Contents lists available at ScienceDirect

Translational Oncology

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

Comprehensive Characterization of the Mutational Landscape in Localized Anal Squamous Cell Carcinoma

Lucía Trilla-Fuertes [a], Ismael Ghanem [b], Joan Maurel [c], Laura G-Pastrián [d,e], Marta Mendiola [e,f], Cristina Peña [d], Rocío López-Vacas [g], Guillermo Prado-Vázquez [a], Elena López-Camacho [g], Andrea Zapater-Moros [g], Victoria Heredia [f,h], Miriam Cuatrecasas [i], Pilar García-Alfonso [j], Jaume Capdevila [k], Carles Conill [l], Rocío García-Carbonero [m], Karen E. Heath [n], Ricardo Ramos-Ruiz [o], Carlos Llorens [p], Ángel Campos-Barros [n], Angelo Gámez-Pozo [g], Jaime Feliu [b,f,q,*], Juan Ángel Fresno Vara [f,g,*]

a Biomedica Molecular Medicine SL, C/Faraday 7, 28049, Madrid, Spain
b Medical Oncology Department, Hospital Universitario La Paz, Paseo de la Castellana 261, 28046, Madrid, Spain
c Medical Oncology Department, Hospital Clinic of Barcelona, Translational Genomics and Targeted Therapeutics in Solid Tumors Group, IDIBAPS, University of Barcelona, Carrer de Villarroel 170, 08036, Barcelona, Spain
d Pathology Department, Hospital Universitario La Paz, Paseo de la Castellana 261, 28046, Madrid, Spain
e Molecular Pathology and Therapeutic Targets Group, Hospital Universitario La Paz-IdiPAZ, Paseo de la Castellana 261, 28046, Madrid, Spain
f Biomedical Research Networking Center on Oncology-CIBERONC, ISCIII, Av. Monforte de Lemos 5, 28029, Madrid, Spain
g Molecular Oncology & Pathology Lab, Institute of Medical and Molecular Genetics-INGEMM, Hospital Universitario La Paz-IdiPAZ, Paseo de la Castellana 261, 28046, Madrid, Spain
h Translational Oncology Lab, Hospital Universitario La Paz-IdiPAZ, Paseo de la Castellana 261, 28046, Madrid, Spain
i Pathology Department, Hospital Clínic Universitari de Barcelona, Carrer de Villarroel 170, 08036, Barcelona, Spain
j Medical Oncology Department, Hospital General Universitario Gregorio Marañón, C/ Dr Esquerdo 46, 28007, Madrid, Spain
k Medical Oncology Service, Vall Hebron University Hospital. Vall Hebron Institute of Oncology (VHIO), Passeig de la Vall d'Hebron, 119-129, 08035, Barcelona, Spain
l Radiotherapy Oncology Department, Hospital Clínic Universitari de Barcelona, Carrer de Villarroel 170, 08036, Barcelona, Spain
m Medical Oncology Department, Hospital Universitario 12 de Ocubre, Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12), UCM, CNIO, CIBERONC, Av. Córdoba s/n, 28041, Madrid, Spain
n Institute of Medical and Molecular Genetics, IdiPAZ, Hospital Universitario La Paz /& CIBERER, Unit 753, ISCIII, Paseo de la Castellana 261, 28046, Madrid, Spain
o Genomics Unit Cantoblanco, Parque Científico de Madrid, C/ Faraday 7, 28049, Madrid, Spain
p Biotechvana SL, Parque Científico de Madrid, C/ Faraday 7, 28049, Madrid, Spain
q Cátedra UAM-Amgen, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049, Madrid, Spain

ARTICLE INFO

ABSTRACT

Anal squamous cell carcinoma (ASCC) is a rare neoplasm. Chemoradiotherapy is the standard of care, with no therapeutic advances achieved over the past three decades. Thus, a deeper molecular characterization of this disease is still necessary. We analyzed 46 paraffin-embedded tumor samples from patients diagnosed with primary ASCC by exome sequencing. A bioinformatics approach focused in the identification of high-impact genetic variants, which may act as drivers of oncogenesis, was performed. The relation between genetics variants and prognosis was also studied. The list of high-impact genetic variants was unique for each patient. However, the pathways in which these genes are involved are well-known hallmarks of cancer, such as angiogenesis or immune pathways. Additionally, we determined that genetic variants in BRCA2, ZNF750, FAM208B, ZNF599, and ZC3H13 genes are related with poor disease-free survival in ASCC. This may help to stratify the patient's prognosis and open new avenues for potential therapeutic intervention. In conclusion, sequencing of ASCC clinical samples appears an encouraging tool for the molecular portrait of this disease.

Introduction

Anal squamous cell carcinoma (ASCC) is a rare tumor. In 2019, an estimated 8300 new cases will occur in the United States, representing approximately 2.5% of all gastrointestinal cancers [1].

Since the 1970s, the standard treatment has consisted of a combination of 5-fluorouracil (5FU) with mitomycin C or cisplatin and radiotherapy [2,3]. Despite this treatment being very effective for early-stage tumors, the disease-free survival (DFS) rate in T3-T4 or N + tumors ranges between 40% and 70% [4,5]. Patients diagnosed with ASCC do not benefit from

---

targeted therapy or immunotherapy. In addition, there is insufficient information on molecular prognostic or response prediction factors.

With the improvements in high-throughput molecular techniques, it is possible to study several variables instead of the classical gene-centered view. These technical advances allow for the study of multiple genetic alterations from clinical samples. Exome sequencing (ES) has contributed to the identification of new disease-causing genes and is now being incorporated into clinical practice [6]. Since the first work reporting ES [7], numerous medical sequencing projects have faced the challenge of identifying molecular alterations related to rare diseases or cancers [8]. The Cancer Genome Atlas is making huge strides in characterizing several tumor types by comprehensive molecular techniques. However, ASCC is not included because this project is focused on more frequent tumors.

Previous studies have analyzed metastatic or primary ASCC tumors by ES or even by gene panels in an attempt to describe the most frequent alterations in this disease. These studies established *PIK3CA* as a frequently mutated gene in ASCC [9–12]. However, the exact relationship between genetic variations, phenotype, and tumor evolution is currently unknown.

In this study, we analyzed 46 ASCC formalin-fixed, paraffin-embedded (FFPE) samples. On the one hand, we characterized the main genetic variants present in these tumors and the main biological processes in which these genes are involved, while on the other hand, we identified those genes in which the presence of a genetic variant is associated with DFS in ASCC.

Materials and Methods

Patients

Forty-six treatment-naive FFPE samples from patients diagnosed with localized ASCC were analyzed by ES. All tumor samples were reviewed by an experienced pathologist. All the samples contained at least 70% invasive tumor cells. Informed consent was obtained for all patients, and the study was approved by the Hospital Universitario La Paz Research Ethics Committee. Patients were required to have a histologically confirmed diagnosis of ASCC; be 18 years of age or older; have an Eastern Cooperative Oncology Group performance status score from 0 to 2; have not received prior radiotherapy or chemotherapy for this malignancy; and present with no distant metastasis. Demographic characteristics related to the tumor and the treatments were collected. The presence of human papillomavirus (HPV) infection was determined using CLART HPV2 (Genomica).

DNA Isolation

One 10-mm section from each FFPE sample was deparaffinized, and DNA was extracted with GeneRead DNA FFPE Kit (Qiagen), following the manufacturer's instructions. Once eluted, DNA was frozen at −80°C until use.

Library Preparation, Exome Capture, and Illumina Sequencing

ES from 46 FFPE samples of ASCC was performed. Purified DNA was quantified by Picogreen, and mean size was determined by gel electrophoresis. Genomic DNA was fragmented by mechanical means (Bioruptor) to a mean size of approximately 200 bp. Then, DNA samples were repaired, phosphorylated, A-tailed, and ligated to specific adaptors, followed by PCR-mediated labeling with Illumina-specific sequences and sample-specific barcodes (Kapa DNA library generation kit).

Exome capture was performed using the VCRome system (capture size of 37 Mb, Roche Nimblegen) under a multiplexing of eight samples per capture reaction. Capture was strictly carried out following manufacturer's instructions. After capture, libraries were purified, quantified, and titrated by real-time PCR before sequencing. Samples were then sequenced to an approximate coverage of 4.5 Gb per sample in Illumina-NextSeq NS500 (Illumina Inc.) using 150-cycle (2×75) high-output cartridges.

Bioinformatics Exome Sequencing Data Processing

The quality of the ES experiments was verified using FASTQC (http://www.bioinformmatics.babraham.ac.uk/projects/fastqc).

First, adaptors were removed using Cutadapt [13], and FASTQ files were filtered by quality using PrinSeq [14]; both tools are included in the GPRO Suite (Biotechvana SL) [15]. Alignment of the sequences was achieved using the human genome h19 as the reference genome. The tools BWA [16], Samtools [17], and Picard Tools (http://picard.sourceforge.net) were used. Variant calling was performed using the MuTect tool from the GATK4 package [18] combined with PicardTools, first, to create a panel of normal samples and, second, for the variant calling [19]. The panel of normal samples was built using 11 samples from Iberian exomes from 1000 genomes (http://www.ncbi.nlm.nih.gov/sra/), and it was used to discard germline variants.

Prioritization of High-Impact Genetic Variants

With the aim of establishing the genetic variants that may act as drivers of the disease, the VarSEQ software (Golden Helix) was used. Variant Call Format files were filtered according to strict criteria: a reading depth of at least $15\times$, a gnomAD global frequency <1%, high impact (frameshift, splice variants, and stop-loss and stop-gain variants), and a detectable presence in at least 15% of the reads [20].

Filtering of the Most Frequent Genetic Variants in Our Cohort

On the other hand, with the aim of establishing the most frequent high- and moderate-impact variants, genetic variants were further annotated using Variant Effect Predictor (VEP) [21] and the Varsome database (https://varsome.com/). Then, the information provided by VEP was used to filter the genetic variants. The filtering criteria were a frequency in the general population, according to the gnomAD database, of less than 1%; a high or moderate impact; and the presence of the variant in at least 10% of the patients in our cohort.

Visual Validation of the High-Impact Genetic Variants

The BAM files containing the prioritized variants were verified using Alamut Visual v2.11 (Interactive Biosoftware).

Pathway Annotation

Pathway annotation of the genes containing the prioritized variants was done using PANTHER database (http://www.pantherdb.org/) and DAVID webtool [22] selecting GO-BP as the category and *Homo sapiens* as the background.

Genes Associated with Prognosis

From the list of genes with high- or moderate-impact variants identified by VEP, those genes in which the presence of a genetic variant was associated with DFS were identified using BRB Array Tools developed by Dr. Richard Simon's team [23]. Genes associated with DFS were selected according to their *P* values by a Kaplan-Meier analysis. DFS was defined as the time from primary tumor surgery until local and/or distant tumor relapse. For survival analyses, only the 41 patients treated with chemoradiotherapy were included (Figure 1).

Statistical Analyses

Statistical analyses were performed using GraphPad Prism v6 and IBM SPSS Statistics v20. All *P* values were two-tailed, and statistical significance was set as .05 or less.
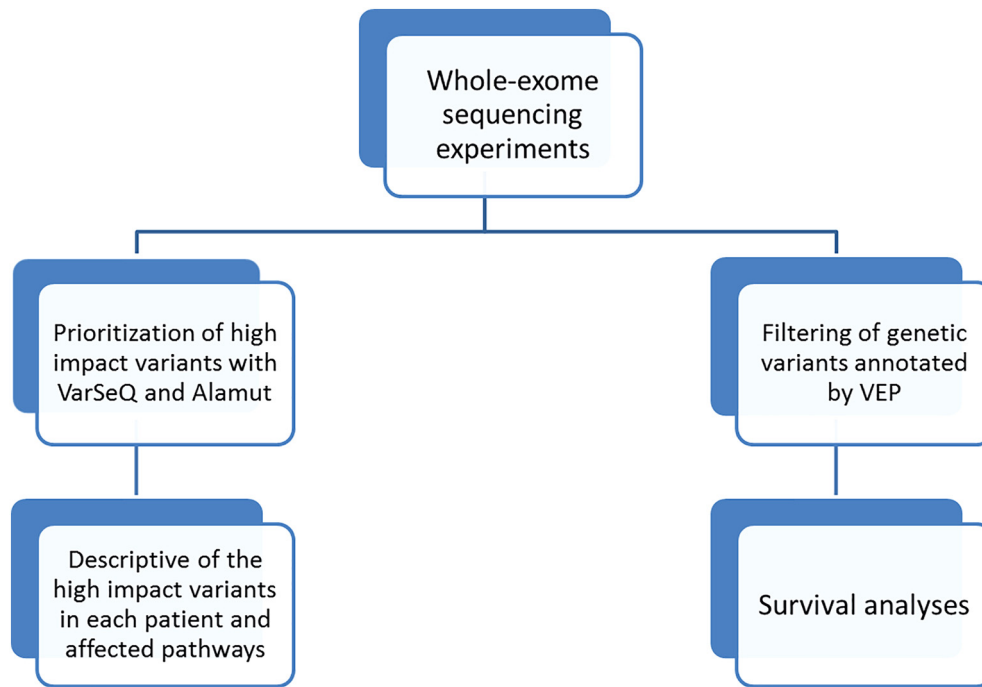
Figure 1. Workflow followed in this study.

Results

Patient Cohort

Forty-six FFPE samples of ASCC were analyzed by ES. Twenty-eight samples were from patients included in the VITAL clinical trial (GEMCAD-09-02, NCT01285778). These patients were treated with panitumumab, 5FU, and mitomycin C, concomitantly with radiotherapy. The remaining 18 patients were retrospectively included from the clinical practice in Hospital Universitario La Paz and Hospital Clinic. Fourteen patients were treated with cisplatin-5FU or mitomycin C-5FU, concomitantly with radiotherapy.

Table 1
Patient Clinical Characteristics

	Number of Patients	Percentage (%)
Number of patients	46	100%
Age at diagnosis (median and range)	61 (41-86)	
Age at diagnosis (mean)	61	
Gender		
Male	23	50%
Female	23	50%
HPV		
16	26	57%
Other subtypes	8	17%
Negative	6	13%
Unknown	6	13%
HIV		
Positive	2	4%
Negative	44	96%
Lymph node status		
N0	19	41%
N positive	24	52%
Unknown	3	7%
TNM stage		
I	3	7%
II	16	35%
III	27	58%
Treatment		
Chemoradiotherapy	14	30%
Chemoradiotherapy-panitumumab	28	61%
Other	4	9%

Three patients who were initially treated only with surgery and one patient treated with radiotherapy alone were excluded from the survival analyses (Table 1).

Exome Sequencing Experiments

The mean coverage obtained in ES experiments was $>42.6\times$, with the exception of one sample with a coverage of $3.57\times$ that was excluded from subsequent analyses. Once this sample was dismissed, the remaining samples presented a mapping efficiency of between 90% and 98% with the exception of one sample (75.4%). The human exome has >195,000 exonic regions, out of which only 23,021 (11.21%) were not mapped in any sample.

Relevant Genes and Their Associated Genetic Variants

VarSEQ and Alamut were used to filter and visualize the variants that may play a significant role in the development of ASCC. For this purpose, we studied all genetic variants that caused a high impact in their respective gene. A total of 333 high-impact variants across 312 genes in the 45 patients were prioritized by the VarSEQ software.

The list of the implicated genes with high-impact variants is shown in Supplementary Table 1.

Within the high-impact variants, the most frequent type of alteration was the nonsense substitution which introduces a premature STOP codon (Table 2).

Table 2
Description of the Type of High-Impact Alterations Detected by WES Analysis

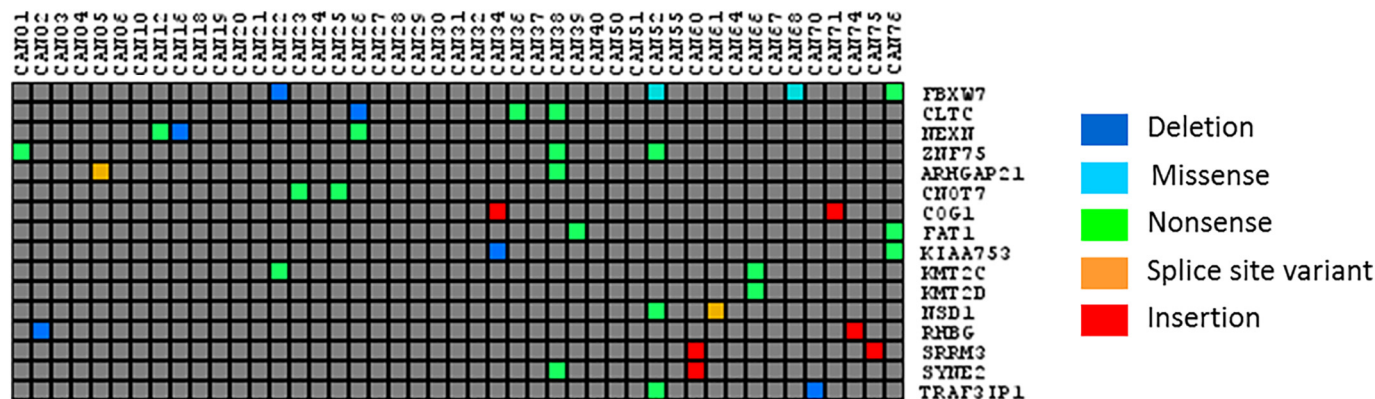| Type of Alteration	Frequency
Duplications	16 (4.8%)
Insertions	32 (9.6%)
Deletions	55 (16.4%)
Nonsense substitutions	204 (62%)
Splice site substitutions	19 (6%)
Intronic deletions	2 (0.6%)
Intronic substitutions	2 (0.6%)

Figure 2. List of high-impact genetic variants identified by VarSEQ in our cohort in more than one patient.

The principal alterations identified by these analyses were different between patients; i.e., each patient presented a unique set of high-impact genetic variants. However, in some cases, the genes affected by these high-impact alterations were common. The list of these genes is shown in Supplementary Table 2 and Figure 2.

Ontology Analysis of the High-Impact Genetic Variants Identified

With the aim of identifying the pathways affected by the genetic alterations, PANTHER and DAVID databases were used. The most frequently implicated pathways were intracellular signal transduction (22% of the genes, 24% of the patients in the cohort), immune (20% of the genes, 15% of the patients), and apoptosis (17% of the genes, 15% of the patients) pathways, although there were also relatively frequent alterations in genes implicated in angiogenesis (11% of the genes, 11% of the patients), metabolism (11% of the genes, 9% of the patients), chromatin modification (8% of the genes, 11% of the patients), EGFR signaling (8% of the genes, 7% of the patients), or Wnt signaling pathways (8% of the genes, 11% of the patients) (Figure 3).

Characterization of the Most Frequently Mutated Genes in Our Cohort

With the aim of characterizing the most frequently mutated genes in ASCC, genetic variants were annotated using VEP, and genes with high-

and moderate-impact genetic variants were analyzed. After filtering by VEP results, 382 genes were found to present at least one high- or moderate-impact genetic variant in at least 10% of the patients in our cohort (Supplementary Table 3). The functions associated with these 382 genes were cytoskeleton, DNA repair, adhesion, and chromatin binding. *PIK3CA* was mutated in 40% of the patients, *FBXW7* in 16%, *FAT1* in 18%, and *ATM* in 27% of the patients. The most frequent variants found in *PIK3CA* are rs104886003 (seven patients), classified as a variant of uncertain significance in Varsome, and rs121913273 (three patients), classified as likely pathogenic by the same database.

Genetic Variants Associated with Prognosis in ASCC

With the aim of determining the genes associated with relapse in ASCC, a Kaplan-Meier analysis was performed. This analysis showed that, in this cohort, the presence of a high- or moderate-impact genetic variant in *BRCA2*, *ZNF750*, *FAM208B*, *ZNF599*, and *ZC3H13* was associated with poor disease-free survival (Figure 4). The genetic variants detected in these genes are summarized in Table 3. Presenting more than one genetic variant in any of these genes implied a worse DFS ($P = .001$) (Figure 5). Patients without genetic variants or with only a single genetic variant in these genes did not reach the median DFS, and DFS percentages at 60 months were 54% and 29%, respectively. The median DFS in those patients with
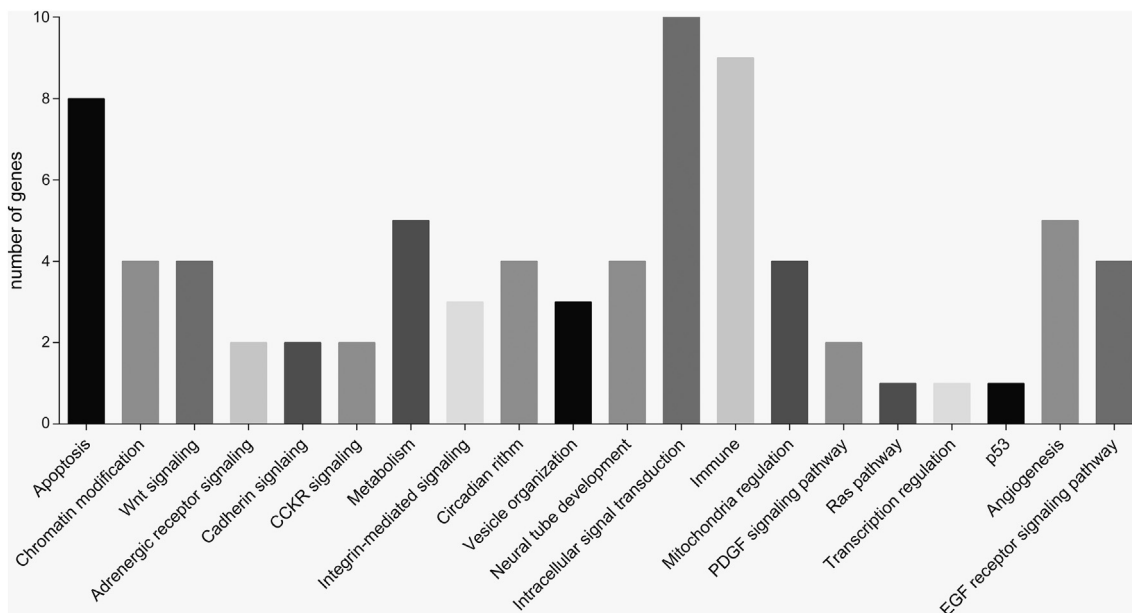

Figure 3. Pathways in which genes presenting high-impact genetic variants are involved.
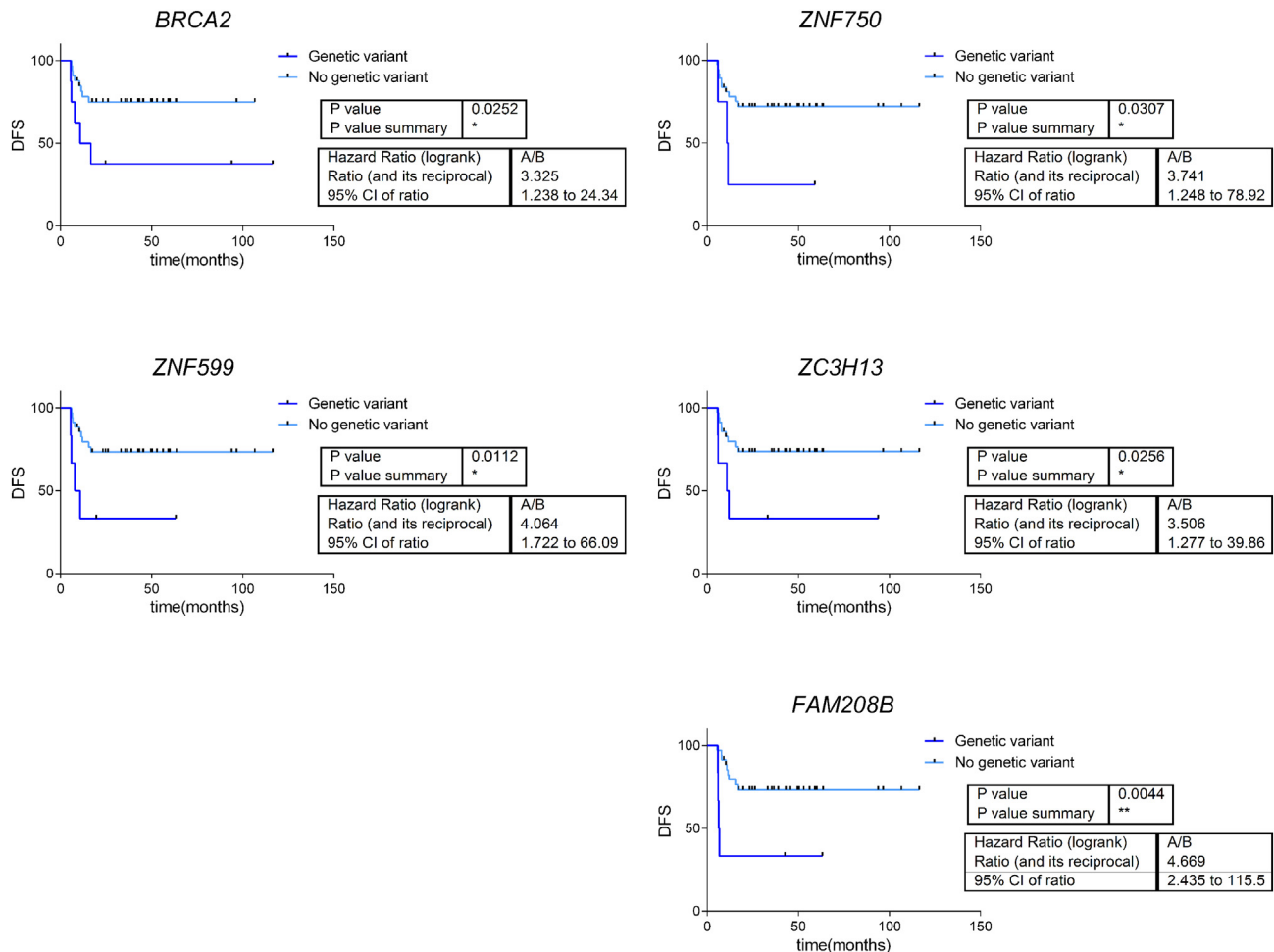
Figure 4. Survival curves for those patients whose high- and moderate-impact genetic variants are associated with DFS.

two or more genetic variant in these genes was 7 months, and DFS percentage at 60 months was 86%.

Discussion

ASCC is a relatively rare type of cancer, although its incidence has increased in recent years. The management of these patients consists of mitomycin C combined with 5FU chemotherapy and radiotherapy, but not all patients obtain a benefit from these treatments. For this reason, it is necessary to delve deeper into the molecular characterization of this disease, with the objective of identifying biomarkers and new potential therapeutic targets. In this study, a cohort of 46 ASCC patients was studied by ES, and a list of genes with high-impact genetic variants that may be drivers of oncogenesis processes was identified. In addition, a characterization of the most frequently mutated genes with high- and moderate-impact variants was performed. Finally, five genes in which the presence of a genetic variant was associated with DFS were identified.

In this study, 46 ASCC paraffin samples were analyzed by ES. The analysis of FFPE samples using next-generation sequencing has been challenging due to the DNA fragmentation and the artificial alterations of the sequence caused by the fixation process. However, paired comparisons between fresh-frozen tissue and FFPE samples showed that although DNA damage in FFPE samples was evident, the results of fresh-frozen tissue and FFPE samples sequencing were comparable [24].

PIK3CA variants with a well-established functional effect were identified in 40% of our cohort samples. This value is consistent with that reported in previous studies [11,12]. The most frequent *PIK3CA* variants were rs121913273 and rs104886003, both described in the COSMIC and

Varsome databases. The rs121913273 variant has been described as likely pathogenic in esophageal and squamous cell carcinoma of the head and neck and in uterine cervical neoplasms (VCV000376244.2). Our results are comparable with the percentage observed in previous studies trying to identify the predominant genetic variants in ASCC [9–12]. Cacheux et al. established variants in *PIK3CA* as a frequent alteration (20.3%) in this type of cancer [10]. Chung et al. studied 70 patients and described recurrent alterations in 40 genes such as *PIK3CA* (40%), *FBXW7* (13%), *PTEN* (14%), and *RICTOR* (9%) [12]. Both studies included localized but also metastatic ASCC. Morris et al. analyzed metastatic ASCC samples combining ES and gene panels. They confirmed that *PIK3CA* is commonly mutated in ASCC (29%), and they used a xenograft model to test EGFR and PIK3CA inhibitors with a decrease in tumor growth [11]. In a later work, Cacheux et al. performed ES on 20 ASCC patients and described *PIK3CA* (25%), *FBXW7* (15%), and *FAT1* (15%) as frequently mutated genes, and *TRIP12* (15%), and chromatin remodeling as a pathway playing an important role in ASCC [9].

In contrast to the gene panel-based analyses, the nondirected ES approach made it possible to propose a list of genetic variants which may play an important role in the development of the disease and to identify five genes in which the presence of a genetic variant is associated with DFS. Moreover, this analysis replicated the identification of genetic variants in some genes previously described in ASCC, such as *FBWX7*, but also suggested the involvement of novel genes, such as *ZNF750*.

In addition, we found that 16% of the patients presented a high- or moderate-impact variant in *FBXW7*, 18% in *FAT1*, and 27% in *ATM*. A remarkable finding was the presence of high-impact genetic variants of the *FBXW7* gene in four patients. *FBXW7* is a cell cycle key regulator, and its

Table 3
Genetic Variants Associated with Disease-Free Survival in ASCC.

Gene	Chromosomal Location	Number of Mutated Samples	%	Genetic Variants	cDNA Level	Protein Level
BRCA2	chr13	8	18%	chr13_32914858_G/T	NM_000059.3:c.6366G > T	p.(Met2122Ile)
				chr13_32936775_G/A	NM_000059.3:c.7921G > A	p.(Glu2641Lys)
				chr13_32972631_-/TAGAC	NM_000059.3:c.9980_9981insTAGAC	p.(Lys3327Asnfs*15)
				chr13_32900790_G/A	NM_000059.3:c.631 + 40G > A	Not predictable
				chr13_32936713_T/C	NM_000059.3:c.7859 T > C	p.(Val2620Ala)
				chr13_32914516_ − /G	NM_000059.3:c.6024dup	p.(Gln2009Alafs*9)
				chr13_32968918_C/T	NM_000059.3:c.9349C > T	p.(His3117Tyr)
				chr13_32953641_A/G	NM_000059.3:c.8942A > G	p.(Glu2981Gly)
ZNF750	chr17	4	9%	chr17_80788075_C/T	NM_024702.2:c.2115G > A	p.(Ala705 =)
				chr17_80789547_C/ −	NM_024702.2:c.784del	p.(Ala262Leufs*104)
				chr17_80789802_C/A	NM_024702.2:c.529G > T	p.(Glu177*)
				chr17_80789503_G/C	NM_024702.2:c.828C > G	p.(Tyr276*)
ZNF599	chr19	6	13%	chr19_35249946_C/G	NM_001007248.2:c.1760G > C	p.(Arg587Thr)
				chr19_35250950_T/A	NM_001007248.2:c.756A > T	p.(Glu252Asp)
				chr19_35251215_G/A	NM_001007248.2:c.491G > A	p.(Gly164Asp)
				chr19_35250352_C/T	NM_001007248.2:c.1354G > A	p.(Glu452Lys)
				chr19_35260373_C/T	NM_001007248.2:c.106G > A	p.(Glu36Lys)
				chr19_35250567_G/T	NM_001007248.2:c.1139C > A	p.(Ser380*)
ZC3H13	chr13	6	13%	chr13_46616368_C/ −	NM_015070.4:c.270del	p.(Asn91Thrfs*37)
				chr13_46539442_C/T	NM_015070.4:c.4447G > A	p.(Asp1483Asn)
				chr13_46542026_CGCG/ −	NM_015070.4:c.3931_3934del	p.(Glu1314Glyfs*48)
				chr13_46563072_G/A	NM_015070.4:c.1105C > T	p.(Arg369Cys)
				chr13_46584507_G/A	NM_015070.4:c.722C > T	p.(Ser241Phe)
				chr13_46563125_C/T	NM_015070.4:c.1052G > A	p.(Arg351Gln)
FAM208B	chr10	6	13%	chr10_5784388_T/C	NM_001321783.1:c.2656 T > C	p.(Leu886 =)
				chr10_5781712_G/A	NM_001321783.1:c.1579G > A	p.(Val527Ile)
				chr10_5762722_G/C	NM_001321783.1:c.31G > C	p.(Glu11Gln)
				chr10_5782394_C/A	NM_001321783.1:c.2261C > A	p.(Ala754Glu)
				chr10_5789623_G/C	NM_001321783.1:c.4239G > C	p.(Glu1413Asp)
				chr10_5781667_G/C	NM_001321783.1:c.1534G > C	p.(Glu512Gln)

function as a tumor suppressor is well known [25]. Genetic variants in this gene have been previously described in 15% of ASCC patients. *FAT1*, which presented high-impact variants in two patients of our cohort, has also been described as a tumor suppressor gene involved in the Wnt pathway [26]. *FAT1* genetic variants have also been described in ASCC patients [9].

Strikingly, each patient presented a unique set of high-impact variants. The most frequent alterations were nonsense stop-gain variants. However, the pathway analysis showed that, despite the individual variant diversity between patients, the affected biological pathways were shared between patients. This has already been seen in other tumors from The Cancer Genome Atlas studies [27]. Moreover, the implicated pathways are traditionally associated with tumor progression and cancer development such as angiogenesis or metabolism pathways [28]. The presence of high-impact variants in the EGFR pathway in 7% of patients or in immunological processes may be relevant when proposing new therapies such as EGFR-targeted therapy or immunotherapy.

Figure 5. Survival curves obtained after considering the presence of high- and moderate-impact genetic variants in *BRCA2*, *FAM208B*, *ZNF750*, *ZNF599*, and *ZC3H13* (0 genetic variant, *n* = 23 patients; 1 genetic variant, *n* = 11 patients; 2-5 genetic variants, *n* = 7 patients).

With the aim of determining the most frequent alterations in our cohort and identifying variants related to the progression of the disease, less restrictive filtering criteria were used. This led us to the identification of high- or moderate-impact genetic variants in 382 genes. These genes are mainly involved in DNA repair, chromatin binding, cytoskeleton, and adhesion processes. Chromatin remodeling was previously suggested as a process with an important role in ASCC [9].

Survival analyses identified five genes associated with DFS. *ZNF750* was recently associated with prognosis in esophageal squamous cell carcinoma [29,30]. Variants of *BRCA2* were also associated with esophageal squamous cell carcinoma and head and neck squamous cell carcinoma, both of which are related to HPV infection [31,32]. Moreover, BRCA mutations are associated with response to PPAR inhibitors in pancreatic, breast, and ovarian tumors [33–35]. Therefore, the prevalence of *BRCA2* mutations may have therapeutic implications. Other genes associated with DFS (*FAM208B*, *ZNF599*, and *ZC3H13*) had not been previously associated with cancer.

The two analyses performed offer complementary insights. The most restrictive filtering identified those genetic variants with a highly deleterious impact upon the resulting protein and proposed a list of variants that may be related to tumor development. In addition, based on the affected genes, the principal altered pathways involved were identified. In contrast, the alternative filtering pipeline yielded the most frequent variants in our cohort, some of them also associated with tumor progression and DFS.

ES analysis allowed us to identify candidate genes potentially involved in the etiology of ASCC without the bias of previous knowledge. On the other hand, our study led to the identification of genes that presented genetic variants associated with DFS in ASCC. As far as we know, this is the first time that an analysis of these characteristics has been performed. Another important point is the sample size. With ASCC being a rare tumor, it is difficult to gather a representative and homogeneous cohort. Our study analyzed 46 samples, the second largest cohort in ASCC from localized tumors treated with chemoradiotherapy so far.

This study has some limitations. A validation with an independent cohort, especially for the prognostic genetic variants, is still needed. Unlike
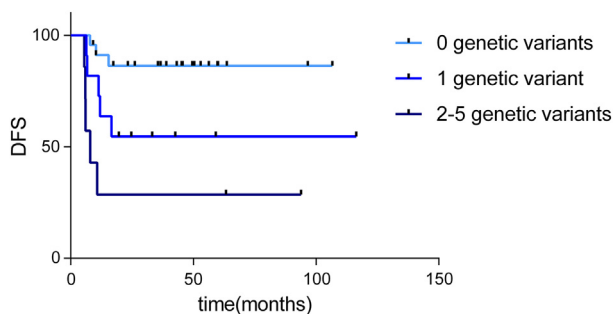
the traditional genetic diseases, where an analysis pipeline is well established, in tumor sequencing data, it is still necessary to establish a consensus analysis workflow.

Conclusions

To summarize, we analyzed 46 ASCC tumor samples by ES, an important cohort taking into account that ASCC is a rare tumor. Our study identified a set of variants having a high impact upon the corresponding protein and proposed a list of candidate genes which may play an important role in the etiology of ASCC. On the other hand, our study yielded a list of genetic variants apparently associated with poor disease progression. In conclusion, sequencing of ASCC clinical samples seems to be a promising tool for the molecular characterization of this pathology, which may aid the clinician in the prognostic stratification of patients and open new avenues for drug development.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tranon.2020.100778.

Declarations

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of Hospital Universitario La Paz (PI-1926). Informed consents were obtained for all patients included in this study.

Availability of Supporting Data

The exome sequencing data are available in SRA (https://www.ncbi.nlm.nih.gov/sra) under the name PRJNA573670.

Funding

Declaration of Interest

J. A. F. V. and A. G.-P. are shareholders in Biomedica Molecular Medicine SL. L. T.-F. and G. P.-V. are employees of Biomedica Molecular Medicine SL. J. C. has received scientific consulting fees (speaker and advisory roles) from Novartis, Pfizer, Ipsen, Exelixis, Bayer, Eisai, Advanced Accelerator Applications, Amgen, Sanofi, and Merck Serono and research support from Eisai, Novartis, Ipsen, Astrazeneca, Pfizer, and Advanced Accelerator Applications. I. G. has received fees and/or travel expenses from Roche, Sanofi, Merck, Servier, Amgen, and Sirtflex, and consulting fees from Merck and Sanofi. J. F. has received honoraria for consulting and advisory role in Amgen, Ipsen, Eissai, Merck, Roche, and Novartis; research funding from Merck. and travel expenses from Amgen and Servier. J. M. reports grants (clinical trial lead principal investigator) from Merck, Amgen, and Servier and grants (project lead principal investigator) from NanoString, Incyte, and Biocartis and also reports personal fees from Sirtex, Pierre-Fabre, Shire, Astra-Zeneca, Bayer, Servier, Sanofi, and Roche, outside the scope of this work. R. G.-C. declares having provided consulting services and/or received fees from AAA, Amgen, Bayer, Ipsen, Lilly, Merck, MSD, Novartis, PharmaMar, Pfizer, Roche, and Sanofi-Aventis and has received

research support from Pfizer and BMS, all beyond the scope of this work. The other authors declare no conflicts of interests.

Authors' Contributions

All the authors have directly participated in the preparation of this manuscript and have approved the final version submitted and declare no ethical conflicts of interest. R. L.-V., M. M., E. L.-C., and V. H. contributed the DNA extraction. M. M. and V. H. contributed the HPV determination. L. G.-P., C. P., and M. C. contributed the pathological review of the samples. R. R.-R. and C. L. contributed the sequencing analyses. I. G., J. M., P. G.-A., J. C., C. C., R. G.-C., and J. F. contributed the clinical data and the analyses related. L. T.-F., I. G., A. G.-P., G. P.-V., and A. Z.-M. contributed in the design of the study and the statistical and prognosis analyses. L. T.-F., A. G.-P., A. C.-B., and K. E. H. contributed to high-impact genetic variants analysis. L. T.-F. drafted the manuscript. I. G., J. M., J. A. F. V., and J. F. conceived of the study and participated in its design and interpretation. A. G.-P., J. M., J. A. F. V., and J. F. reviewed the manuscript. J. F. and J. A. F. V. coordinated the study. All authors read and approved the final manuscript.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2019, CA Cancer J. Clin. 69 (1) (2019) 7–34, Epub 2019/01/08 https://doi.org/10.3322/caac.21551, PubMed 30620402.

[2] A. Benson, A. Venook, M. Al-Hawary, L. Cederquist, Y. Chen, K. Ciombor, et al., Anal carcinoma, version 2.2018, NCCN clinical practice guidelines in oncology, J Natl Compr Canc Netw (2018) 852–871.

[3] R. Glynne-Jones, P.J. Nilsson, C. Aschele, V. Goh, D. Peiffert, A. Cervantes, et al., Anal cancer: ESMO-ESSO-ESTRO clinical practice guidelines for diagnosis, treatment and follow-up, Eur. J. Surg. Oncol. 40 (10) (2014) 1165–1176, https://doi.org/10.1016/j.ejso.2014.07.030, PubMed 25239441.

[4] L.L. Gunderson, K.A. Winter, J.A. Ajani, J.E. Pedersen, J. Moughan, A.B. Benson, et al., Long-term update of US GI intergroup RTOG 98-11 phase III trial for anal carcinoma: survival, relapse, and colostomy failure with concurrent chemoradiation involving fluorouracil/mitomycin versus fluorouracil/cisplatin, J Clin Oncol 30 (35) (2012) 4344–4351, Epub 2012/11/13 https://doi.org/10.1200/JCO.2012.43.8085, PubMed 23150707 PubMed Central. PMCPMC3515768.

[5] J.A. Ajani, K.A. Winter, L.L. Gunderson, J. Pedersen, A.B. Benson, C.R. Thomas, et al., Prognostic factors derived from a prospective database dictate clinical biology of anal cancer: the intergroup trial (RTOG 98-11), Cancer 116 (17) (2010) 4007–4013, https://doi.org/10.1002/cncr.25188, PubMed 20564111 PubMed Central. PMCPMC3831519.

[6] M. Tetreault, E. Bareke, J. Nadaf, N. Alirezaie, J. Majewski, Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities, Expert. Rev. Mol. Diagn. 15 (6) (2015) 749–760, Epub 2015/05/09 https://doi.org/10.1586/14737159.2015.1039516, PubMed 25959410.

[7] S. Ng, E. Turner, P. Robertson, S. Flygare, A. Bigham, C. Lee, et al., Targeted capture and massively parallel sequencing of 12 human exomes, Nature 461 (7261) (2009) 272–276, Epub 2009 Aug 16 https://doi.org/10.1038/nature08250.

[8] B. Rabbani, M. Tekin, N. Mahdieh, The promise of whole-exome sequencing in medical genetics, J. Hum. Genet. 59 (1) (2014) 5–15, Epub 2013/11/07 https://doi.org/10.1038/jhg.2013.114, PubMed 24196381.

[9] W. Cacheux, V. Dangles-Marie, E. Rouleau, J. Lazartigues, E. Girard, A. Briaux, et al., Exome sequencing reveals aberrant signalling pathways as hallmark of treatment-naive anal squamous cell carcinoma, Oncotarget 9 (1) (2018) 464–476, Epub 2017/12/08 10.18632/oncotarget.23066, PubMed 29416628 PubMed Central. PMCPMC5787481.

[10] W. Cacheux, E. Rouleau, A. Briaux, P. Tsantoulis, P. Mariani, M. Richard-Molard, et al., Mutational analysis of anal cancers demonstrates frequent PIK3CA mutations associated with poor outcome after salvage abdominoperineal resection, Br J Cancer 114 (12) (2016) 1387–1394, Epub 2016/05/24 https://doi.org/10.1038/bjc.2016.144, PubMed 27219019 PubMed Central. PMCPMC4984471.

[11] V. Morris, X. Rao, C. Pickering, W.C. Foo, A. Rashid, K. Eterovic, et al., Comprehensive genomic profiling of metastatic squamous cell carcinoma of the anal canal, Mol Cancer Res 15 (11) (2017) 1542–1550, Epub 2017/08/07 https://doi.org/10.1158/1541-7786.MCR-17-0060, PubMed 28784613 PubMed Central. PMCPMC5991496.

[12] J.H. Chung, E. Sanford, A. Johnson, S.J. Klempner, A.B. Schrock, N.A. Palma, et al., Comprehensive genomic profiling of anal squamous cell carcinoma reveals distinct genomically defined classes, Ann. Oncol. 27 (7) (2016) 1336–1341, Epub 2016/04/06 https://doi.org/10.1093/annonc/mdw152, PubMed 27052656.

[13] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnetjournal 17 (1) (2011) 10–12, https://doi.org/10.14806/ej.17.1.200.

[14] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, Bioinformatics 27 (6) (2011) 863–864.

[15] R. Futami, A. Muñoz-Pomer, J. Viu, R. Domínguez-Escribá, L. Covelli, G. Bernet, et al., GPRO The professional tool for annotation, management and functional analysis of omic databases, Biotechvana Bioinformatics: SOFT3. (2011).

[16] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 Epub 2009 May 18.

[17] H. Li, B. Handsaker, B. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., The Sequence Alignment/Map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.

[18] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (9) (2010) 1297–1303, https://doi.org/10.1101/gr.107524.110 Epub 2010 Jul 19.

[19] K. Cibulskis, M. Lawrence, S. Carter, A. Sivachenco, D. Jaffe, C. Sougnez, et al., Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, Nat. Biotechnol. 31 (3) (2013) 213–219, https://doi.org/10.1038/nbt.2514 Epub 2013 Feb 10.

[20] Jalali Sefid Dasthi M, J G. A practical guide to filtering and prioritizing genetic variants. BioTechniques 2017;62(1):18–30. doi: https://doi.org/10.2144/000114492.

[21] Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. Nucleic Acids Res. 2015;43(Database issue):D662-9. Epub 2014/10/28. doi: https://doi.org/10.1093/nar/gku1010. PubMed PMID: 25352552; PubMed Central PMCID: PMCPMC4383879.

[22] dW Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (1) (2009) 44–57, https://doi.org/10.1038/nprot.2008.211, PubMed 19131956.

[23] R. Simon, Roadmap for developing and validating therapeutically relevant genomic classifiers, J. Clin. Oncol. 23 (29) (2005) 7332–7341, Epub 2005/09/06 https://doi.org/10.1200/JCO.2005.02.8712, PubMed 16145063.

[24] E. Oh, Y.L. Choi, M.J. Kwon, R.N. Kim, Y.J. Kim, J.Y. Song, et al., Comparison of accuracy of whole-exome sequencing with formalin-fixed paraffin-embedded and fresh frozen tissue samples, PLoS One 10 (12) (2015) e0144162, Epub 2015/12/07 https://doi.org/10.1371/journal.pone.0144162 PubMed. 26641479. PubMed Central. PMCPMC4671711.

[25] S. Takeishi, K.I. Nakayama, Role of Fbxw7 in the maintenance of normal stem cells and cancer-initiating cells, Br J Cancer 111 (6) (2014) 1054–1059, Epub 2014/05/22 https://doi.org/10.1038/bjc.2014.259, PubMed 24853181 PubMed Central. PMCPMC4453837.

[26] L.G. Morris, A.M. Kaufman, Y. Gong, D. Ramaswami, L.A. Walsh, Ş. Turcan, et al., Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation, Nat Genet 45 (3) (2013) 253–261, Epub 2013/01/27 https://doi.org/10.1038/ng.2538, PubMed 23354438 PubMed Central. PMCPMC3729040.

[27] M.H. Bailey, C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, et al., Comprehensive characterization of cancer driver genes and mutations, Cell 174 (4) (2018) 1034–1035, https://doi.org/10.1016/j.cell.2018.07.034, PubMed 30096302.

[28] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (5) (2011) 646–674, https://doi.org/10.1016/j.cell.2011.02.013, PubMed 21376230.

[29] R. Otsuka, Y. Akutsu, H. Sakata, N. Hanari, K. Murakami, M. Kano, et al., ZNF750 expression is a potential prognostic biomarker in esophageal squamous cell carcinoma, Oncology 94 (3) (2018) 142–148, Epub 2017/12/08 https://doi.org/10.1159/000484932, PubMed 29216641 PubMed Central. PMCPMC5969074.

[30] M. Hazawa, D.C. Lin, H. Handral, L. Xu, Y. Chen, Y.Y. Jiang, et al., ZNF750 is a lineage-specific tumour suppressor in squamous cell carcinoma, Oncogene 36 (16) (2017) 2243–2254, Epub 2016/11/07 https://doi.org/10.1038/onc.2016.377, PubMed 27819679 PubMed Central. PMCPMC5415641.

[31] M.R. Akbari, R. Malekzadeh, P. Lepage, D. Roquis, A.R. Sadjadi, K. Aghcheli, et al., Mutations in Fanconi anemia genes and the risk of esophageal cancer, Hum. Genet. 129 (5) (2011) 573–582, Epub 2011/01/30 https://doi.org/10.1007/s00439-011-0951-7, PubMed 21279724.

[32] G.M. Low, D.S. Thylur, V. Yamamoto, U.K. Sinha, The effect of human papillomavirus on DNA repair in head and neck squamous cell carcinoma, Oral Oncol. 61 (2016) 27–30, Epub 2016/08/18 https://doi.org/10.1016/j.oraloncology.2016.08.002, PubMed 27688101.

[33] M. Munroe, J. Kolesar, Olaparib for the treatment of BRCA-mutated advanced ovarian cancer, Am. J. Health Syst. Pharm. 73 (14) (2016) 1037–1041, https://doi.org/10.2146/ajhp150550, PubMed 27385701.

[34] M. Robson, C. Goessl, S. Domchek, Olaparib for metastatic germline BRCA-mutated breast cancer, N. Engl. J. Med. 377 (18) (2017) 1792–1793, https://doi.org/10.1056/NEJMc1711644, PubMed 29091556.

[35] T. Golan, P. Hammel, M. Reni, E. Van Cutsem, T. Macarulla, M.J. Hall, et al., Maintenance olaparib for germline BRCA-mutated metastatic pancreatic cancer, N Engl J Med. (2019) https://doi.org/10.1056/NEJMoa1903387, Epub 2019/06/02. PubMed 31157963.