Mutations, Differential Gene Expression, and Chimeric Transcripts in Esophageal Squamous Cell Carcinoma Show High Heterogeneity

Paulo Thiago de Souza-Santos*, Sheila Coelho Soares Lima*, Pedro Nicolau-Neto*, Mariana Boroni†, Nathalia Meireles Da Costa‡, Lilian Brewer‡, Albert Nobre Menezes¶, Carolina Furtado§, Miguel Angelo Martins Moreira§, Hector N. Seuanez§, Tatiana de Almeida Simão‡ and Luis Felipe Ribeiro Pinto*,

*Molecular Carcinogenesis Program, Instituto Nacional de Câncer–INCA, Rua Andre Cavalcanti, 37–6° andar, Centro, Rio de Janeiro, RJ, Brasil, 20231-050; †Bioinformatics and Computational Biology Laboratory, Instituto Nacional de Câncer–INCA, Rua Andre Cavalcanti, 37–1° andar, Centro, Rio de Janeiro, RJ, Brasil, 20231-050; ‡Biochemistry Department, Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Boulevard 28 de Setembro, 77-Maracanã, Rio de Janeiro, RJ, Brasil, 20551-030; §Genetics Program, Instituto Nacional de Câncer–INCA, Rua Andre Cavalcanti, 37–4° andar, Centro, Rio de Janeiro, RJ, Brasil, 20231-050; ¶College of Medical and Dental Sciences, University of Birmingham, Vicent Drive, Edgbaston, Birmingham, B15 2TT, UK

Abstract

Esophageal squamous cell carcinoma (ESCC) is a frequent and lethal neoplasia. As recent advances in targeted therapy have not improved ESCC prognosis, characterization of molecular alterations associated to this tumor is of foremost relevance. In this study, we analyze, for the first time, the complete genomic profile of ESCC by RNA-seq. TP53 was the most frequently mutated gene in the investigation and validation sets (78.6% and 67.4%, respectively). Differential expression analysis between tumor and nontumor adjacent mucosa showed 6698 differentially expressed genes, most of which were overexpressed (74%). Enrichment analysis identified overrepresentation of Wnt pathway, with overexpressed activators and underexpressed inactivators, suggesting activation of canonical and noncanonical Wnt signaling pathways. Higher WNT7B expression was associated with poor prognosis. Twenty-one gene fusions were identified in 50% of tumors, none of which involving the same genes in different patients; 71% of fusions involved syntenic genes. Comparisons with TCGA data showed co-amplification of seven gene pairs involved in fusions in the present study (~33%), suggesting that these rearrangements might have been driven by chromoanagenesis. In conclusion, genomic alterations in ESCC are highly heterogeneous, impacting negatively in target therapy development.

Introduction

Esophageal carcinoma (EC) is a highly frequent and lethal tumor, representing the eighth most common and the sixth most fatal cancer worldwide [1]. Esophageal squamous cell carcinoma (ESCC) is the main histopathology subtype, accounting to approximately 80% of all EC cases, mainly in developing countries, such as Brazil [2]. ESCC patients show a poor prognosis, with a 5-year survival rate...
were filtered by quality and unique mapping features with Samtools intersection counting. As cutoff, only genes with ≥10 read counts were considered for further analyses. Differential expression analysis was performed with DESeq2 package 1.2.5 [12]. Genes were considered to be differentially expressed with adjusted P value < .001 and |log fold-change| ≥ 1. Principal component analysis (PCA) was estimated, and heatmaps were constructed with regularized-logarithmic transformation (rlog) provided by the DESeq2 package.

### Functional Annotation

Enrichment analysis was carried out with R package KEGG.db [13]. Following application of Fisher exact test, only pathways with $P < 0.1$ adjusted by Benjamin-Hochberg’s procedure were considered to be enriched.

### qPCR Validation

Four genes related to the WNT-signaling pathway (Supplementary Table 1) were selected for validation by quantitative PCR (qPCR) as described previously [14].

### Variant Calling

RNA-seq reads were initially mapped to the reference genome and to all known splice junctions using STAR [15]. Uniquely mapped reads were subsequently used for calling the initial set of candidate variants with Genome Analyses Toolkit (GATK) following the best practices recommended by the Broad Institute. Subsequently, these candidates were subjected to several filtering procedures, including removal of variations present in the nontumor adjacent mucosa; InDel alterations and A:T>G:C conversions were excluded in view of their high probability of representing RNA editing products [16]. We used ANNOVAR [17] for annotating variants based on gene models from GENCODE, RefSeq, Ensembl, and the UCSC Genome Browser. All RNA-seq variants present in 1000 genomes were defined as SNPs and excluded from the analyses.

### Validation of TP53 Mutations

Genomic DNA was isolated using the DNeasy Blood and Tissue kit (Qiagen) as recommended by the manufacturer. TP53 was amplified using three primer pairs (Supplementary Table 1), 50 ng of DNA and Taq Platinum (Invitrogen), and amplified products purified with Purelink genomic DNA purification kit (Invitrogen). Subsequently, DNA libraries were constructed with 1 ng of PCR product and Nextera XT DNA sample preparation (Illumina) following the manufacturer’s instructions and sequenced in a HiSeq 2500 platform. Reads were submitted to the above described quality control for cDNA reads. Good quality reads were mapped to the human genome with BWA aligner [18].

### Fusion Analysis

The FusionMap package [19] was used with the following parameters: Sample.SeedCount > 3, SplicePatternClass = “CanonicalPattern[Major]” or “CanonicalPattern[Minor]”, Filter = empty, FrameShiftClass = “InFrame” and OnExonBoundary = “Both”.

### Comparisons with TCGA data

TCGA data were used to compare our findings with other populations and datasets. The cBioportal for cancer genomics [20,21] was used for selecting and analyzing ESCC from the TCGA provisional dataset.

### Further Statistical Analyses

Differential expression of selected genes between ESCC and paired nontumor mucosa was evaluated with paired $t$ test or Wilcoxon test.
and chi-square. Survival analyses were carried out using clinical and pathology data were analyzed with underexpressed in our dataset (Supplementary Table 2).

mucosa, while overexpressed in ESCC with respect to the nontumor adjacent esophageal mucosa but a spread distribution among ESCC samples revealed 6698 differentially expressed genes (DEG), 4966 of which (74%) were overexpressed and 1732 (26%) underexpressed in tumors with respect to the nontumor adjacent mucosa (Figure 1C). The Wnt signaling pathway was also found to be deregulated in the TCGA dataset, although this finding was not analyzed in the TCGA report [7]. Reanalysis of TCGA data revealed that, among the Wnt-related DEG herein analyzed, NDK2 and LRP5 were mutated or amplified (23% and 21%, respectively), while SFRP5, WNT7B, CDC25G, RHOC, PAX2, and SOST did not show alterations (Supplementary Figure 2).

The expression of genes selected for validation was subsequently analyzed in association to age and clinical and pathology variables (Table 2). Low MAPK10 expression was associated with local/distant metastases (P = .046), while associations with age, tumor stage, or lymph node metastases were not observed for any other gene. Interestingly, patients whose tumors showed higher WNT7B expression showed a decreased overall survival (median = 7.73 months) when compared with patients with lower WNT7B expression in tumors (median = 17.17 months) (Figure 1D). Multivariate analysis further revealed that the impact of WNT7B expression on the prognosis of ESCC patients was independent of age and tumor stage (HR = 5.5 (1.7-18.0), P = .005).

Mutational ESCC Landscape

RNA-seq data from IS revealed a median of 65 point mutations per tumor (33-199) following removal of AT>GC conversions (Figure 2A). GC>AT was the most common conversion (57.2%), followed by GC>CG (17.7%), GC>TA (12.9%), AT>TA (7.5%), and AT>CG (4.7%) (Figure 2B). Among GC>AT, 58.6% occurred in a CpG context. TP53 was the most commonly mutated gene, in 78.6% of tumors (Figure 2C), while mutations were not detected in the nontumor adjacent mucosa. Other frequently mutated genes were LOC389831 (42.9%), PIK4A (21.4%), MSI1L (21.4%), HERC2 (21.4%), and NBP9 (21.4%). Comparisons between our findings and the TCGA dataset revealed that, except for TP53, mutations occurred in different genes. In our tumor samples, expression analyses of the 10 most frequently TCGA mutated genes (Supplementary Table 2) showed that 4 of them (CSMD3, MUC16, DNAH5 and PHD1L1) were overexpressed, while NFE2L2 was underexpressed (adjusted P value < .05).

TP53 was selected for further analysis by DNA sequencing in the validation set (VS) where its mutation frequency was slightly lower (67.4%) than in IS, with single mutations representing the most frequent class in both cases (Figure 3A). TP53 analysis identified four tumors with two mutations, and three and six mutations in single

Based on our DEG dataset, pathway enrichment analysis using the KEGG database retrieved a list of 71 overrepresented pathways (adjusted P value < .1), among which the “Wnt signaling pathway” showed a total of 54 DEG of the 150 genes included in this pathway. Their expression patterns in tumors and nontumor adjacent mucosa are shown in Figure 1A. Most genes (87%) were found to be overexpressed in tumors, with a median log fold-change of 1.75 (1.08-7.55). Most underexpressed genes in tumors corresponded to pathway inhibitors, like CTTNB1P1 and DKK1, indicating reinforcement of Wnt pathway activation. Additionally, we found that, in tumors, activation was not restricted to the canonical pathway represented by the WNT7B ligand and its downstream CCND2 target leading to cell cycle progression, but also involved the noncanonical pathway activated by WNT16 leading to MAPK10 activation and apoptosis (Figure 1B). Validation of these findings confirmed overexpression of WNT7B (P < .001), CCND2 (P = .041), WNT16 (P = .037), and MAPK10 (P < .001) in tumors with respect to the nontumor adjacent mucosa (Figure 1C). The Wnt signaling pathway was also found to be deregulated in the TCGA dataset, although this finding was not analyzed in the TCGA report [7]. Reanalysis of TCGA data revealed that, among the Wnt-related DEG herein analyzed, NDK2 and LRP5 were mutated or amplified (23% and 21%, respectively), while SFRP5, WNT7B, CDC25G, RHOC, PAX2, and SOST did not show alterations (Supplementary Figure 2).

The expression of genes selected for validation was subsequently analyzed in association to age and clinical and pathology variables (Table 2). Low MAPK10 expression was associated with local/distant metastases (P = .046), while associations with age, tumor stage, or lymph node metastases were not observed for any other gene. Interestingly, patients whose tumors showed higher WNT7B expression showed a decreased overall survival (median = 7.73 months) when compared with patients with lower WNT7B expression in tumors (median = 17.17 months) (Figure 1D). Multivariate analysis further revealed that the impact of WNT7B expression on the prognosis of ESCC patients was independent of age and tumor stage (HR = 5.5 (1.7-18.0), P = .005).

Mutational ESCC Landscape

RNA-seq data from IS revealed a median of 65 point mutations per tumor (33-199) following removal of AT>GC conversions (Figure 2A). GC>AT was the most common conversion (57.2%), followed by GC>CG (17.7%), GC>TA (12.9%), AT>TA (7.5%), and AT>CG (4.7%) (Figure 2B). Among GC>AT, 58.6% occurred in a CpG context. TP53 was the most commonly mutated gene, in 78.6% of tumors (Figure 2C), while mutations were not detected in the nontumor adjacent mucosa. Other frequently mutated genes were LOC389831 (42.9%), PIK4A (21.4%), MSI1L (21.4%), HERC2 (21.4%), and NBP9 (21.4%). Comparisons between our findings and the TCGA dataset revealed that, except for TP53, mutations occurred in different genes. In our tumor samples, expression analyses of the 10 most frequently TCGA mutated genes (Supplementary Table 2) showed that 4 of them (CSMD3, MUC16, DNAH5 and PHD1L1) were overexpressed, while NFE2L2 was underexpressed (adjusted P value < .05).

TP53 was selected for further analysis by DNA sequencing in the validation set (VS) where its mutation frequency was slightly lower (67.4%) than in IS, with single mutations representing the most frequent class in both cases (Figure 3A). TP53 analysis identified four tumors with two mutations, and three and six mutations in single

Association between gene expression and the TP53 mutation status with clinical and pathology data were analyzed with t-test, Mann-Whitney test, and chi-square. Survival analyses were carried out using “survival” package for R, and hazard ratios (HRs) were adjusted by Cox regression. All analyses were performed in the R environment.

Results

Clinical and Pathology Data

The 55 patients included in this study showed a median age of 59 years (39-79) (Table 1). Most patients were male (74.5%) and showed a median overall survival of 8 months (1.2-42.7). Tumors were moderately differentiated (70.9%), most frequently located in the middle third of the esophagus (78.2%), and diagnosed at late stages (80%).

High-Throughput Evaluation of Gene Expression in ESCC

Global gene expression of tumor and nontumor adjacent mucosa was analyzed in the investigation set (IS). RNA-seq provided an average of 37 million sequences for nontumor adjacent samples, 33 million (89%) of which of good quality. Comparatively, an average of 40 million sequences was generated from tumor samples, with 38.5 million (96%) of good quality. Sequences of good quality, aligned to the reference human genome, were correctly mapped, accounting for 80% and 87% of sequences from nontumor and tumor samples, respectively. Reads were mapped to the 63,677 genes of the Ensembl database, and 44,402 (70%) of them were considered for further analyses following mapping of at least 10 reads. Principal component analysis revealed a homogenous distribution among nontumor adjacent esophageal mucosa but a spread distribution among ESCC samples (Supplementary Figure 1). Comparisons between total paired samples revealed 6698 differentially expressed genes (DEG), 4966 of which (74%) were overexpressed and 1732 (26%) underexpressed in tumors. The differential expression of the 10 most commonly amplified or deleted genes in the TCGA study was analyzed. Four amplified genes (FGES, FGFI9, TP63, and TFRC) were found to be overexpressed in ESCC with respect to the nontumor adjacent mucosa, while CDKN2B, found to be deleted in TCGA, was underexpressed in our dataset (Supplementary Table 2).

| Table 1. Sample Data and Clinical and Pathology Profiles of 55 ESCC Patients Included in This Study |
|-----------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Gender                                   |                 |                 |                 |                 |                 |
| Male                                     | 41 (74.5%)     |                 |                 |                 |                 |
| Female                                   | 14 (25.4%)     |                 |                 |                 |                 |
| Age (median and range, years)            | 59 (39-79)     |                 |                 |                 |                 |
| Follow-up (median and range, months)     | 8 (2.4-17.7)   |                 |                 |                 |                 |
| Tumor central localization               |                 |                 |                 |                 |                 |
| Upper third                              | 5 (9.1%)       |                 |                 |                 |                 |
| Middle third                             | 43 (78.2%)     |                 |                 |                 |                 |
| Lower Third                              | 7 (12.7%)      |                 |                 |                 |                 |
| Differentiation                          |                 |                 |                 |                 |                 |
| Well                                     | 1 (1.8%)       |                 |                 |                 |                 |
| Moderately                               | 39 (70.9%)     |                 |                 |                 |                 |
| Poorly                                   | 15 (27.3%)     |                 |                 |                 |                 |
| Tumor stage                              |                 |                 |                 |                 |                 |
| I                                         | 2 (3.6%)       |                 |                 |                 |                 |
| II                                        | 5 (9%)         |                 |                 |                 |                 |
| III                                       | 9 (16.4%)      |                 |                 |                 |                 |
| IV                                        | 18 (32.7%)     |                 |                 |                 |                 |
| NA                                        | 21 (38.2%)     |                 |                 |                 |                 |

NA, not available.
tumors (Figure 3A). The most common conversion was AT\(\rightarrow\)GC (31.7%), not considered in IS (Figure 3B). GC\(\rightarrow\)AT was the second most frequent conversion (29.3%) in VS and the most frequent one in IS (45.5%); AT\(\rightarrow\)TA and GC\(\rightarrow\)TA were found in at least 10% of cases in both sets; InDel and GC\(\rightarrow\)CG and AT\(\rightarrow\)CG conversions showed smaller frequencies (Figure 3B). Figure 3C shows the distribution of TP53 mutations per exon in IS and VS. The vast majority of mutations was found in exons 5 to 8 (altogether comprising 82% in the IS and 90% in the VS), while mutations were also observed in exons 4 (3% in the VS only), 9 (9% the IS only), 10 (9% the IS and 3% in the VS), and 11 (3% in the VS only). In IS and VS, 64% and 86% of missense mutations were observed, respectively, while all others comprised nonsense mutations in both datasets (Supplementary Table 3). Mutations were also identified in TP53 introns in VS, always coexisting with exonic mutations. The presence of TP53 mutations was not associated with clinical and pathology data (data not shown).

**Gene Fusions Analysis in ESCC**

RNA-seq data were further used for investigating the presence of gene fusions in ESCC. Interestingly, although 21 fusions were detected, none of them was shared by different tumors, and the vast majority of them (71%) involved syntenic genes (Table 3). These fusions had not been previously described in the Fusion Cancer Database and Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer.

Five of seven patients of IS showed one or two fusions in tumors, and two patients showed a higher number of fusions. Patient 4 was shown to carry eight gene fusions, including two involving the same genes, MAP4 and SPINK8, but with different breakpoints (Table 3), while patient 14 showed six fusions. Mutations were not detected in genes involved in fusions, and fusions were not associated with clinical and pathology data (not shown). However, TCGA data of all genes herein involved in fusions showed that seven gene pairs were co-amplified, albeit with frequencies ranging from 1% to 22%, while five gene pairs showed co-deletion (in 1%-4% of samples), and one gene pair was co-amplified and co-deleted in different samples (Supplementary Figure 3). Moreover, as fusions might be a consequence of dysregulated DNA repair, mainly by nonhomologous end-joining (NHEJ), analysis of 14 NHEJ-related genes from the KEGG database showed that only ATR was overexpressed in tumor samples with fusions \((P = .034)\) (Figure 4).
This study reports, for the first time, a comprehensive analysis of the ESCC transcriptome, mutational landscape, and fusion events. It shows that the Wnt signaling pathway was frequently activated through dysregulation of gene expression. Furthermore, TP53 was the only gene with a high mutation frequency, while gene fusions were apparently random.

RNA-seq is a high-throughput procedure that provides extensive data on genomic alterations with numerous advantages over hybridization-based transcriptome analysis, like detection of rare

### Table 2. Association Between Characteristics of the ESCC Patients Included in This Study and Gene Expression Levels of the Wnt Signaling Pathway Components, Assessed by RT-qPCR in the Validation Set (n=28)

| Clinical and Pathology Data | WNT7B Expression | P Value | WNT16 Expression | P Value | CCND2 Expression | P Value | MAPK10 Expression | P Value |
|-----------------------------|------------------|---------|------------------|---------|------------------|---------|-------------------|---------|
| Age <60                     | 4.9x10^{-2}      | .722    | 5.4x10^{-3} to 1.3x10^{-5} | .722    | 1.6x10^{-3}      | .602    | 5.7x10^{-4} to 1.5x10^{-3} | .763    |
| ≥60                         | 3.7x10^{-2}      | .999    | 3.1x10^{-8} to 1.0x10^{-4} | .999    | 2.4x10^{-3} to 1.0x10^{-4} | .813    | 1.6x10^{-4} to 4.1x10^{-3} | .389    |
| Tumor stage                 |                  |         |                  |         |                  |         |                   |         |
| I/II                        | 8.0x10^{-2}      | .72     | 2.4x10^{-6} to 1.3x10^{-4} | .72     | 8.5x10^{-3}      | .183    | 9.7x10^{-4} to 1.5x10^{-3} | .389    |
| III/IV                      | 3.7x10^{-2}      | .859    | 3.1x10^{-8} to 9.2x10^{-5} | .859    | 7.1x10^{-3}      | .149    | 6.1x10^{-3} to 9.7x10^{-5} | .318    |
| Lymph node metastases       |                  |         |                  |         |                  |         |                   |         |
| No                          | 5.9x10^{-2}      | .859    | 2.8x10^{-7} to 1.3x10^{-4} | .859    | 7.1x10^{-3}      | .149    | 6.1x10^{-3} to 4.1x10^{-3} | .318    |
| Yes                         | 6.2x10^{-2}      | .859    | 1.1x10^{-7} to 9.9x10^{-4} | .859    | 5.7x10^{-3} to 5.7x10^{-3} | .149    | 1.4x10^{-4} to 1.5x10^{-3} | .318    |
| Local/distant metastases    |                  |         |                  |         |                  |         |                   |         |
| No                          | 3.7x10^{-2}      | .370    | 1.1x10^{-7} to 1.3x10^{-4} | .370    | 5.7x10^{-3} to 5.7x10^{-3} | .105    | 6.1x10^{-3} to 4.1x10^{-3} | .016    |
| Yes                         | 7.3x10^{-2}      | .370    | 3.1x10^{-8} to 9.2x10^{-5} | .370    | 3.2x10^{-3}      | .105    | 1.4x10^{-4} to 1.2x10^{-3} | .316    |
| Cox regression               |                  | .33     |                  | .33     |                  | .57     |                  | .16     |

**Figure 2.** RNA-seq mutational analysis in ESCC. Only missense, nonsense, and synonymous point mutations were considered. SNPs, InDel, and A>T to G>C conversions were removed due to presumptive association with RNA edition. (A) Graphical representation of the number of mutations per tumor sample in the investigation set. (B) Graphical representation of the mutational profiles. (C) Graphical representation of the mutation frequency of the most commonly altered genes in tumors of the investigation set. IS, investigation set.
and alternative transcripts, splice variants, mutations, and quantification of transcript expression [22]. Although mutational analysis of transcripts is restricted following removal of InDel mutations and A:T to G:C conversions largely produced by RNA editing [16], all other mutations in coding regions can be successfully analyzed, as was the case in liver and prostate cancer [23,24].

Less than 10 studies of ESCC alterations by RNA-seq have been reported in the literature, most of which of Chinese patients and with similar number of patients to those included in our study [25–31]. These studies also showed different but comparable estimates of DEG (1425 to 6150) to our findings (6698), depending on sample size, fold-change and P value cutoffs, with a predominance of overexpression. The difference in DEG among the different studies may, at least in part, be caused by the heterogeneity among tumor samples, as shown in this study through the comparison of principal component analysis between nontumor adjacent mucosa and tumor samples. Nevertheless, RNA-seq experiments with at least 12 biological replicates were found to be satisfactory for identifying differentially expressed genes for all fold changes [32]. Recently, the most comprehensive RNA-seq study on ESCC carried out by the TCGA consortium [7] included 90 tumors and 3 samples of nontumor esophageal mucosa from Asian, North American, South American, and European populations. Although this study did not compare

Figure 3. TP53 mutational analysis by RNA-seq and DNA sequencing. (A) Graphical representation of the frequency of samples without mutations and with different numbers of TP53 mutations in the investigation set (RNA-seq; n = 14) and validation set (DNA-seq; n = 41). (B) TP53 InDel and conversions in the investigation and validation sets. (C) Distribution of TP53 mutations per exon, in the investigation and validation sets. WT = wild type. # not applicable; InDel and A>T to G>C conversions were removed from analysis in the investigation set due to presumptive association with RNA edition.

Table 3. Gene Fusions (n = 21) Identified in the 14 Samples of the Investigation Set by RNA-seq

| Sample | 3' Gene | 5' Gene | 5' Breakpoint Position (chr:nt) | 3' Breakpoint Position (chr:nt) |
|--------|---------|---------|---------------------------------|---------------------------------|
| IS01   | PL53    | IKZF3   | X:114795587                    | 17:37949186                     |
|        | SLC25A43| IMPG1   | X:118540664                    | 6:76640688                      |
| IS02   | CCDC127 | AHRB    | 5:216844                       | 5:376725                        |
| IS04   | RAB3IP  | CTC2    | 12:70132811                    | 12:69985839                     |
| IS05   | HNRNPC  | CHD8    | 14:21731470                    | 14:21869672                     |
| IS06   | POLK2A  | NLGN2   | 17:7416242                     | 17:7417663                      |
| IS07   | NDUFA10 | MYEOV2  | 2:240957970                    | 2:241070505                     |
| IS08   | MAP4    | SPINK8  | 3:48130263                     | 3:48351436                      |
| IS09   | MAP4    | SPINK8  | 3:48130263                     | 3:48361108                      |
| IS10   | FLSN    | SLMAP   | 3:58134579                     | 3:57893611                      |
| IS11   | ZNE3    | TAF6    | 10:8238:00                      | 7:99711891                      |
| IS12   | SPAG9   | CA10    | 17:49197715                    | 17:49825178                     |
| IS13   | NDUFA2  | ZSWIM6  | 5:60241209                     | 5:60768508                      |
| IS14   | SGMS1   | GACRN2  | 10:52220433                    | 10:18439812                     |
|        | CTBP2   | EPAC1B  | 10:126848888                    | 17:28378136                     |
|        | LOC100133315| DEFBI3 | 11:71627120                    | 4:9452086                       |
|        | YWHAZ   | DACH1   | 8:101964157                    | 13:72256042                     |
|        | TRAF3   | CDRK12  | 14:103244012                   | 17:37665958                     |
|        | PHF2    | CHM2F102| 9:96429522                     | 9:98718196                      |
|        | IS15    | CTTNBP2NL| ST7L                           | 1:112958864                     |
|        |         | SHBP1   | SLC20A2                        | 8:41160980                      |
|        |         |         |                                 | 8:42302280                      |

Chr, chromosome; nt, nucleotide; IS, investigation set.
ESCC tumor with nontumor adjacent tissues, three different molecular subtypes (ESCC1, ESCC2, and ESCC3) were suggested. The ESCC2 subtype, more common among Eastern European and South American patients, was characterized by loss of function of KDM6A, KTM2D, PTEN and PIK3R1, CDK6 amplification and BST2 overexpression. Interestingly, in our dataset, CDK6 and BST2 overexpression was observed in tumors with respect to the nontumor adjacent mucosa (logFC of 1.86 and 2.28, respectively), while PTEN was significantly underexpressed (logFC of -0.47). These findings suggested that these alterations might play a role in ESCC development in South American patients.

Although enrichment of alterations of the Wnt pathway has been reported in ESCC studies with RNA-seq [7,26,27], this pathway had not been further explored. Our findings showed 54 DEG in the Wnt pathway, most of which showing overexpression and associated with activation of Wnt signaling. Deregulation of Wnt signaling has been found in other tumors, albeit by different mechanisms. Recently, Guinney et al. [33] proposed a new molecular classification for colorectal cancer, showing that APC mutations were present in 70% of cases, being more common (83%) in the consensus molecular subtype 2 (CMS2), followed by CMS3 (78%), CMS4 (66%), and CMS1 (40%). Conversely, breast cancer showed a more similar
Deregulation of CTNNB1, APC, and DVL1 expression was found in 21% of all invasive breast carcinomas, while this frequency was much higher (56%) in PAM50 Basal subtype [36]. Although different genes were found to be involved in ESCC and breast cancer, Wnt pathway activation was associated with poor outcome in both tumors [present data; 34-36]. In this study, WNT7B overexpression was shown to be an independent prognostic marker in ESCC (HR = 5.5). WNT7B overexpression has been reported in breast cancer at mRNA and protein levels, with immunoreactivity in tumor and myeloid cells [37]. These authors showed that WNT7B silencing in myeloid cells resulted in reduction of breast tumor mass, volume, and lung metastases in a murine model. Additionally, WNT16, a gene involved in Wnt pathway activation by a noncanonical mechanism leading to JNK activation [38], was also found to be overexpressed in our study, as well as its downstream target MAPK10, also known as JNK3. The activation of this noncanonical pathway has been shown to increase keratinocyte proliferation [38], while JNK has been shown to be necessary for UV-mediated apoptosis [39].

Mutation analysis revealed that TP53 was the most frequently mutated gene in ESCC, with approximately 70% of tumors showing at least one mutation, mainly missense (86%), and in exons 5 to 8 (90%). Most mutations resulted in a nonfunctional protein. Our RNA-seq analyses were more sensitive than previous ones based on Sanger sequencing pointing to TP53 mutations in 35% of Brazilian patients with ESCC [40]. It also confirmed TCGA and the International Cancer Genome Consortium (ICGC) studies showing TP53 as the only gene that was very frequently affected by mutations in ESCC (93% and 62%, respectively). LOC389831, a gene coding for uncharacterized isoforms, was the second most frequently mutated gene (43%) in our study, followed by PI4KA, HERC2, NBPF9, and MST1L, each with 21.4%. Phosphatidylinositol 4-Kinase Alpha (PI4KA) catalyzes the first step of the biosynthesis of phosphatidylinositol 4,5-bisphosphate. It has been found to be overexpressed in hepatocarcinoma, showing a positive and significant association between PCNA and KI67 proliferation markers and associated to a poor prognosis [41]. HECT Domain And RCC1-Like Domain-Containing Protein 2 (HERC2) facilitates the assembly of the RNF8-UBC13 complex to recruit BRCA1 to DNA damaged sites, and it also plays a role in p53 oligomerization [42]. Its overexpression has been previously associated with a favorable prognosis in non–small cell lung cancer [43]. Neuroblastoma Breakpoint Family Member 9 (NBPF9) has been found to be overexpressed in lung adenocarcinoma [44], while Macrophage Stimulating 1 Like (MST1L) codes for a serine-type endopeptidase of unknown status in tumors (www.genecards.org, GC01M016754).

Our findings were not coincident with TCGA and ICGC studies because none of these four above-mentioned genes were included among the 20 most mutated genes in these databases. A likely explanation for this discrepancy might be the exclusion of A:T>G:C conversions and InDel in our RNA-seq mutational analysis and the ESCC heterogeneity between populations since only 4 of the 10 most frequently mutated genes were shared by TCGA and ICGC data, albeit with different mutation frequencies, probably due to the different ethnic composition of patients included in these studies, viz., Chinese in ICGC and a heterogeneous set of patients in TCGA. These findings pointed to the heterogeneous mutational landscape of ESCC, without a major mutation accounting for the likely activation of a driver oncogene, thus making the development of a target-specific tyrosine kinase inhibitor difficult.

Although half of the tumors herein studied showed fusions, none of them was recurrent or involved the same genes in different patients, and most of them involved syntenic loci. Approximately 33% of the gene pairs involved in fusions were co-amplified in TCGA, like the syntenic genes CCDC127 and AHR. This pointed to a likely occurrence of chromoanagenesis and formation of micronuclei initiated by error in mitotic joining, a common event in malignancies with a highly dysregulated cell cycle like ESCC. Newly formed micronuclei frequently show a reduced nuclear import consequently to which defective response to DNA damage signaling and delayed DNA replication occur. Cells may thus enter in mitosis with micronuclei still undergoing DNA replication, resulting in chromosome fragmentation and repair of double-strand breaks by nonhomologous end joining (NHEJ), leading to deletions, translocations, and formation of double minute chromosomes. These small, circular DNA fragments might contain one or more genes, usually oncogenes, while lacking centromeres and telomerases [45], often present at many copies per cell. Double minute chromosomes have been reported in ESCC in association to amplification of two candidate oncogenes, FGFR1 and LETM2 [46]. Furthermore, the association between ATR overexpression and presence of fusions in our dataset further suggested the occurrence of chromoanagenesis in ESCC, although further studies are necessary for confirming this phenomenon.

This study showed that there was a high heterogeneity among ESCC patients, either for each type of genomic alteration or between them, and also when compared to other studies, such as TCGA. Therefore, we conclude that genomic alterations in ESCC are highly heterogeneous, impacting negatively in target therapy development for ESCC.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tranon.2018.08.002.

Acknowledgements
The authors would like to thank the National Tumor Bank from Instituto Nacional de Cancer (BNT/INCA) for sample collection and the Pathology Division (DIPAT/INCA) for diagnosis confirmation.

Funding
This work was supported by Swiss Bridge Foundation, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), and Ministério da Saúde (INCA). The sponsors had no involvement in either study conception or execution.

References
[1] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, and Bray F (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136, E359-386.
[2] Rustgi AK and El-Serag HB (2014). Esophageal carcinoma. N Engl J Med 371, 2499-2509.
[3] Cohen DJ and Ajani J (2011). An expert opinion on esophageal cancer therapy. Expert Opin Pharmacother 12, 225–239.
[4] Mauer AM, Kraut EH, Krauss SA, Ansari RH, Kasza K, Szeto L, and Vokes EE (2005). Phase II trial of oxaliplatin, leucovorin and fluorouracil in patients with advanced carcinoma of the esophagus. Ann Oncol 16, 1320–1325.
[5] Souza-Santos et al. Translational Oncology Vol. 11, No. 6, 2018
Ge H, Liu K, Juan T, Fang F, Newman M, and Hoeck W (2011). FusionMap: Translational Oncology

Ma Y, Miao Y, Peng Z, Sandgren J, De Ståhl TD, Huss M, Lennartsson L, Liu Y, Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Tong M, Chan KW, Bao JY, Wong KY, Chen JN, Kwan PS, Tang KH, Fu L, Anders S and Huber W (2010). Differential expression analysis for sequence count data. Genome Biol 11, R106.

Carlson M (2016). KEGG.db: a set of annotation maps for KEGG. R package version 3.2.3; 2016.

Nicolau-Neto P, Da Costa NM, de Souza Santos PT, Gonçaga IM, Ferreira MA, Guaraldi S, Moreira MA, Seuánez HN, Brewer L, and Bergmann A, et al (2018). FOXM1 on patient outcome through novel PIK3R3 mediated activation of PI3K signaling pathway. Oncotarget 9, 16634–16647.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Wang K, Li M, and Hakonarson H (2010). ANNOVAR: functional annotation of SNVs and INVs. Bioinformatics 26, 448–455.

Picardi E, Manzari C, Mastropasqua F, Aiello I, D’Erchia AM, and Pesole G (2015). Profiling RNA editing in human tissues: towards the inosinome atlas. Sci Rep 5, 14941.

Wang K, Li M, and Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164.

Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM; 2013 [arXiv:1303.3997v1 [q-bio.GN]].

Ge H, Liu K, Juan T, Fang F, Newman M, and Hoeck W (2011). FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. Bioinformatics 27, 1922–1928.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, and Larsson E, et al (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6, p11.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, and Larsson E, et al (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2, 401–404.

Wang Z, Gerstein M, and Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63.

Choi JH, Kim MJ, Park YK, Im JY, Kwon SM, Kim HC, Woo HG, and Wang HJ (2017). Mutations acquired by hepatocellular carcinoma recurrence give rise to an aggressive phenotype. Oncotarget 8, 22903–22916.

Ma Y, Miao Y, Peng Z, Sandgren J, De Stald TH, Huss M, Lennartsson L, Liu Y, Nistér M, and Nilsson S, et al (2016). Identification of mutations, gene expression changes and fusion transcripts by whole transcriptome RNA-seq in docetaxel resistant prostate cancer cells. Springerplus 5, 1861–1872.

Ma S, Bao JY, Kwon PS, Chan YP, Tong CM, Fu L, Zhang N, Tong AYH, Qin YR, and Tsao SW, et al (2012). Identification of PTK6, via RNA sequencing analysis, as a suppressor of esophageal squamous cell carcinoma. Gastroenterology 143, 675–686.e612.

Tong M, Chan KW, Bao JY, Wong KY, Chen JN, Kwon PS, Tang KH, Fu L, Qin YR, and Lok S, et al (2012). Rab25 is a tumor suppressor gene with antiangiogenic and anti-invasive activities in esophageal squamous cell carcinoma. Cancer Res 72, 6024–6035.

Jiang YZ, Li QH, Zhao QJ, and Lv JF (2014). Identification of a novel fusion gene (HLA-E and HLA-B) by RNA-seq analysis in esophageal squamous cell carcinoma. Asian Pac J Cancer Prev 15, 2309–2312.

Chen Y, Yin D, Li L, Deng YC, and Tian W (2015). Screening aberrant methylation profile in esophageal squamous cell carcinoma for Kazakhs in Xinjiang area of China. Mol Biol Rep 42, 457–464.

Wei G, Luo H, Sun Y, Li J, Tian L, Wu W, Liu L, Luo J, He J, and Chen R (2015). Transcriptome profiling of esophageal squamous cell carcinoma reveals a long noncoding RNA acting as a tumor suppressor. Onco Targets Ther 6, 17065–17080.

Fu JH, Wang LQ, Li T, and Ma GJ (2015). RNA-sequencing based identification of crucial genes for esophageal squamous cell carcinoma. J Cancer Res Ther 11, 420–425.

Li CQ, Huang GW, Wu YZ, Xu YJ, Li XC, Xue YJ, Zha Y, Zhao JM, Li M, and Zhang J, et al (2017). Integrative analyses of transcriptome sequencing identify novel functional IncRNAs in esophageal squamous cell carcinoma. Oncogenesis 6, e297.

Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, and Owen-Hughes T, et al (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 22, 839–851.

Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, and Angelino P, et al (2015). The consensus molecular subtypes of colorectal cancer. Nat Med 21, 1350–1356.

Geyer F, Lacroix-Triki M, Savage K, Arnedos M, Lambros MB, MacKay A, Natrajan R, and Reis-Filho JS (2011). β-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. Mod Pathol 24, 209–231.

Dey N, Barwick BG, Moreno CS, Ordonic-Kodani M, Chen Z, Oppera-Illes G, Tang W, Catzavelos C, Kerstann KF, and Sledge GW, et al (2013). Wnt signaling in triple negative breast cancer is associated with metastasis. BMC Cancer 13, 537–551.

De P, Carlson JH, Wu H, Marcus A, Leyland-Jones B, and Dey N (2016). Wnt/β-catenin pathway signals metastasis-associated tumor cell phenotypes in triple negative breast cancers. Oncotarget 7, 43124–43149.

Yeo EJ, Cassetta L, Qian BZ, Lewkowich I, Li JF, Stetler JA, Smith AN, Wiechmann LS, Wang Y, andPollard JW, et al (2014). Myeloid WNTβ mediated the angiogenic switch and metastasis in breast cancer. Cancer Res 74, 2962–2973.

Tch MT, Blaydon D, Ghali LR, Briggs V, Edmunds S, Pantazi E, Barnes MR, Leigh IM, Kelsh DP, and Philpot MP (2007). Role for WNT16β in human epidermal keratinocyte proliferation and differentiation. J Cell Sci 120, 330–339.

Tourcier C, Hes P, Yang XD, Xu J, Turner TK, Nimnuan A, Bar-Sagi D, Jones SN, Flavell RA, and Davis RJ (2000). Requirement of JNK for stress-induced activation of the cytochrome c-mediated death pathway. Science 288, 870–874.

Rossini A, de Almeida Simão T, Marques CB, Soares-Lima SC, Herster S, Rapozo DC, Andreollo NA, Ferreira MA, El-Jaick KB, and Teixeira R, et al (2010). TP53 mutation profile of esophageal squamous cell carcinomas of patients from Southeastern China. Mutat Res 696, 10–15.

Ilboudo A, Nault JC, Dubois-Pot-Schneider H, Corlu A, Zucman-Rossi J, Samson M, and Le Seyec J (2014). Overexpression of phosphatidylinositol 4-kinase type IIα is associated with undifferentiated status and poor prognosis of human hepatocellular carcinoma. BMC Cancer 14, 7–14.

Cubillos-Rojas M, Amair-Pinedo F, Peiró-Jordán R, Barrón R, Ventura F, and Ross J (2016). The E3 ubiquitin protein ligase HERC2 modulates the activity of tumor protein p53 by regulating its oligomerization. J Biol Chem 289, 14782–14795.

Bonanno L, Costa C, Majmud M, Sanchez JJ, Rodriguez I, Gimenez-Capitan A, Molina-Vila MA, Vergnenegre A, Masutti B, and Favaoreto A, et al (2016). Combinatory effect of BRCA1 and HERC2 expression on outcome in advanced non-small-cell lung cancer. BMC Cancer 16, 312–317.

Zhang Y, Wang H, Wang J, Bao L, Wang L, Hoo J, and Wang X (2015). Global analysis of chromosome 1 genes among patients with lung adenocarcinoma, squamous carcinoma, large-cell carcinoma, small-cell carcinoma, or non-cancer. Cancer Metastasis Rev 34, 249–264.

Holland AJ and Cleveland DW (2012). Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. Nat Med 18, 1630–1638.

Cheng C, Zhou Y, Li H, Xiong T, Li S, Bi Y, Kong P, Wang F, Cui H, and Li Y, et al (2016). Whole-genome sequencing reveals diverse models of structural variations in esophageal squamous cell carcinoma. Am J Hum Genet 98, 256–274.