequently in developed countries5,6–7. ESCC affects the middle third of the esophagus and shares several characteristics with head and neck squamous cell carcinoma5,8. The main risk factors are tobacco and alcohol consumption (especially in combination), hot food and beverage intake5,9,10.

Recent extensive comprehensive molecular studies determine the genomic landscape of ESCC11–17. The TP53 was the most frequent mutated gene observed all over the studies, and genes involved in other essential cancer pathways, such as the cell cycle, PI3K, and NOTCH pathways, were also reported11–17. Of note, the integrative TCGA consortium identified three molecular subtypes: ESCC1, associated with alterations in the NRF2 pathway, which regulates adaptation to oxidative stressors, and gene expression profile resembles lung cancer and head
and neck squamous carcinoma; ESCC2, with higher NOTCH1 mutation rate and deregulation of apoptosis; and ESCC3 with upregulation of the PI3K pathway and lower TP53 mutation rate. Interestingly, the three ESCC subtypes tended geographic associations, being the ESCC1 more prevalent in the Asian population, the ESCC2 in Eastern European and South American (Brazil), and the ESCC3 observed only in North America. The characterization of genes involved in ESCC tumorigenesis is crucial to understand its biology and help identify putative cancer biomarkers and targeted therapies.

The molecular profile of ESCC is mainly unknown in Brazil, being mainly restricted to TP53 mutation status. Therefore, this study aimed to identify the somatic mutational profile of Brazilian ESCC patients by analyzing the entire coding region of a panel of 150 cancer-related genes by next-generation sequencing.

**Results**

**Description of the clinicopathological features.** The summary of clinicopathological features of the esophageal squamous cell carcinoma studied is reported in Table 1. We observed that the population was primarily composed of men (39 cases, 84.8%) with an average age of 60 years (minimum 39 years and maximum 77 years); most of them consumed alcohol (37 cases, 82.6%) and tobacco (36 cases, 80.0%). The most used these substances in combination (32 cases, 71.1%). Most patients presented moderately differentiated tumors (26 cases, 59.1%), whose clinical stage was most advanced III and IV (36 cases, 87.8%). Patients’ median survival concerning the tumor diagnosis data until the last information (death or follow-up) was 9.40 months.

| Variable                          | ESCC (n=46) |
|----------------------------------|-------------|
| **Category**                     | n | %   |
| Age (years)                      | Mean (SD) 60 | –   |
|                                 | Min–Max 39–77 | –   |
| Gender                           | Female 7 | 15.2 |
|                                 | Male 39 | 84.8 |
| Alcohol consumption              | No 8 | 17.4 |
|                                 | Yes 37 | 82.6 |
|                                 | Missing 1 | –   |
| Tobacco consumption              | No 9 | 19.6 |
|                                 | Yes 36 | 80.4 |
|                                 | Missing 1 | –   |
| Tobacco and alcohol in combination | No 13 | 28.3 |
|                                 | Yes 32 | 71.7 |
|                                 | Missing 1 | –   |
| Tumor differentiation            | Little 14 | 31.8 |
|                                 | Moderate 26 | 59.1 |
|                                 | Well 4 | 9.1 |
|                                 | Missing 2 | –   |
| TNM Staging*                     | I e II 5 | 10.8 |
|                                 | III e IV 36 | 89.2 |
|                                 | Missing 5 | –   |
| Life status                      | Alive (without cancer) 1 | 2.2 |
|                                 | Alive (with cancer) 6 | 13.0 |
|                                 | Dead (by cancer) 39 | 84.8 |

Table 1. Clinical-pathological features of esophageal squamous cell carcinoma patients. *TNM 7 edition staging. n number of cases, ESCC esophageal squamous cell carcinoma.

**Description of the mutation profile.** We sequenced the whole coding region of 150 cancer-related genes in 46 cases of esophageal squamous cell carcinoma. The mean read depth of sequencing was 911 × per gene and 310.9 × per variant. We found a mean of 1.9 driver mutations per patient (range 0–7), and we identified at least one driver somatic variant in 42 tumor samples. Driver mutations in single genes were found in 41.3% of tumors (19/46), whereas 26.1% (12/46) showed driver mutations in two genes, 17.4% (8/46) in three genes, 6.5% (3/46) in four or more genes and 8.7% (4/46) showed driver mutations in none of the genes analyzed. In total, 25 genes were found to harbor driver somatic mutations (Fig. 1). A complete list of variants (missense, frameshift, nonsense, in-frame, and splice mutations) identified is presented in Supplementary Table S1.

The gene with the highest frequency of driver mutations was TP53 (39 cases, 85%), followed by NOTCH1 (5 cases, 11%), NFE2L2 (5 cases, 11%), RB1 (3 cases, 7%), PTEN (3 cases, 7%), CDKN2A (3 cases, 7%), PCH1 (1 case, 2%) and PIK3CA (2 cases, 4%) (Fig. 1). Below, we describe in more detail the most affected genes and pathways.
Genes involved in the cell cycle. Of the 25 genes with driver mutations identified, ten were involved in the cell cycle (Fig. 1; Supplementary Table S1). Overall, 41 (89.1%) cases showed mutation of this crucial cancer pathway.

A total of 44 somatic TP53 mutations were identified in 39 tumors (85%) (Fig. 1; Supplementary Table S1). The mutations were 28 missense, 9 nonsense, 4 splice sites, and 3 frameshift. The most frequent changes were p.His193Arg, p.His179Arg, p.Arg248Trp, p.Arg273Leu, p.Tyr107Asp, and p.Tyr220Cys, all of them previously reported. We also identified two changes (p.Thr211AsnfsTer5 and p.Thr256HisfsTer8) that have not been previously identified (Fig. 2A).

Five cases (11%) showed somatic mutations in the NFE2L2 gene (Figs. 1, 2 and Supplementary Table S1). The mutations were all missense, being p.Arg34Gln and p.Val32Gly in 2 cases, and p.Glu79Lys and p.Leu30Phe, which was present in one case each (Fig. 2B). Three tumors (7%) exhibited somatic changes in the RB1 gene (Fig. 1; Supplementary Table S1). The mutations were two frameshifts and one splice site (Fig. 2D). The CDKN2A gene was mutated in 3 cases (7%) (Fig. 1; Supplementary Table S1). The types of changes identified in the CDKN2A gene were two nonsense and one frameshift (Fig. 2E). We also observed the PTCH1 gene's somatic mutation in two tumors (2%) (Fig. 1; Supplementary Table S1). The types of changes identified were one missense and one nonsense. Moreover, we observed mutations in TERT, STAT3, DDX3X, CDK12, and BAP1 genes in one case each (Supplementary Table S1).

NOTCH signaling pathway alterations. The Notch signaling pathway involving the genes NOTCH1 and FBXW7 was altered in 6 cases (13%) (Fig. 1; Supplementary Table S1). The most affected gene was NOTCH1 (11%), and the types of changes identified were 3 missense, one frameshift, and one splice site (Fig. 2C). Only one missense mutation was identified in FBXW7.
Figure 2. Lollipop plot of the genes with the highest frequency of mutations.
PIK3-AKT signaling pathway alterations. The PIK3-AKT signaling pathway involving the genes PTEN, PIK3CA, and TSC2 was altered in 6 (13%) cases (Fig. 1; Supplementary Table S1).

PTEN was the gene with the most mutations identified with alterations in 3 cases (7%). The types of changes identified were one missense, one nonsense, and one translocation (Fig. 2F). Mutations in the PIK3CA gene were identified in 2 cases (4%), at hotspot codons p.Glu542Lys and p.Glu545Lys in one case each. The TSC2 gene was also mutated in one case (2%), and the alteration was a p.Thr831Met.

MAPK signaling pathway alterations. The MAPK signaling pathway, with genes KRAS, NF1, and BRAF, was altered in 3 ESCC (7%) (Fig. 1; Supplementary Table S1). All three genes were changed in one case each (2%). All changes were missense type. The type of alterations in the KRAS gene was p.Gly12Ser; in the NF1 gene, it was p.Pro2115Leu, and in the BRAF gene, it was p.Arg260Cys, outside the known hotspot V600 region.

Genes involved in the DNA repair. Of the 25 mutated genes, only 2 (MSH6 and PALB2) were involved in the DNA repair (Fig. 1; Supplementary Table S1).

RTKs signaling pathway alterations. All four genes of the RTKs signaling pathway had a mutation each (2%). The RET gene is the missense p.Arg1050Gln mutation; the EPHA7 gene harbor the splice site p.Val777Leu mutation and the ERBB2 gene have a missense mutation. In the KIT gene, the only mutation was for p.Val825Gly.

Classification of TP53 mutations. TP53 mutations were further classified using the evolutionary action score of p53 (EAp53), which considers only missense mutations such as “low-risk” and “high-risk”19. Of the 23 mutations, we observe 14/23 (60%) as high-risk and 9/23 (40%) as low-risk (Table 2). A second classification was done following Poeta et al., which categorize mutations as “disruptive” and “non-disruptive”20,21, and we observed 21/39 (54%) as disruptive and 18/39 (46%) as non-disruptive (Table 2).

Association of mutation profile and clinicopathological features. Next, we performed an analysis of the association between clinicopathological features and mutation status. Due to the small number of cases analyzed, this analysis was performed only for genes harboring more than 10% of mutations, namely TP53, NFE2L2, and NOTCH1 (Table 3). No significant association was found (Table 3). We also analyzed the association of TP53 categories of (EAp53 risk score and disruptive and non-disruptive) with patients’ clinicopathological features, yet, no significant association was found (Table 4).

In addition, a Kaplan–Meier survival analysis was performed, and no association was observed with TP53, NFE2L2, and NOTCH1, neither with TP53 EAp53 score and disruptive and non-disruptive categories (Table 5 and Fig. 3).

Discussion
In the present study, we performed, for the first time, the somatic mutational profile of a panel of 150 cancer-related genes in a series of 46 Brazilian esophageal squamous cell carcinomas. We identified 25 genes with alterations, the genes with the highest frequencies being TP53, NOTCH1, and NFE2L2.

The frequency of mutation observed was compared with the literature, and similar frequencies were observed with other populations (Table 6). As expected, the TP53 gene was the highest mutated (84.8%, 39 cases). In our study population, its frequency was similar to that reported in other studies such as TCGA14-17,22-25 (Table 6) and higher than previously reported in the Brazilian ESCC population (34–40%)18,24,26. This discrepancy with previous Brazilian studies is probably due to the distinct methodologies used since Sanger sequencing and hotspot TP53 regions were used on those studies, at variance with all Brazilian studies is probably due to the distinct methodologies used since Sanger sequencing and hotspot TP53 regions were used on those studies, at variance with all Brazilian studies.

We also applied for the first time a risk-system score based on Evolutionary Trace (ET) method proposed and validated in head and neck tumors (EAp53)19 to evaluate its impact in ESCC. We showed that 36% of TP53 mutation in squamous cell carcinoma would be considered high-risk; however, there was no significant association between the clinical and pathological condition of the patient and the prognosis. This lack of significance could be related to the limited number of cases analyzed, and further studies are needed to validate our findings. We also classified TP53 mutations in disruptive and non-disruptive, as reported by Poeta et al.,20 and Molina et al.,21 for head and neck and lung cancer, respectively. We observed that 54% were disruptive, yet no significant statistical association was observed with the clinical-pathological characteristics and patient prognosis.

Another important gene of the cell cycle is the CDKN2A gene, also known as p16, which the loss of its function can result from homozygous deletions and mutations. According to the TCGA and genomic studies, the significant alteration is homozygous deletion, leading to 76% loss of function of the CDKN2A gene in ESCC27. At variance with high copy number alterations, CDKN2A mutation frequency is much lower, ranging from 3–8% (Table 6).

The second most mutated gene was NOTCH1 (11%), which is part of the NOTCH signaling pathway, important to regulate the cell cycle senescence27. In esophageal squamous cell carcinoma, the frequency of mutations in this gene varies from 8 to 19% (Table 6), so our population is within the reported range. These findings are in concordance with the ESCCC2 TCGA molecular subtype, observed in the 15 patients from South America (Brazil) analyzed17.

The NFE2L2 gene encodes the NRF2 protein, a transcription factor regulating the expression of antioxidant proteins that protect against damage caused by injuries and inflammation28. Genetic deletion of NRF2 the susceptibility to the development of cancer increases, causing the tumor cell to survive the oxidative stress caused
by chemoradiation leading to resistance to treatment 29–31. The frequency of mutations in the NRF2 gene has been reported in 3–17% of ESCC (Table 6). Driver mutations in NRF2 are believed to be late events in the process of developing esophageal squamous cell carcinoma 22,32. In our population, we observed a frequency of 10.8% of alterations in this gene, and in accordance with TCGA, it is associated with the ESCC1 subtype, typical of squamous carcinomas associated with tobacco exposure 17.

The PI3K-AKT signaling is a complex pathway that regulates cell growth, proliferation, motility, apoptosis, and cell growth and is often related to the development of esophageal squamous cell carcinoma 15,16,33. The frequency of mutations reported in the literature varied from 4 in the present to 13% (Table 6). Recently our group analyzed another series of 38 ESCC and observed a frequency of 10.5% of PIK3CA mutations 34. The majority of the mutations are present in the hotspot exons 9 (E542K and E545K) and 20 (H1047R) 34–36. (Supplementary Table S1). Another gene of this PI3K-AKT pathway is the PTEN, which is frequently mutated in some tumor types, such as melanomas and glioblastomas, yet its mutation rate in ESCC is lower, varying from 1 to 6.5%, as observed in our Brazilian series (Table 6). Of note, the TCGA ESCC3 subgroup is characterized by upregulation of this pathway, and in our study, it was alternated in 13% (PTEN, PIK3CA, and TSC2).

Table 2. Classification of TP53 mutations. NA not applicable, n number of cases. a http://mammoth.bcm.tmc.edu/cgi-bin/panos/EAp53.cgi. b Neskey et al. Evolutionary Action Score of TP53 Identifies High-Risk Mutations Associated with Decreased Survival and Increased Distant Metastases in Head and Neck Cancer, Cancer Res. 2015 Apr 1;75(7):1527–36; n number of variants. c Poeta et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. N Engl J Med 357, 2552–2561, https://doi.org/10.1056/NEJMoa073770 (2007).
The KRAS gene is an essential biomarker in cancer, mainly because it predicts the efficacy in therapies targeting the growth factor EGFR in tumors such as colorectal cancer. According to the TCGA, the frequency of mutations in the KRAS gene is low, 7%, which is in line with the results obtained in our study population, which was 2%.

ESCC is usually diagnosed late, and the minority of patients can benefit from treatments such as chemotherapy and radiation therapy. Target therapies are important approaches for several tumors, including ESCC. Significantly, in the present work, the identification of patients harboring PIK3CA mutations could potentially benefit from PI3K and mTOR inhibitors, such as buparlisib, alpelisib, and everolimus. Moreover, the recent development of anti-KRAS agents, such as sotorasib and adagrasib, can bring some hope for patients with KRAS mutations (Table 6).

The present study has several limitations, being the relatively small number of cases analyzed the major issue, and it could explain the lack of significant association of mutation status and patients’ clinical-pathological features. Additionally, a limited panel of 150 cancer-related genes, not whole-genome nor whole-exome sequencing, was performed, so a complete picture of the mutated landscape is lacking. Therefore, further studies with a larger population and broader mutation analysis are needed. Despite these issues, we performed paired germline/tumor analysis, and it is the first to the somatic landscape of an admixture population such as the Brazilian ESCC population. Our findings align with the frequencies reported in other populations, namely Occidental and Asian, and will contribute to understanding the mutational profile of esophageal squamous cell carcinoma in Brazil.

### Material and methods

#### Tissue samples.

Forty-six patients diagnosed with esophageal squamous cell carcinoma treated at the Barretos Cancer Hospital’s upper-digestive department, Barretos, SP, Brazil, were evaluated. The main clinicopathological features were collected from patients’ medical records.

The tumor and blood samples were obtained from biopsy or surgery and immediately processed and stored at – 80 °C in the Barretos Cancer Hospital Biobank. The present study was approved by the Barretos Cancer Hospital Institutional Review Board (Project No. 1.454.967/2016), and all patients included signed an Informed Consent Form. All methods were performed following the relevant guidelines and regulations.

#### DNA isolation.

Tumor DNA was isolated from fresh-frozen tissue using QIAsymphony DNA Mini Kit following the Tissue_200 protocol for automated isolation in the QIAsymphony (QIAGEN, Hilden, Germany). DNA from leukocytes of peripheral blood was isolated using the QIAmp DNA Blood Mini Kit (QIAGEN,

### Table 3.

| Variable                  | Category | ESCC (n = 46) | NFE2L2 gene (n = 7) | NOTCH1 gene (n = 41) |
|---------------------------|----------|--------------|---------------------|----------------------|
| Age (years)               | Mean     | 59% 57% 0.597| 65% 58% 0.205       | 62% 59% 0.495        |
|                           | Min - Max| 39 - 75% 41 - 77% | 52 - 72% 39 - 77% | 52 - 75% 39 - 77%  |
| Gender                    | Female   | 6% 15.4% 1% 14.3% | 1% 20.0% 6% 14.6% | 0% 0.0% 7% 17.1%    |
|                           | Male     | 33% 84.6% 6% 85.7% | 4% 80.0% 35% 85.4% | 5% 100% 34% 82.9%   |
|                           | Missing  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%   |
| Alcohol consumption       | No       | 6% 15.8% 2% 28.6% | 1% 20.0% 7% 17.5% | 1% 20.0% 7% 17.5%   |
|                           | Yes      | 32% 84.2% 5% 71.4% | 4% 80.0% 33% 82.5% | 4% 80.0% 33% 82.2%  |
|                           | Missing  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%   |
| Tobacco consumption       | No       | 7% 18.4% 2% 28.6% | 2% 40.0% 7% 17.5% | 1% 20.0% 8% 20.0%   |
|                           | Yes      | 31% 81.6% 5% 71.4% | 4% 80.0% 33% 82.5% | 4% 80.0% 32% 80.0%  |
|                           | Missing  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%   |
| Tobacco and alcohol in combination | No | 11% 28.9% 2% 28.6% | 2% 40.0% 11% 27.5% | 0.617 1% 20.0% 12% 30.0% |
|                           | Yes      | 27% 71.1% 5% 71.4% | 3% 60.0% 29% 72.5% | 4% 80.0% 32% 70.0%  |
|                           | Missing  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%   |
| Tumor differentiation     | Little   | 13% 35.1% 1% 14.3% | 2% 40.0% 12% 30.8% | 1.000 0% 0.0% 14% 35.0% |
|                           | Moderate | 20% 54.1% 6% 85.7% | 3% 60.0% 23% 59.0% | 3% 75.0% 23% 57.5%  |
|                           | Well     | 4% 10.8% 0% 0.0% | 0% 0.0% 4% 10.3% | 1% 25.0% 3% 7.5%    |
|                           | Missing  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%  | 0% 0% 0% 1% 0% 0%   |
| TNM Staging*              | I e II   | 5% 14.3% 0% 0.0% | 1% 25.0% 4% 10.8% | 0.418 2% 40.0% 3% 8.3% |
|                           | III e IV | 30% 85.7% 6% 100.0% | 3% 75.0% 33% 89.2% | 3% 60.0% 33% 91.7%  |
|                           | Missing  | 5% 0% 0% 5% 0% 0%  | 5% 0% 0% 5% 0% 0%  | 5% 0% 0% 5% 0% 0%   |

* TNM 7 edition staging. n number of cases. ESCC esophageal squamous cell carcinoma. MUT mutated. WT wild type.
### Table 4. Association between patients' clinicopathological features with the classification of TP53 mutations.

| Variable                          | Category                  | EAp53 status<sup>a,b</sup> | Disruptive/Non-disruptive classification<sup>c</sup> | p-value |
|----------------------------------|---------------------------|-----------------------------|-------------------------------------------------------|---------|
|                                  | High                      | Low                         | p-value                                               |         |
|                                  | n  | %   | n  | %   | n  | %   | n  | %   | p-value |
| Age (years)                      | Mean (SD)                 | 58 (10.35)                  | 64 (7.38)                                             | 0.196   | 60 (8.99) | 60 (10.08) | 0.660   |
|                                  | Min–Max                   | 39–73                       | 53–75                                                 |         | 44–75 | 39–73 |         |         |
|                                  | Missing                   | 0                           | 0                                                     |         |        |       |        |         |
| Gender                           | Female                    | 4                           | 22.2                                                  | 0.268   | 0     | 66.4 | 2     | 0.667   |
|                                  | Male                      | 14                          | 77.8                                                  | 100.0   | 17    | 51.5 | 16    | 48.5    |
|                                  | Missing                   | 0                           | 0                                                     |         |        |       |        |         |
| Alcohol consumption              | No                        | 3                           | 17.6                                                  | 11.1    | 1.000 | 3    | 42.9 | 4      | 0.682   |
|                                  | Yes                       | 14                          | 82.4                                                  | 88.9    |         | 18   | 56.3 | 14     | 43.8    |
|                                  | Missing                   | 1                           | 0                                                     |         |        |       |        |         |
| Tobacco consumption              | No                        | 3                           | 17.6                                                  | 11.1    | 1.000 | 5    | 62.5 | 3      | 0.702   |
|                                  | Yes                       | 14                          | 82.4                                                  | 88.9    |         | 16   | 51.6 | 15     | 48.4    |
|                                  | Missing                   | 1                           | 0                                                     |         |        |       |        |         |
| Tobacco and alcohol in combination | No                      | 5                           | 29.4                                                  | 11.1    | 0.380 | 7    | 58.3 | 3      | 0.742   |
|                                  | Yes                       | 12                          | 70.6                                                  | 88.9    |         | 14   | 51.9 | 13     | 48.1    |
|                                  | Missing                   | 1                           | 0                                                     |         |        |       |        |         |
| Tumor differentiation            | Little                    | 4                           | 22.2                                                  | 57.1    | 0.251 | 7    | 53.8 | 6      | 1.000   |
|                                  | Moderate                  | 12                          | 66.7                                                  | 28.6    |         | 11   | 55.0 | 9      | 45.0    |
|                                  | Well                      | 2                           | 11.1                                                  | 14.3    |         | 2    | 50.0 | 2      | 50.0    |
|                                  | Missing                   | 1                           | 0                                                     |         |        |       |        |         |
| TNM Staging*                     | I e II                    | 3                           | 18.8                                                  | 25.0    | 1.000 | 1    | 25.0 | 3      | 0.312   |
|                                  | III e IV                  | 13                          | 81.3                                                  | 75.0    |         | 18   | 53.1 | 13     | 14.9    |

<sup>a</sup> TNM 7 edition staging; n – number of cases. <sup>b</sup> http://mammoth.bcm.tmc.edu/cgi-bin/panos/EAp53.cgi. <sup>c</sup>Neskey et al. Evolutionary Action Score of TP53 Identifies High-Risk Mutations Associated with Decreased Survival and Increased Distant Metastases in Head and Neck Cancer, Cancer Res. 2015 Apr 1;75(7):1527–36.

### Table 5. Estimation of Global Survival by the Kaplan–Meier method considering patients' time follow-up and mutation status and multivariable survival analysis (Cox regression model).

| Variable                          | Category                  | Overall survival 95% confidence interval | 6 months | 1 year | 2 years | 5 years | p-value |
|----------------------------------|---------------------------|------------------------------------------|----------|--------|---------|---------|---------|
| Global                           |                           | Mean (months)                            | 9.40     | 6.101 to 12.715 | 65.5 | 38.8 | 21.9 | 7.3 | 0.373 |
| TP53                             | MUT                       | 9.40                                     | 6.282     | 12.534 | 62.5 | 36.9 | 19.9 | 5.0 | 0.822 |
|                                  | WT                        | 8.65                                     | 0.000 to 24.681 | 83.3 | 50.0 | 33.3 | 16.7 | 0.627 |
| NFE2L2                           | MUT                       | 8.71                                     | 8.364 to 9.070 | 80.0 | 20.0 | 20.0 | 0.0 | 0.0   |
|                                  | WT                        | 9.40                                     | 5.555 to 13.261 | 63.3 | 41.6 | 22.2 | 8.4 | 0.0   |
| NOTCH1                           | MUT                       | 3.42                                     | 2.220 to 4.622 | 40.0 | 40.0 | 40.0 | 0.0 | 0.589 |
|                                  | WT                        | 9.40                                     | 5.570 to 13.246 | 68.7 | 38.5 | 19.3 | 6.4 | 0.0   |
| EAp53 status<sup>a,b</sup>       | High                      | 11.28                                    | 5.023 to 17.543 | 64.7 | 51.8 | 12.9 | 12.9 | 0.0   |
|                                  | Low                       | 8.71                                     | 0.857 to 16.577 | 53.3 | 26.7 | 26.7 | 0.0 | 0.497 |
| TP53 disruptive/non-disruptive mutations<sup>c</sup> | Disruptive             | 9.40                                     | 5.169 to 13.647 | 86.7 | 41.0 | 19.2 | 6.4 | 0.0   |
|                                  | Non-disruptive            | 9.93                                     | 1.320 to 18.548 | 59.3 | 35.7 | 11.9 | 6.0 | 0.0   |

<sup>a</sup>TP53 mutations and survival in squamous-cell carcinoma of the head and neck. N Engl J Med 357, 2552–2561, https://doi.org/10.1056/NEJMoaa073770 (2007).
Figure 3. Kaplan–Meier curve for assessing the estimated overall survival probability of follow-up time among esophageal squamous cell carcinoma patients concerning mutation profile. A – TP53 gene; B – NOTCH1 gene; C – NFE2L2 gene; D – Eap53 score; E – Disruptive and non-disruptive classification of TP53 mutations.
DNA quantity and quality were assessed by Qubit (Thermo Scientific, Wilmington, DE, USA).

The mutation profile of a commercial panel of 150 cancer-related genes was conducted at Mendelics Genetics company (São Paulo, SP, Brazil, https://www.mendelics.com/oncologia/) as previously reported. The panel analyzed all coding sequence of the following genes:

- ABL1, AKT1, AKT2, AKT3, ALK, APC, AR, ARAF, ATM, AURKA, AURKB, AXIN1, AXL, BAP1, BARD1, BCL2, BCL2L1, BCL2L2, BLM, BRAF, BRCA1, BRCA2, BRD2, BRD3, BRD4, BRTD, BRIP1, BTK, CBFB, CCND1, CCND2, CCND3, CCNE1, CD22, CD274, CD79A, CD79B, CDH1, CDK12, CDK4, CDKN1A, CDKN2A, CHEK1, CHEK2, CREBBP, CRKL, DDR2, DDX3X, EGFR, EPHA7, ERBB2, ERBB3, ERBB4, ESR1, EZH2, FAM175A, FAS, FBXW7, FGFR1, FGFR2, FGFR3, FGFR4, FHT, FLT1, FLT3, FLT4, FRB2, GATA2, GNA11, GNAQ, HDAC1, HDAC4, HDAC7, HGF, HRAS, IDH1, IDH2, IGFR1, JAK1, JAK2, JAK3, KDR, KIT, KRAS, MAP2K1, MAP2K2, MAP3K1, MAPK1, MCL1, MDM2, MET, MLH1, MPL, MRE11A, MS4A1, MSH2, MSH6, MTO, MTHYR, MYC, MYD88, NBN, NF1, NF2, NFKBIA, NOTCH1, NOTCH2, NOTCH3, NRAS, NT5C2, NTRK1, PALB2, PDGFR, PDGFRB, PDK1, PIK3CA, PIK3CB, PIK3CG, PMS2, PTCH1, PTEN, PTPN1, RAD51, RAF1, RANBP2, RASA, RB1, RET, RICTOR, ROS1, RRM1, RUNX1, SDHB, SMO, SOX2, SRC, STAT3, STAT5B, STAT5A, TERC, TERT, TGFB2, TP53, TSC1, TSC2, VEGFA, WT1, and XPO1.

For sequencing, paired tumor and blood DNA libraries were prepared using Nextera Rapid Capture Custom Enrichment kit (Illumina, San Diego, CA, USA). Qubit Fluorometer quantified libraries, and Agilent 2100 Bioanalyzer evaluated their quality. The cluster generation and sequencing were performed in Illumina HiSeq 4000 following the manufacturer’s instructions. Paired-end reads from Illumina sequencing were obtained by script bcl2fastq (v. 2.17.1.14), and data pre-processing was performed following the recommended best practices, i.e., alignment against the human genome reference build GRCh37 using Burrows-Wheeler Aligner (BWA, version 0.7.13), duplicates were marked, and a further base quality score recalibration step was applied.

The VarScan2 algorithm called the somatic variants. The variants with artifacts due to indel reads at their position, or less than 10% or more than 90% of variant supporting reads on one strand, were removed. The

Table 6. Comparison of mutation frequencies in esophageal squamous cell carcinoma patients. cbioPortal for Cancer Genomics [http://cbioportal.org]; ICGC International Cancer Genome Consortium [https://dcc.icgc.org/]; n number of cases, ESCC esophageal squamous cell carcinoma. Cancer Genome Atlas Research et al. Integrated genomic characterization of oesophageal carcinoma. Nature 541, 169–175, https://doi.org/10.1038/nature20805 (2017). Sawada et al. Genomic Landscape of Esophageal Squamous Cell Carcinoma in a Japanese Population. Gastroenterology 150, 1171–1182, https://doi.org/10.1053/j.gastro.2016.01.035 (2016).
variants were further filtered to remove those with fewer than ten reads covering the variant and less than 5% of the variant allele frequency.

A second algorithm was applied using the somatic SNV and indell caller MuTect2 from GATK. MuTect2 combines the somatic genotyping engine of the original MuTect with the assembly-based machinery of HaplotypeCaller provided by GATK, detecting somatic mutations using a Bayesian classifier approach. The variants were detected by comparing the likelihood of the site to be sequencing noise and filtered by the alterations of the normal paired control (blood) from the same patient, by a pool of normals of 291 local samples using the same NGS technology and according to the allelic frequencies provided by the Genome Aggregation Database (gnomAD) datasets to reduce miscalled germline calls. In this study, only variants found by both variant callers (VarScan2 and MuTect2) were considered. Finally, the annotation of variants was done using Variant Effect Prediction.

To identify driver mutations in tumors, we applied the Cancer Genome Interpreter-CGI. Briefly, CGI annotates potential driver mutations detected in tumors by identifying known tumorigenic variants and classifying variants of unknown significance via OncodriveMUT. After the CGI classification, we maintained variants classified as tumorigenic and variants predicted as Tier1 or Tier2 for tumorigenesis. Variants that were not classified as cancer driver mutation or not predicted as Tier1 or Tier2 as driver by the OncodriveMut algorithm were excluded. Therefore, mutations identified as polymorphism (high allele frequency) or predicted as neutral or passenger for oncogenesis and found in DNA sequence outside coding regions were excluded.

Classification of TP53 mutations. To assess the impact of TP53 missense mutations, we used a risk-system score based on the Evolutionary Trace (ET) method proposed and validated in head and neck tumors by Neskey et al., called evolutionary action score of p53 (EAp53) (http://mammoth.bcm.tmc.edu/cgi-bin/panos/EAp53.cgi). In this system, mutations are scored from 0 to 100, with higher scores representing more deleterious mutations. A threshold score of 75 is used to classify variants as low-risk (EAp score < 75) or as high-risk (EAp score > 75). Mutations classified as high risk was associated with a poor prognosis, decreased survival, and increased development of distant metastases in head and neck tumors.

Additionally, we evaluated a second classification proposed for head and neck tumors by Poeta et al. and validated in lung cancer by Molina et al. According to this system, TP53 mutations are divided into “disruptive” and “non-disruptive”. Disruptive mutations are stop-codon all over the coding region and missense mutations within the L2 and L3 sites, codons 163–195 and 236–251 with an amino acid polarity shift. Non-disruptive mutations are missense mutations within the L2 and L3 sites and do not change in polarity between the amino acids.

Statistical analysis. Characterization of the study population was analyzed through frequency tables for qualitative variables and measures of central tendency and dispersion (mean, standard deviation, minimum, and maximum) for the quantitative variables. Regarding the clinicopathological association analyzes with profile mutation (TP53, NF2L2, and NOTCH1) and classification of TP53 mutations (EAp53 score and disruptive and non-disruptive), we used the Mann–Whitney test for age, and for other categorical variables, the Chi-square test.

The level of significance adopted was 5% (p < 0.05). Statistical analyses were performed using the SPSS software v.21.0 (SPSS, Chicago, IL).

Data availability
Data that support the findings are available upon reasonable request and with the permission of Dr. Rui Manuel Reis.

Received: 24 April 2021; Accepted: 8 October 2021
Published online: 18 October 2021

References
1. Ferlay, J. et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int. J. Cancer 144, 1941–1953. https://doi.org/10.1002/ijc.31937 (2019).
2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424. https://doi.org/10.3322/caac.21492 (2018).
3. INCA. Estimativa 2020: Incidência de Câncer no Brasil, http://www.inca.gov.br/estimativa/2016/ (2020).
4. Wu, S. G. et al. Surgery combined with radiotherapy improved survival in metastatic esophageal cancer in a surveillance epidemiology and end results population-based study. Sci. Rep. 6, 28280. https://doi.org/10.1038/srep28280 (2016).
5. Abnet, C. C., Arnold, M. & Wei, W. Q. Epidemiology of esophageal squamous cell carcinoma. Gastroenterology 154, 360–373. https://doi.org/10.1053/j.gastro.2017.08.023 (2018).
6. Mahboubi, E. et al. Oesophageal cancer studies in the Caspian littoral of Iran: the Caspian cancer registry. Br. J. Cancer 28, 197–214. https://doi.org/10.1038/bjc.1973.138 (1973).
7. Hull, R. et al. A multinational review: Oesophageal cancer in low to middle-income countries. Oncol. Lett. 20, 42. https://doi.org/10.3892/ol.2020.11902 (2020).
8. Tirumani, H. et al. Esophageal Carcinoma: Current concepts in the role of imaging in staging and management. Can. Assoc. Radiol. J. 66, 130–139. https://doi.org/10.1016/j.carj.2014.08.006 (2015).
9. Islami, F. et al. High-temperature beverages and foods and esophageal cancer risk: A systematic review. Int. J. Cancer 125, 491–524. https://doi.org/10.1002/ijc.24445 (2009).
10. Prabhu, A., Obi, K. O. & Rubenstein, J. H. The synergistic effects of alcohol and tobacco consumption on the risk of esophageal squamous cell carcinoma: a meta-analysis. Am. J. Gastroenterol. 109, 822–827. https://doi.org/10.1038/ajg.2014.71 (2014).
11. Cheng, C. et al. Whole-genome sequencing reveals diverse models of structural variations in esophageal squamous cell Carcinoma. Am. J. Hum. Genet. 98, 236–274. https://doi.org/10.1016/j.ajhg.2015.12.013 (2016).
12. Qin, H. D. et al. Genomic characterization of esophageal squamous cell carcinoma reveals critical genes underlying tumorigenesis and poor prognosis. *Am. J. Hum. Genet.* **98**, 709–727. https://doi.org/10.1016/j.ajhg.2016.02.021 (2016).

13. Sawada, G. et al. Genomic landscape of esophageal squamous cell carcinoma in a Japanese population. *Gastroenterology* **150**, 1171–1182. https://doi.org/10.1053/j.gastro.2016.01.035 (2016).

14. Song, Y. et al. Identification of genomic alterations in esophageal squamous cell cancer. *Nature* **509**, 91–95. https://doi.org/10.1038/nature13176 (2014).

15. Lin, D. C. et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat. Genet.* **46**, 467–473. https://doi.org/10.1038/ng.2935 (2014).

16. Gao, Y. B. et al. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet* **46**, 1097–1102. https://doi.org/10.1038/ng.3076 (2014).

17. Cancer Genome Atlas Research, N. et al. Integrated genomic characterization of esophageal carcinoma. *Nature* **541**, 169–175. https://doi.org/10.1038/nature20805 (2017).

18. Lacerda, C. F. et al. Mutational profile of TP53 in esophageal squamous cell carcinoma associated with chagasic megaesophagus. *Dis. Esophagus* **30**, 1–9. https://doi.org/10.1093/dote/dow040 (2017).

19. Neskey, D. M. et al. Evolutionary action score of TP53 identifies high-risk mutations associated with decreased survival and increased distant metastases in head and neck cancer. *Cancer Res.* **75**, 1527–1536. https://doi.org/10.1158/0008-5472.CAN-14-2735 (2015).

20. Poeta, M. L. et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* **357**, 2552–2561. https://doi.org/10.1056/NEJMoa073770 (2007).

21. Molina-Vila, M. A. et al. Nondisruptive p53 mutations are associated with shorter survival in patients with advanced non-small cell lung cancer. *Clin. Cancer Res.* **20**, 4647–4659. https://doi.org/10.1158/1078-0432.CCR-13-2391 (2014).

22. Chen, X. X. et al. Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by multi-region whole-exome sequencing. *Nat. Commun.* **8**, 524. https://doi.org/10.1038/s41467-017-00650-0 (2017).

23. Deng, J. et al. Comparative genomic analysis of esophageal squamous cell carcinoma between Asian and Caucasian patient populations. *Nat. Commun.* **8**, 1533. https://doi.org/10.1038/s41467-017-01730-x (2017).

24. Putz, A. et al. TP53 mutation pattern of esophageal squamous cell carcinomas in a high risk area (Southern Brazil): Role of lifestyle factors. *Int. J. Cancer* **99**, 99–105. https://doi.org/10.1002/ijc.10128 (2002).

25. Hu, N. et al. Genomic landscape of somatic alterations in esophageal squamous cell carcinoma and gastric cancer. *Cancer Res.* **76**, 1714–1723. https://doi.org/10.1158/0008-5472.CAN-15-0338 (2016).

26. Rossini, A. et al. TP53 mutation profile of esophageal squamous cell carcinomas of patients from Southeastern Brazil. *Mutat. Res.* **696**, 10–15. https://doi.org/10.1016/j.mrgentox.2009.11.005 (2010).

27. Kaga, S. et al. Cellular senescence checkpoint function determines differential Notch1-dependent oncogenic and tumor-suppressor activities. *Oncogene* **34**, 2347–2359. https://doi.org/10.1038/onc.2014.169 (2015).

28. Kessler, T. W., Wakabayashi, N. & Biswas, S. Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. *Annu. Rev. Pharmacol. Toxicol.* **47**, 89–116. https://doi.org/10.1146/annurev.pharmaco.46.120604.141046 (2007).

29. Satoh, H., Moriguchi, T., Takai, J., Ebina, M. & Yamamoto, M. Nrf2 prevents initiation but accelerates progression through the Kras signaling pathway during lung carcinogenesis. *Cancer Res.* **73**, 4158–4168. https://doi.org/10.1158/0008-5472.CAN-12-4499 (2013).

30. Wang, X. J. et al. Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2. *Carcinogenesis* **29**, 1235–1243. https://doi.org/10.1093/carcin/bgn093 (2008).

31. Sporn, M. B. & Liby, K. T. Nrf2 and cancer: The good, the bad and the importance of context. *Cancer Genomics & Proteomics* **12**, 568–576. https://doi.org/10.1101/gr.129684.111 (2012).

32. Wang, L. et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am. J. Hum. Genet.* **98**, 597–611. https://doi.org/10.1016/j.ajhg.2015.02.017 (2015).

33. Du, P. et al. Comprehensive genomic analysis of Oesophageal Squamous Cell Carcinoma reveals clinical relevance. *Sci. Rep.* **7**, 15324. https://doi.org/10.1038/s41598-017-14099-5 (2017).

34. Canon, J. et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. Nature **575**, 217–223. https://doi.org/10.1038/s41586-019-1694-1 (2019).

35. Heist, L. et al. Activity and safety of Adagrasib (MRTX8499) in advanced/metastatic non-small-cell lung cancer (NSCLC) harboring KRAS G12C mutation. *Eur. J. Cancer* https://doi.org/10.1016/S0959-8049(20)31076-5 (2020).

36. Dos Santos, W. et al. Mutation profiling of cancer drivers in Brazilian colorectal cancer. *Sci. Rep.* **9**, 13687. https://doi.org/10.1038/s41598-019-49611-1 (2019).

37. Zhang, L. et al. Targeted therapy in esophageal cancer. *Expert Rev. Gastroenterol. Hepatol.* **10**, 595–604. https://doi.org/10.1586/e16.140036 (2016).

38. Zhang, L. et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am. J. Hum. Genet.* **96**, 597–611. https://doi.org/10.1016/j.ajhg.2015.02.017 (2015).

39. Du, P. et al. Comprehensive genomic analysis of Oesophageal Squamous Cell Carcinoma reveals clinical relevance. *Sci. Rep.* **7**, 15324. https://doi.org/10.1038/s41598-017-14099-5 (2017).

40. Canon, J. et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223. https://doi.org/10.1038/s41586-019-1694-1 (2019).

41. Heist, L. et al. Activity and safety of Adagrasib (MRTX8499) in advanced/metastatic non-small-cell lung cancer (NSCLC) harboring KRAS G12C mutation. *Eur. J. Cancer* https://doi.org/10.1016/S0959-8049(20)31076-5 (2020).

42. Dos Santos, W. et al. Mutation profiling of cancer drivers in Brazilian colorectal cancer. *Sci. Rep.* **9**, 13687. https://doi.org/10.1038/s41598-019-49611-1 (2019).

43. Zhang, L. et al. Targeted therapy in esophageal cancer. *Expert Rev. Gastroenterol. Hepatol.* **10**, 595–604. https://doi.org/10.1586/e16.140036 (2016).

44. Zhang, L. et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am. J. Hum. Genet.* **96**, 597–611. https://doi.org/10.1016/j.ajhg.2015.02.017 (2015).
52. Katsonis, P. & Lichtarge, O. A formal perturbation equation between genotype and phenotype determines the evolutionary action of protein-coding variations on fitness. Genome Res. 24, 2050–2058. https://doi.org/10.1101/gr.176214.114 (2014).

Acknowledgements
This study was financially supported by the Public Ministry of Labor Campinas (Research, Prevention, and Education of Occupational Cancer, Brazil). FFM was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [Grant No. 2015/20077-3 to FFM]. The authors would like to thank Barretos cancer Hospital Biobank for helping in frozen tissue and blood processing. We also would like to thank Dr. Pedro Ferreira from I3S, Porto, Portugal, for the study’s initial discussions.

Author contributions
F.F.M. performed experiments, analyzed the data, and wrote the draft of the manuscript; W.S. and A.F.E. performed experiments, analyzed the data. A.C.C. contributed to the interpretation of the results. P.A.P., D.B., D.R.W., C.S., D.P.G., and A.L.F. edited and revised the manuscript critically for important intellectual content. R.M.R. conceived the original idea, supervised the project, and wrote the manuscript. All authors critically reviewed and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00208-7.

Correspondence and requests for materials should be addressed to R.M.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021