Genetic profile and functional proteomics of anal squamous cell carcinoma: proposal for a molecular classification

Lucía Trilla-Fuertes[1#], Ismael Ghanem[2#], Angelo Gámez-Pozo[3], Joan Maurel[4], Laura G-Pastrián[5,6], Marta Mendiola [6,7], Cristina Peña [5], Rocío López-Vacas[3], Guillermo Prado-Vázquez[1], Elena López-Camacho[3], Andrea Zapater-Moros[3], Victoria Heredia[7,8], Miriam Cuatrecasas [9], Pilar García-Alfonso [10], Jaume Capdevila[11], Carles Conill[12],Rocío García-Carbonero[13], Ricardo Ramos-Ruiz[14], Claudia Fortes[15], Carlos Llorens[16], Paolo Nanni[15], Juan Ángel Fresno Vara [3,7], Jaime Feliu[2,7,17*]

[1] Biomedica Molecular Medicine SL, C/ Faraday 7, 28049, Madrid, Spain.

[2] Medical Oncology Department, Hospital Universitario La Paz, Paseo de la Castellana 261, 28046, Madrid, Spain

[3] Molecular Oncology & Pathology Lab, Institute of Medical and Molecular Genetics-INGEMM, Hospital Universitario La Paz -IdiPAZ, Paseo de la Castellana 261, 28046, Madrid, Spain.

[4] 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

[5] Pathology Department, Hospital Universitario La Paz, Paseo de la Castellana 261, 28046, Madrid, Spain

[6] Molecular Pathology and Therapeutic Targets Group, Hospital Universitario La Paz-IdiPAZ, Paseo de la Castellana 261, 28046, Madrid, Spain.

[7] Biomedical Research Networking Center on Oncology-CIBERONC, ISCIII, Av. Monforte de Lemos 5, 28029, Madrid, Spain

[8] Translational Oncology Lab, Hospital Universitario La Paz -IdiPAZ, Paseo de la Castellana 261, 28046, Madrid, Spain.

[9] Pathology Department, Hospital Clínic Universitari de Barcelona, Carrer de Villarroel 170, 08036, Barcelona, Spain

[10] Medical Oncology Department, Hospital General Universitario Gregorio Marañón, /Dr. Esquerdo 46, 28007, Madrid, Spain.

[11] Medical Oncology Service, Vall Hebron University Hospital. Vall Hebron Institute of Oncology (VHIO), Paseigg de la Vall d'Hebron 119, 08035, Barcelona, Spain.

[12] Radiotherapy Oncology Department, Hospital Clínic Universitari de Barcelona, Carrer de Villarroel 170, 08036, Barcelona, Spain.

[13] Medical Oncology Service, Hospital Universitario 12 de Ocubre, Av. de Córdoba s/n, 28041, Madrid, Spain.

[14] Genomics Unit Cantoblanco, Parque Científico de Madrid, C/ Faraday 7, 28049, Madrid, Spain.

[15] Functional Genomics Center Zurich, University of Zurich/ETH Zurich, Winterthurerstrasse 190, CH-8057, Zurich, Switzerland

[16] Biotechvana SL, Parque Científico de Madrid, C/ Faraday 7, 28049, Madrid, Spain.

[17]Cátedra UAM-Amgen, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049, Madrid, Spain.

[#] LT-F and IG contributed equally to this work.

[*] Corresponding author: Jaime Feliu, Hospital Universitario La Paz, Paseo de la Castellana 261, 28046, Madrid, Spain. Phone: +34 912071138 , jaimefeliu@hotmail.com

45

ABSTRACT

Background: Anal squamous cell carcinoma is a rare tumor. Chemo-radiotherapy yields a 50% 3-year relapse-free survival rate in advanced anal cancer, so improved predictive markers and therapeutic options are needed.

Methods: High-throughput proteomics and whole-exome sequencing were performed in 46 paraffin samples from anal squamous cell carcinoma patients. Hierarchical clustering was used to establish groups *de novo*. Then, probabilistic graphical models were used to study the differences between groups of patients at the biological process level.

Results: A molecular classification into two groups of patients was established, one group with increased expression of proteins related to adhesion, T lymphocytes and glycolysis; and the other group with increased expression of proteins related to translation and ribosomes. The probabilistic graphical model showed that these two groups presented differences in metabolism, mitochondria, translation, splicing and adhesion processes. Additionally, these groups showed different frequencies of genetic variants in some genes, such as *ATM, SLFN11* and *DST*. Finally, genetic and proteomic characteristics of these groups suggested the use of some possible targeted therapies, such as PARP inhibitors or immunotherapy.

Conclusions: In this study, a molecular classification of anal squamous cell carcinoma using high-throughput proteomics and whole-exome sequencing data was proposed. Moreover, differences between the two established groups suggested some possible therapies.

Keywords: anal squamous cell carcinoma, whole-exome sequencing, proteomics, molecular classification, personalized medicine.

Background

Anal squamous cell carcinoma (ASCC) is a relatively rare cancer. In the United States there are 8,000 new estimated cases per year, accounting for approximately 2.7% of all gastrointestinal cancers. Of these, more than 1,000 cases will result in death(1).

Since the 1970s, the standard treatment has consisted of the combination of 5-fluorouracil (5FU), mitomycin C or cisplatin and radiotherapy (2). This treatment is particularly effective in T1/T2 tumors, achieving complete regression in 80-90% of cases. However, in advanced anal cancers (T3-4 or N+), the disease-free survival (DFS) rate is only around 50% (3). Therefore, due to the lack of advances in the last years, new therapeutic strategies are needed to improve these outcomes.

Whole-exome sequencing (WES) focused on the identification of disease-causing genes is now being implemented into clinical practice (4). The first work that announced entire exome sequencing was published by Ng et al (5). Since then, personalized medicine has focused on identifying the cause of rare diseases and cancers.

With the recent improvements in mass-spectrometry (MS) techniques, high-throughput proteomics has made it possible to identify thousands of proteins (6). Proteins are the effectors of biological processes, being closer to the phenotype than genes or transcripts. On the other hand, probabilistic graphical models (PGMs) were successfully used in previous studies to characterize tumors from a functional perspective (7-9). Moreover, when used in combination, proteomics and genomics provide complementary information.

Previous studies in ASCC were focused in the characterization of genetic variants in this disease using next-generation sequencing techniques. The most frequent mutated genes, such as *PIK3CA*, *FBXW7*, *FAT1* or *ATM,* were characterized (10-13). On the other hand, Herfs et al. used MS proteomics in microdissected anal samples to establish differential protein expression

92 patterns depending of the location (squamous or transitional) (14). However, until date, a
93 molecular classification of ASCC has not been established

94 In this study, we combined WES with high-throughput proteomics to further characterize a
95 cohort of 46 ASCC tumors. This is the first time that a combined study of these characteristics
96 in ASCC has been done. Genomics provides information about the genetic causes of disease
97 and proteins are the ultimate effectors of biological processes. Therefore, a study of these two
98 –omics allows us to obtain a broader picture of the molecular features of ASCC tumors.

99

Methods

Patient cohort

Forty-six formalin-fixed, paraffin-embedded (FFPE) samples from patients diagnosed with ASCC were analyzed by WES and MS proteomics. The study was approved by the Ethical Committee of Hospital Universitario La Paz. Informed consent was obtained for all patients in the study. Samples were reviewed by an experienced pathologist and all the samples included at least 70% invasive tumor cells. 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 (ECOG-PS) of 0 to 2; have received no prior radiotherapy or chemotherapy for this malignancy and present with no metastasis. Demographic information related to the tumor and the treatments was collected. Human papilloma virus (HPV) infection was determined by CLART® HPV2 (Genomica).

DNA isolation

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

Protein isolation

Proteins were extracted from FFPE samples as previously described (15). Briefly, FFPE sections were deparaffinized in xylene and washed twice with absolute ethanol. Protein extracts were prepared in a 2% SDS buffer by a protocol based on heat-induced antigen retrieval. Protein concentration was measured using the MicroBCA Protein Assay Kit (Pierce-Thermo Scientific). Protein extracts (10 µg) were digested with trypsin (1:50) and SDS was removed from the digested lysates using Detergent Removal Spin Columns (Pierce). Peptides were desalted using

123 self-packed C18 stage tips, dried and resolubilized with 10 µl of 3% acetonitrile, 0.1% formic
124 acid.

125 *Whole-exome sequencing experiments*

126 WES from 46 ASCC FFPE samples was performed. The isolated DNA was evaluated by
127 Picogreen and mean size was controlled by gel electrophoresis. Genomic DNA was divided by
128 mechanical methods (Bioruptor) to a mean size of approximately 200 bp. At that point, DNA
129 tests were fixed, phosphorylated, A-followed and ligated to explicit connectors, trailed by PCR-
130 interceded naming with Illumina-explicit successions and test explicit standardized
131 identifications (Kapa DNA library age unit).

132 Exome capture was performed utilizing the VCRome framework (catch size of 37 Mb,
133 Nimblegen) under a multiplexing of 8 tests for every capture response. Capture was performed
134 entirely in accordance to Nimblegen enhancement instructions. After the capture, libraries
135 were cleansed, measured and titrated using Real Time PCR before sequencing. Tests were then
136 sequenced to a surmised inclusion of 4.5 Gb per test in Illumina-NextSeq NS500 (Illumina Inc.)
137 utilizing 150 cycles (2x75) High Output cartridges.

138 Raw data files were available in Sequence Read Archive (SRA,
139 https://www.ncbi.nlm.nih.gov/sra) under the name PRJNA573670.

140 *Bioinformatics analyses of whole-exome sequencing data*

141 The quality of the WES experiments was verified using FASTQC (http://www.bioinformmatics.
142 babraham.ac.uk/projects/fastqc). First, primers were removed using Cutadapt. Then, FASTQ
143 files were filtered by quality using PrinSeq. Both tools are available in GPRO tool (16).
144 Sequence alignment was performed using the human genome h19 as the reference and BWA
145 tools (17), Samtools (18) and Picard Tools (http://picard.sourceforge.net). Variant calling was

performed using the MuTect tool from the GATK4 package (19) combined with PicardTools, first, to create a panel of normal samples (PON) and second, to perform the variant calling (20). The PON was built using 11 samples from Iberian exomes from a 1000genomes database (http://www.ncbi.nlm.nih.gov/sra/).

Finally, variants were annotated using Variant Effect Predictor (VEP) (21). The information about the genetic variants provided by VEP was used to filter the genetic variants. The filtering criteria were: a frequency in the general population, according gnomAD database, of less than 1%, a high or moderate impact, and the presence of a variant of this gene in our cohort in at least 10% of the patients.

Liquid chromatography-mass spectrometry analysis

MS analysis was performed using a Q Exactive HF-X mass spectrometer (Thermo Scientific) equipped with a Digital PicoView source (New Objective) and coupled to a M-Class UPLC (Waters). Solvent composition at the two channels was 0.1% formic acid for channel A and 0.1% formic acid, 99.9% acetonitrile for channel B. For each sample 3 µL of peptides were loaded on a commercial MZ Symmetry C18 Trap Column (100Å, 5 µm, 180 µm x 20 mm, Waters) followed by nanoEase MZ C18 HSS T3 Column (100Å, 1.8 µm, 75 µm x 250 mm, Waters). The peptides were eluted at a flow rate of 300 nL/min by a gradient from 8 to 27% B in 85 min, 35% B in 5 min and 80% B in 1 min. Samples were acquired in a randomized order. The mass spectrometer was operated in data-dependent acquisition mode (DDA), acquiring full-scan MS spectra (350–1'400 m/z) at a resolution of 120'000 at 200 m/z after accumulation to a target value of 3'000'000, followed by HCD (higher-energy collision dissociation) fragmentation on the twenty most intense signals per cycle. HCD spectra were acquired at a resolution of 15'000 using normalized collision energy of 28 and a maximum injection time of 22 ms. The automatic gain control (AGC) was set to 100'000 ions. Charge state screening was enabled. Singly, unassigned, and charge states higher than seven were rejected. Only those

171 precursors with an intensity above 110'000 were selected for MS/MS. Precursor masses
172 previously selected for MS/MS measurement were excluded from further selection for 30 s,
173 and the exclusion window was set at 10 ppm. The samples were acquired using internal lock
174 mass (22) calibration on m/z 371.1012 and 445.1200.

175 The MS proteomics results were handled using the local laboratory information management
176 system (LIMS) (22) and all relevant data have been deposited to Chorus under the project
177 name "Anal squamous cell carcinoma proteomics".

178 *Protein identification and label-free protein quantification*

179 The acquired raw MS data was processed by MaxQuant (version 1.6.2.3), followed by protein
180 identification using the integrated Andromeda search engine (23). Spectra were searched
181 against a Uniprot reference proteome (taxonomy 9606, canonical version from 2016-12-09),
182 concatenated to its reversed decoyed fasta database and common protein contaminants.
183 Methionine oxidation and N-terminal protein acetylation were set as variable modifications.
184 Enzyme specificity was set to trypsin/P allowing for a minimal peptide length of 7 amino acids
185 and a maximum of two missed-cleavages. MaxQuant Orbitrap default search settings were
186 used. The maximum false discovery rate (FDR) was set to 0.01 for peptides and 0.05 for
187 proteins. Label-free quantification was enabled and a 2 minutes window for match between
188 runs was applied. In the MaxQuant experimental design template, each file is kept separate in
189 the experimental design to obtain individual quantitative values.

190 As quality criteria, the detectable measurement in at least 75% of the samples and the
191 presence of two unique peptides were applied. Log2 of the data was calculated and missing
192 values were imputed to a normal distribution using Perseus software (24).

193 *Probabilistic graphical models and functional node activity*

194 With the aim of studying proteomics data from a functional point of view, PGMs compatible
195 with high-dimensional data, were used. *grapHD* (25) and R v3.2.5 were used to generate the
196 PGM from proteomics expression without any *a priori* information, based on correlation as
197 associative measurement. The network was built in two steps: first, the spanning tree with
198 maximum likelihood was found and, then, the edges was chosen based on the reduction of the
199 Bayesian Information Criteria (BIC) and the preservation of the decomposability of the graph
200 (26). The resulting network was analyzed to define a functional structure by gene ontology
201 analyses, as in previous works (7-9). Briefly, the branches of the network were analyzed by
202 gene ontology and a majority function for each branch was assigned, thereby determining
203 different functional nodes in the network. Gene ontology analyses were performed using
204 DAVID 6.8 webtool (27) using "homo sapiens" as background and GOTERM-FAT, Biocarta and
205 KEGG as categories.

206 Once each branch had been assigned a function, functional node activities were calculated as
207 the mean of the proteins of each branch related to the main function of that branch (8, 9).
208 Then, comparisons between groups using Mann-Whitney test were done.

209 *Metabolic modeling and estimation of tumor growth rate*

210 Flux Balance Analysis (FBA) is a method used to model the flow of metabolites through
211 biochemical networks (28). It allows the growth rate or production rate of a given metabolite
212 to be estimated using as input gene or protein expression data. In this study we used the
213 whole human reconstruction Recon2 and the biomass reaction included in this model as the
214 objective function and as representative of tumor growth (29). Proteomics data was
215 introduced into the model to make accurate predictions by solving Gene-Protein-Reaction
216 rules (GPRs), which contain the relationships between genes and enzymes, using a modified
217 algorithm of Barker et al. (30) and a modified E-flux (7, 31). FBA calculations were performed
218 using the COBRA Toolbox library, available for MATLAB (32).

219 *Statistical analyses*

220 Statistical analyses were performed in GraphPad Prism 6 and SPSS IBM Statistics 20. Network
221 analyses were performed using Cytoscape software. Hierarchical cluster and Significance
222 Analysis of Microarrays (SAM) were performed using MeV software. SAM analysis allows the
223 identification of differential proteins between groups by a t-test corrected by permutations
224 over the number of samples. The significance was determined using the False Discovery Rate
225 (FDR) (33). The Genomics of Drug Sensitivity in Cancer database
226 (https://www.cancerrxgene.org/) was used to find possible therapeutic targets. P-values are
227 two-sided and considered statistically significant under 0.05.

228

Results

Patient cohort

Forty-six patients diagnosed with non-metastatic ASCC were recruited for this study. Twenty-eight patients came from the VITAL clinical trial (GEMCAD-09-02, NCT01285778), treated with panitumumab, 5FU, Mitomycin C, and radiotherapy. The other 18 patients were included from the routine clinical practice at Hospital Universitario La Paz and Hospital Clinic and were treated with cisplatin-5FU or Mitomycin C-5FU, and concomitant radiotherapy.

For the survival analyses, 4 patients that could not receive chemo-radiotherapy were excluded (two of them had stage I anal carcinomas, and the other two had stage III tumors). The median follow-up was 33.18 months (5.53-116.4) and there are 13 relapse events. All clinical characteristics are shown in Table 1.

Whole-exome sequencing experiments

Forty-six FFPE samples were analyzed by WES. The mean coverage of the samples was 42.6x, with the exception of one sample that presented a coverage of 3.57x. This sample was dismissed from the subsequent analyses. Once this sample was dismissed, all the samples presented a mapping efficiency of 90-98%, with the exception of one sample (with a mapping efficiency of 75.4%). Human exome has > 195,238 exonic regions, of which only 23,021 (11.21% of the human exome) have not been mapped in any sample.

After VEP analysis and filtering, 382 genes that presented a genetic variant with high or moderate impact in at least 10% of our cohort were identified (Sup Table 1). These genes were mostly related to DNA repair, chromatin binding and focal adhesion processes. *PIK3CA* was mutated in the 40% of the patients of our cohort, *FBXW7* in 16%, *FAT1* in 18%, and *ATM* was mutated in 27% of the patients. Figure 1 summarizes the high and moderate impact alteration landscape in ASCC.

Proteomics experiments

After dismissing one sample in the WES experiments, 45 FFPE samples were analyzed by MS and 6,035 proteins were identified. After applying quality criteria (detectable measurement in at least 75% of the samples and at least two unique peptides), 1,954 proteins were used for the subsequent analyses.

De novo identification of groups based on differential proteomics profiles

With the aim of defining *de novo* molecular groups of patients, a hierarchical cluster was used. Two different molecular groups were obtained. A SAM was performed to define the differential proteins between these two groups, yielding 318 proteins which were differentially expressed between these groups (Sup Table 2). Group 1 underexpressed proteins related to translation and ribosomal processes and overexpressed proteins related to metabolism, specially glycolysis, T lymphocytes, and adhesion. On the other hand, Group 2 underexpressed proteins related to metabolism, T lymphocytes, and adhesion processes and overexpressed proteins related to translation and ribosomes (Figure 2).

With respect to the clinical data distribution between these two groups, both were comparable; there were no significant differences in the distribution of clinical parameters (Sup Table 3). In addition, there were not significant differences in disease-free survival or overall survival (Sup Fig 1).

A search in the Genomics of Drug Sensitivity in Cancer database (https://www.cancerrxgene.org/) suggested RAC1 (overexpressed in Group1 and underexpressed in Group 2) as a possible therapeutic target. The drug associated with this gene is EHT-1864.

Functional characterization of proteomics data

To study proteomics data from a functional perspective, a PGM was created using the 1,954 proteins obtaining from the MS experiments. The resulting network was divided into 10 functional nodes, one of them with an overrepresentation of two biological functions (metabolism and mitochondria) (Figure 3).

Then, functional node activities were used, as in previous works (7-9), to study the differences in biological processes between the two identified groups of patients. There were significant differences between the two groups in membrane category, in the two nodes related to metabolism, and in the nodes associated with adhesion, ribosomes, translation, extracellular matrix and splicing (Figure 4).

Metabolism nodes

We found two different nodes related to metabolism, both of which showed a higher expression in Group 1. The metabolism 1 node was formed by 158 proteins, mostly related to mitochondrial metabolism, especially the tricarboxylic acid cycle. Among them, three were also identified as differentially expressed by the SAM analysis: P09622 (DLD), P06744 (GPI) and P14550 (AKR1A1). The metabolism 2 node included 104 proteins related to mitochondria and metabolism, especially oxidative phosphorylation, such as P04406 (GAPDH), P06733 (ENO1), P07954 (FH) or Q9UI09 (NDUFA12).

Adhesion node

The adhesion node showed a higher expression in Group 1. The adhesion node included 654 proteins, 25 of them classified by SAM as differentially expressed between the two groups of patients.

Genetic variants with different frequencies between the two groups of patients established by proteomics

299 Using a Chi-squared test, the frequencies of genetic variants for each gene were compared to
300 determine whether the groups of patients defined using proteomics data also showed
301 differences in genetic variants. Twelve genes presented different frequencies of genetic
302 variants between the two groups (Figure 5).

303 *Tumor growth rate predicted by metabolic modeling*

304 FBA allows for the comparison of the tumor growth rate between groups of tumors. The tumor
305 growth rate predicted for Group 1 was significantly higher than the tumor growth rate
306 predicted for Group 2 (Figure 6).

307

Discussion

ASCC is an infrequent tumor. With no targeted therapy yet established, the molecular characterization of these tumors is still necessary. In this study, we combined the two main –omics, WES and proteomics, to further characterize a cohort of 46 patients diagnosed with primary ASCC. To our knowledge, this is the first study combining WES and proteomics in ASCC. The results of this study allow us to establish two molecular subgroups in ASCC with different molecular features. Moreover, the analyses of these two groups pointed out some drug-susceptible processes, such as metabolism, and suggested other possible therapeutic targets, like *ATM* and its relationship with PARP inhibitors (PARPi).

Previous studies have analyzed both primary and metastatic ASCC paraffin samples using WES or gene panels (10-13). A previous study using proteomics data to characterize different locations of anal cancer was also existed (14). However, this is the first study that performs proteomics experiments in localized ASCC samples and combines proteomics data with WES information. Previous WES studies served to identify frequently-occurring mutations in this disease, including mutations in the *PIK3CA, FBXW7*, *FAT1* and *ATM* genes (11-13). In our cohort, *PIK3CA* presented a genetic variant with a high or modifier impact in 40% of the patients, *FBXW7* in 16%, *FAT1* in 18%, and *ATM* in 27% of the patients.

Using proteomics data and HCL, it was possible to establish two different molecular groups of patients. Differential proteins were mainly related to the metabolism of glucose, translation and ribosomes, T lymphocytes and adhesion. Although these molecular groups have not been associated with any clinical or prognostic features, these processes may be relevant in the development of new therapeutic strategies. For instance, those tumors that overexpressed proteins related to glycolysis may be candidates for drugs targeting metabolism such as metformin, which has been shown to have cytostatic effects on other tumors, such as breast or bladder carcinoma (7). On the other hand, one of the main differences between these two

groups is that Group 1 had a higher expression of proteins related to T lymphocytes. With the bloom of immunotherapies, immune proteins have acquired great relevance. Therefore, this group of patients may be good candidates for immunotherapy. In fact, nivolumab has been reported to be an effective therapy in metastatic ASCC and its efficacy is related to the presence of cytotoxic T cells (34). Moreover, pembrolizumab has demonstrated its antitumor activity in PD-L1-positive advanced ASCC (35).On the other hand, the search in Genomics of Drug Sensitivity in Cancer database to establish possible therapeutic targets suggested RAC1 (overexpressed in Group 1 and underexpressed in Group 2) as a potential therapeutic target. RAC1 has as associated drug, EHT-1864, which affects the cytoskeleton (36).

In addition, MS experiments and PGMs allow for the functional characterization of these two groups of patients. In this functional analysis, differences in metabolism were confirmed. There were also differences at the mitochondria level, which is the target for metformin. Metabolism nodes showed a higher expression in Group 1. P06744 (GPI, glucose 6-phosphate isomerase), included in the metabolism 1 node, is the enzyme that converts glucose – phosphate into fructose 6-phosphate and a higher expression has been associated with tumorigenesis and poor prognosis in gastric cancer (37).

The adhesion node also had a higher expression in Group 1 and contained 25 proteins identified by SAM as differentially expressed. P46940 (IQGAP1) has been associated with poor prognosis in head and neck squamous cell carcinoma (38) and has also been associated with response to chemo-radiotherapy in rectal adenocarcinomas (39). P27797 (CALR) induces an immune response in esophageal squamous cell carcinoma (40). P08133 (ANXA6) promotes EGFR deactivation (41). P62829 (RPL23) negatively regulates apoptosis and also inhibits growth in colorectal cancer (42). On the other hand, RPL23 has been identified as an oncogene in head and neck squamous cell carcinoma (43). O43707 (ACTN4) increases cell motility and invasion in colorectal cancer (44). Previous studies have described how P84077 (ARF1) forms a complex

with EGFR and promotes invasion in head and neck squamous cell carcinoma (45). P35579 (MYH9) plays an important role in adhesion and migration, and its overexpression is correlated with metastasis in colorectal cancer through the MAPK pathway (46) . Aberrant activity of P63000 (RAC1), which is involved in metastasis and proliferation, is a hallmark in cancer, (47). At the same time, O75131 (CPNE3) promotes cell migration through RAC1 (48). Finally, P61586 (RHOA), a tumor suppressor gene, plays a relevant role in colorectal cancer, being associated with metastasis and is deactivated in a significant number of colorectal tumors (49).In conclusion, the majority of the proteins included in this adhesion node play well-established roles in metastasis processes.

Moreover, the combination of the proteomics and genetic variants information showed that the two molecular groups defined by proteomics also had a different mutational profile. Group 2 showed a higher frequency of *ATM* genetic variants. Previous studies have described a high response rate to PARPi, as olaparib, in prostate tumors with mutations in *ATM* (50). Therefore, Group 2 patients may also be candidates for the treatment with PARPi.

In addition, FBA predicted a higher tumor growth rate for Group 1 than for Group 2. It may be possible that, given their higher proliferation, the tumors of Group 1 may also be more responsive to chemotherapy.

This study has some limitations. The results need to be validated in an independent cohort. The information in ASCC is scarce so a prospective validation will be needed. The number of proteins detected by MS still needs technical improvement to be considered to be at the same level as genomics. However, proteomics offers a more direct measurement of the effectors of biological processes. Finally, a consensus analysis pipeline to apply in cancer sequencing data is still necessary.

Conclusions

In conclusion, two different molecular groups of patients have been proposed based on proteomics expression. This may be the first step toward a personalized therapy approach in ASCC. In addition, some possible targeted therapies, such as PARPi or immunotherapy, according to the molecular features (genetic and protein-based) defined in the two proteomics groups were suggested.

List of abbreviations

ASCC: Anal squamous cell carcinoma

5FU: 5-fluorouracil

DFS: disease-free survival

WES: whole-exome sequencing

MS: mass-spectrometry

PGM: probabilistic graphical model

FFPE: formalin-fixed paraffin-embedded

ECOG-PS: Eastern Cooperative Oncology Group performance status

HPV: Human papilloma virus

VEP: Variant Effect Predictor

DDA: Data-dependent acquisition

FDR: False Discovery Rate

FBA: Flux Balance Analysis

BIC: Bayesian Information Criterion

GPR: Gene-Protein-Reaction rules

SAM: Significance Analysis of Microarrays

PARPi: PARP inhibitors

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.

Consent for publication: Not applicable.

Availability of data and material: The proteomics data generated during the current study are available in Chorus repository (https://chorusproject.org/pages/index.html). The exome sequencing data are available in SRA https://www.ncbi.nlm.nih.gov/sra) under the name PRJNA573670.

Competing interests: JAFV and AG-P are shareholders in Biomedica Molecular Medicine SL. LT-F and GP-V are employees of Biomedica Molecular Medicine SL. JC has received honoraria for scientific consulting (as 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. IG has received honoraria and/or travel expenses from Roche, Sanofi, Merck, Servier, Amgen and Sirtflex, and for advisory role from Merck and Sanofi. JF has received consulting and advisory honoraria from Amgen, Ipsen, Eissai, Merck, Roche and Novartis; research funding from Merck, and travel and accommodation expenses from Amgen and Servier. The other authors declare no conflicts of interest.

Funding: This study was supported by the Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry, Spain and co-sponsored by the FEDER program, "Una forma de hacer Europa" (PI15/01310), a Roche Farma grant, Cátedra UAM-Amgen and a grant of Grupo Español Multidisciplinar en Cáncer Digestivo (GEMCAD1403). LT-F is supported by the Spanish Economy and Competitiveness Ministry (DI-15-07614). GP-V is supported by the Consejería de Educación, Juventud y Deporte of Comunidad de Madrid (IND2017/BMD7783); AZ-M is supported by Jesús Antolín Garciarena fellowship from IdiPAZ. The sponsors were not involved in the study design, in data collection and analysis, in the decision to publish or in the preparation of this manuscript.

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. RL-V, MM, EL-C, and VH contributed the DNA and protein extraction. MM and VH contributed the HPV determination. LG-P, CP, and MC contributed the pathological review of the samples. PN and CF contributed the mass spectrometry data. RR-R and CL contributed the sequencing analyses. LT-F, AG-P and JAFV contributed the probabilistic graphical models. LT-F contributed the FBA analyses. IG, JM, PG-A, JC, CC, RG-C, and JF contributed the clinical data and the analyses related. LT-F, IG, AG-P, GP-V, and AZ-M contributed in the design of the study and the statistical and gene ontology analyses. L-T-F drafted the manuscript. IG, JM, JAFV, and JF conceived of the study, and participated in its design and interpretation. AG-P, JM, JAFV, and JF reviewed the manuscript. JF coordinated the study. All authors read and approved the final manuscript.

448

449

450 **REFERENCES**

451 1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30.

452 2. Benson A, Venook A, Al-Hawary M, Cederquist L, Chen Y, Ciombor K, et al. Anal
453 carcinoma, Version 2.2018, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc
454 Netw. 2018. p. 852-71.

455 3. Gunderson LL, Winter KA, Ajani JA, Pedersen JE, Moughan J, Benson AB, et al. Long-
456 term update of US GI intergroup RTOG 98-11 phase III trial for anal carcinoma: survival,
457 relapse, and colostomy failure with concurrent chemoradiation involving
458 fluorouracil/mitomycin versus fluorouracil/cisplatin. J Clin Oncol. 2012;30(35):4344-51.

459 4. Tetreault M, Bareke E, Nadaf J, Alirezaie N, Majewski J. Whole-exome sequencing as a
460 diagnostic tool: current challenges and future opportunities. Expert Rev Mol Diagn.
461 2015;15(6):749-60.

462 5. Ng S, Turner E, Robertson P, Flygare S, Bigham A, Lee C, et al. Targeted capture and
463 massively parallel sequencing of 12 human exomes. Nature. 2009;461 (7261):272-6.

464 6. Meier F, Geyer PE, Virreira Winter S, Cox J, Mann M. BoxCar acquisition method
465 enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. Nat Methods.
466 2018.

467 7. Trilla-Fuertes L, Gámez-Pozo A, Arevalillo JM, Díaz-Almirón M, Prado-Vázquez G,
468 Zapater-Moros A, et al. Molecular characterization of breast cancer cell response to metabolic
469 drugs. Oncotarget. 2018;9(11):9645-60.

470 8. Gámez-Pozo A, Trilla-Fuertes L, Berges-Soria J, Selevsek N, López-Vacas R, Díaz-Almirón
471 M, et al. Functional proteomics outlines the complexity of breast cancer molecular subtypes.
472 Scientific Reports. 2017;7(1):10100.

473 9. Gámez-Pozo A, Berges-Soria J, Arevalillo JM, Nanni P, López-Vacas R, Navarro H, et al.
474 Combined label-free quantitative proteomics and microRNA expression analysis of breast
475 cancer unravel molecular differences with clinical implications. Cancer Res; 2015. p. 2243-53.
476 10. Cacheux W, Rouleau E, Briaux A, Tsantoulis P, Mariani P, Richard-Molard M, et al.
477 Mutational analysis of anal cancers demonstrates frequent PIK3CA mutations associated with
478 poor outcome after salvage abdominoperineal resection. Br J Cancer. 2016;114(12):1387-94.
479 11. Cacheux W, Dangles-Marie V, Rouleau E, Lazartigues J, Girard E, Briaux A, et al. Exome
480 sequencing reveals aberrant signalling pathways as hallmark of treatment-naive anal
481 squamous cell carcinoma. Oncotarget. 2018;9(1):464-76.
482 12. Morris V, Rao X, Pickering C, Foo WC, Rashid A, Eterovic K, et al. Comprehensive
483 Genomic Profiling of Metastatic Squamous Cell Carcinoma of the Anal Canal. Mol Cancer Res.
484 2017;15(11):1542-50.
485 13. Chung JH, Sanford E, Johnson A, Klempner SJ, Schrock AB, Palma NA, et al.
486 Comprehensive genomic profiling of anal squamous cell carcinoma reveals distinct genomically
487 defined classes. Ann Oncol. 2016;27(7):1336-41.
488 14. Herfs M, Longuespée R, Quick CM, Roncarati P, Suarez-Carmona M, Hubert P, et al.
489 Proteomic signatures reveal a dualistic and clinically relevant classification of anal canal
490 carcinoma. J Pathol. 2017;241(4):522-33.
491 15. Gámez-Pozo A, Ferrer NI, Ciruelos E, López-Vacas R, Martínez FG, Espinosa E, et al.
492 Shotgun proteomics of archival triple-negative breast cancer samples. Proteomics Clin Appl.
493 2013;7(3-4):283-91.
494 16. Futami R, Muñoz-Pomer A, Viu J, Domínguez-Escribá R, Covelli L, Bernet G, et al. GPRO
495 The professional tool for annotation, management and functional analysis of omic databases.
496 Biotechvana Bioinformatics: SOFT3. 2011.
497 17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
498 transform. Bioinformatics. 2009;25(14):1754-60.

499 18. Li H, Handsaker B, Wysoker B, Fennell T, Ruan J, Homer N, et al. The Sequence
500 Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078-9.
501 19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
502 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
503 sequencing data. Genome Research. 2010;20(9):1297-303.
504 20. Cibulskis K, Lawrence M, Carter S, Sivachenco A, Jaffe D, Sougnez C, et al. Sensitive
505 detection of somatic point mutations in impure and heterogeneous cancer samples. Nature
506 Biotechnology. 2013;31(3):213-9.
507 21. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015.
508 Nucleic Acids Res. 2015;43(Database issue):D662-9.
509 22. Turker CA, F, Joho D, Panse B, Oesterreicher B, Rehrauer H, Schlapbach R. B-Fabric: The
510 Swiss Army Knife for Life Sciences. Lausanne, Switzerland EDBT; 2010.
511 23. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized
512 p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol.
513 2008;26(12):1367-72.
514 24. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus
515 computational platform for comprehensive analysis of (prote)omics data. Nat Methods.
516 2016;13(9):731-40.
517 25. Abreu G, Edwards D, Labouriau R. High-Dimensional Graphical Model Search with
518 the gRapHD R Package Journal of Statistical Software2010. p. 1-18.
519 26. Lauritzen S. Graphical Models. Oxford,UK.: Oxford University Press1996.
520 27. Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene
521 lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.
522 28. Orth J, Thiele I, Palsson B. What is flux balance analysis? : Nat Biotechnol; 2010. p. 245-
523 8.

524 29. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, et al. A community-
525 driven global reconstruction of human metabolism. Nat Biotechnol. 2013;31(5):419-25.
526 30. Barker BE, Sadagopan N, Wang Y, Smallbone K, Myers CR, Xi H, et al. A robust and
527 efficient method for estimating enzyme complex abundance and metabolic flux from
528 expression data. Comput Biol Chem. 2015;59 Pt B:98-112.
529 31. Colijn C, Brandes A, Zucker J, Lun D, Weiner B, Farhat M, et al. Interpreting expression
530 data with metabolic flux models: Predicting Mycobacterium tuberculosis mycolic acid
531 production. PLOS Comput Bio; 2009.
532 32. Schellenberger J, Que R, Fleming R, Thiele I, Orth J, Feist A, et al. Quantitative
533 prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.
534 Nature Protocols; 2011. p. 1290-307.
535 33. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the
536 ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98(9):5116-21.
537 34. Morris VK, Salem ME, Nimeiri H, Iqbal S, Singh P, Ciombor K, et al. Nivolumab for
538 previously treated unresectable metastatic anal cancer (NCI9673): a multicentre, single-arm,
539 phase 2 study. Lancet Oncol. 2017;18(4):446-53.
540 35. Ott PA, Piha-Paul SA, Munster P, Pishvaian MJ, van Brummelen EMJ, Cohen RB, et al.
541 Safety and antitumor activity of the anti-PD-1 antibody pembrolizumab in patients with
542 recurrent carcinoma of the anal canal. Ann Oncol. 2017;28(5):1036-41.
543 36. Onesto C, Shutes A, Picard V, Schweighoffer F, Der CJ. Characterization of EHT 1864, a
544 novel small molecule inhibitor of Rac family small GTPases. Methods Enzymol. 2008;439:111-
545 29.
546 37. Ma YT, Xing XF, Dong B, Cheng XJ, Guo T, Du H, et al. Higher autocrine motility
547 factor/glucose-6-phosphate isomerase expression is associated with tumorigenesis and poorer
548 prognosis in gastric cancer. Cancer Manag Res. 2018;10:4969-80.

549 38. Wu CC, Li H, Xiao Y, Yang LL, Chen L, Deng WW, et al. Over-expression of IQGAP1
550 indicates poor prognosis in head and neck squamous cell carcinoma. J Mol Histol.
551 2018;49(4):389-98.
552 39. Holck S, Nielsen HJ, Hammer E, Christensen IJ, Larsson LI. IQGAP1 in rectal
553 adenocarcinomas: localization and protein expression before and after radiochemotherapy.
554 Cancer Lett. 2015;356(2 Pt B):556-60.
555 40. Liu X, Song N, Liu Y, Li J, Ding J, Tong Z. Efficient induction of anti-tumor immune
556 response in esophageal squamous cell carcinoma via dendritic cells expressing MAGE-A3 and
557 CALR antigens. Cell Immunol. 2015;295(2):77-82.
558 41. Koese M, Rentero C, Kota BP, Hoque M, Cairns R, Wood P, et al. Annexin A6 is a
559 scaffold for PKCα to promote EGFR inactivation. Oncogene. 2013;32(23):2858-72.
560 42. Qi Y, Li X, Chang C, Xu F, He Q, Zhao Y, et al. Ribosomal protein L23 negatively regulates
561 cellular apoptosis via the RPL23/Miz-1/c-Myc circuit in higher-risk myelodysplastic syndrome.
562 Sci Rep. 2017;7(1):2323.
563 43. Russo N, Wang X, Liu M, Banerjee R, Goto M, Scanlon C, et al. A novel approach to
564 biomarker discovery in head and neck cancer using an autoantibody signature. Oncogene.
565 2013;32(42):5026-37.
566 44. Honda K, Yamada T, Hayashida Y, Idogawa M, Sato S, Hasegawa F, et al. Actinin-4
567 increases cell motility and promotes lymph node metastasis of colorectal cancer.
568 Gastroenterology. 2005;128(1):51-62.
569 45. He L, Gao L, Shay C, Lang L, Lv F, Teng Y. Histone deacetylase inhibitors suppress
570 aggressiveness of head and neck squamous cell carcinoma via histone acetylation-independent
571 blockade of the EGFR-Arf1 axis. J Exp Clin Cancer Res. 2019;38(1):84.
572 46. Wang B, Qi X, Liu J, Zhou R, Lin C, Shangguan J, et al. MYH9 Promotes Growth and
573 Metastasis via Activation of MAPK/AKT Signaling in Colorectal Cancer. J Cancer.
574 2019;10(4):874-84.

575 47. Kazanietz MG, Caloca MJ. The Rac GTPase in Cancer: From Old Concepts to New
576 Paradigms. Cancer Res. 2017;77(20):5445-51.
577 48. Ahmat Amin MKB, Shimizu A, Zankov DP, Sato A, Kurita S, Ito M, et al. Epithelial
578 membrane protein 1 promotes tumor metastasis by enhancing cell migration via copine-III and
579 Rac1. Oncogene. 2018;37(40):5416-34.
580 49. Dopeso H, Rodrigues P, Bilic J, Bazzocco S, Cartón-García F, Macaya I, et al.
581 Mechanisms of inactivation of the tumour suppressor gene RHOA in colorectal cancer. Br J
582 Cancer. 2018;118(1):106-16.
583 50. Mateo J, Carreira S, Sandhu S, Miranda S, Mossop H, Perez-Lopez R, et al. DNA-Repair
584 Defects and Olaparib in Metastatic Prostate Cancer. N Engl J Med. 2015;373(18):1697-708.

585

586

587 **TABLE AND FIGURE LEGENDS**

588 Figure 1: High and moderate impact genetic variants located in genes mutated in at least 20%
589 of the ASCC patients of this cohort.

590 Figure 2: Significance Analysis of Microarrays identified 318 differential proteins between two
591 groups of ASCC patients. Green=underexpressed. Red= overexpressed. In green, Group 1. In
592 blue, Group 2.

593 Figure 3: Functional network created using the proteomics data from the ASCC patients. Ten
594 nodes with different biological functions were identified.

595 Figure 4: Functional node activities between the two groups defined by the hierarchical
596 cluster. ****: $p<0.0001$; ***: $0.0001 \geq p \leq 0.001$; **: $0.01 \geq p \leq 0.05$. G1= Group 1; G2= Group 2.

597 Figure 5: Genes with different frequency of genetic variants between the two groups of
598 patients identified in proteomics. The mutation rate was calculated as the total number of
599 mutations of a gene in each group divided by the number of patients assigned to that group.

600 Figure 6: Tumor growth rate predicted using metabolic models. a.u. = arbitrary units. *: p
601 <0.05

602 Sup Fig 1: A. Disease-free survival curves for the two groups defined by proteomics. B. Overall
603 survival curves for the two groups defined by proteomics.

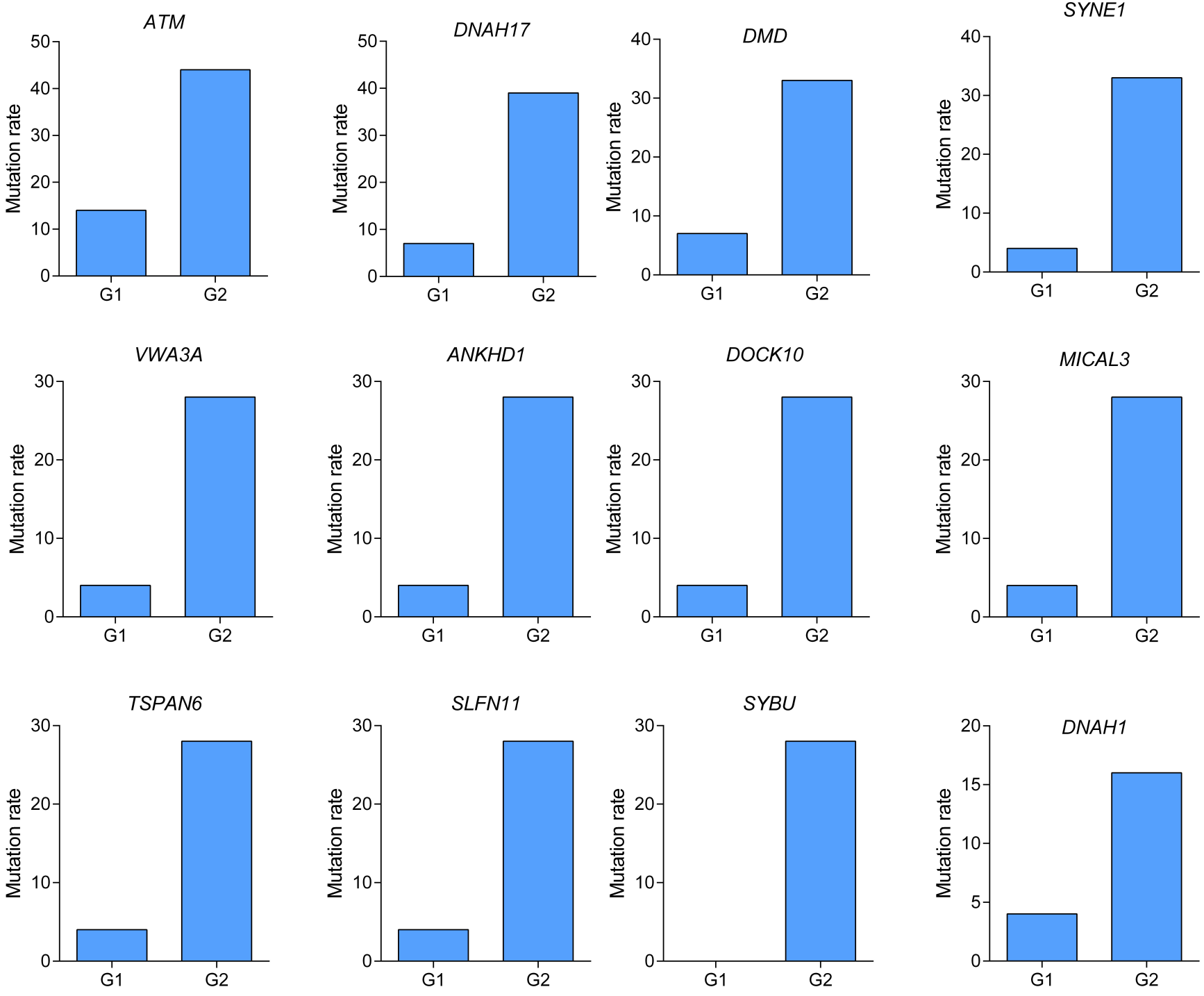
604 Table 1: Patient characteristics.

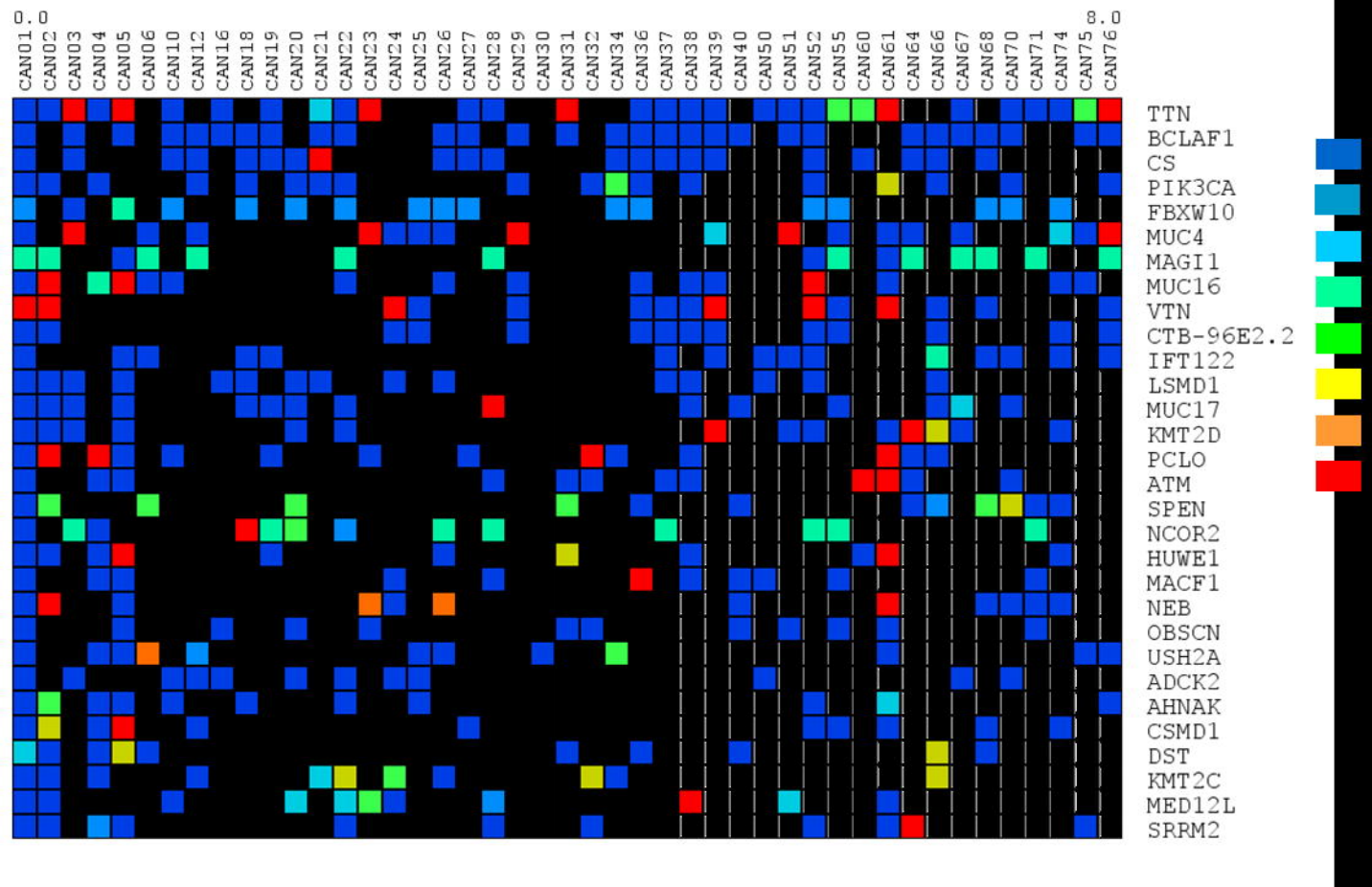
605 Sup Table 1: Genes with a detected genetic variant in at least 10% of the patients.

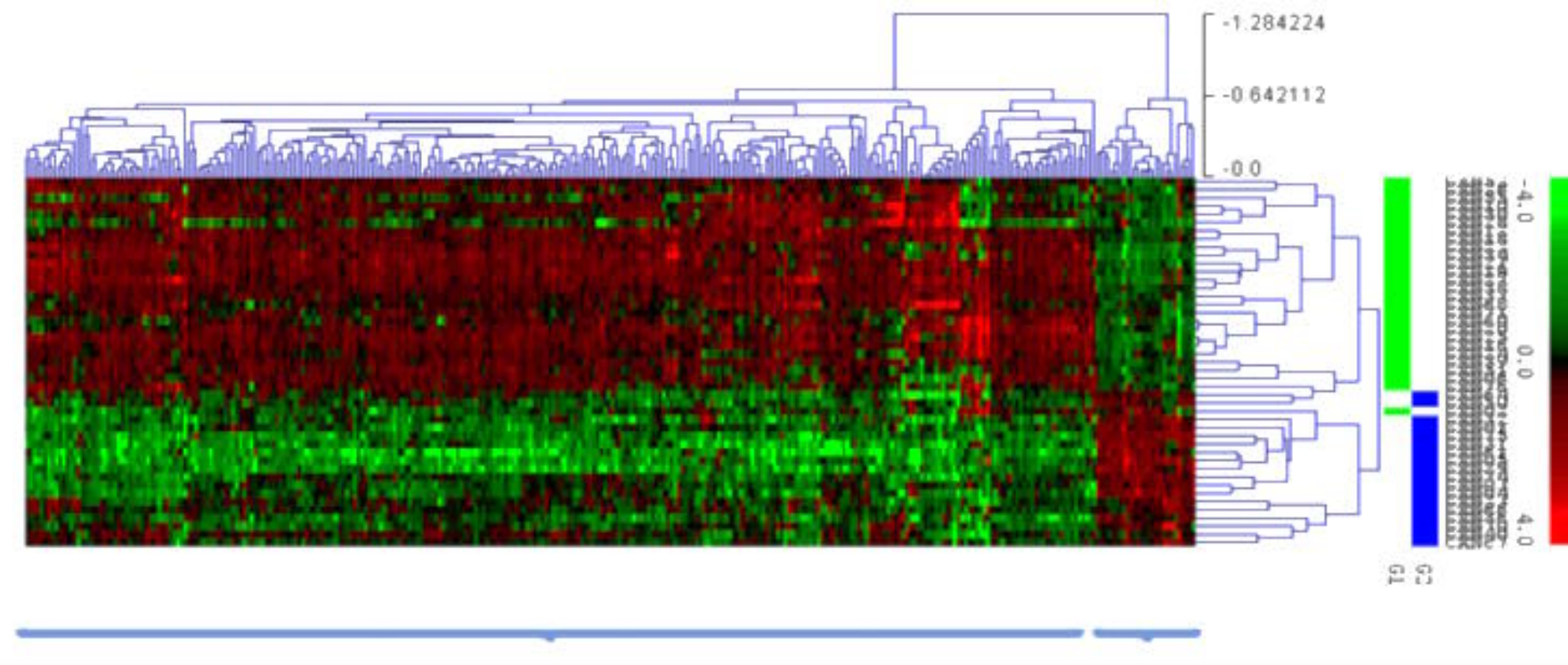
606 Sup Table 2: 318 differential proteins obtained by SAM between the two groups.

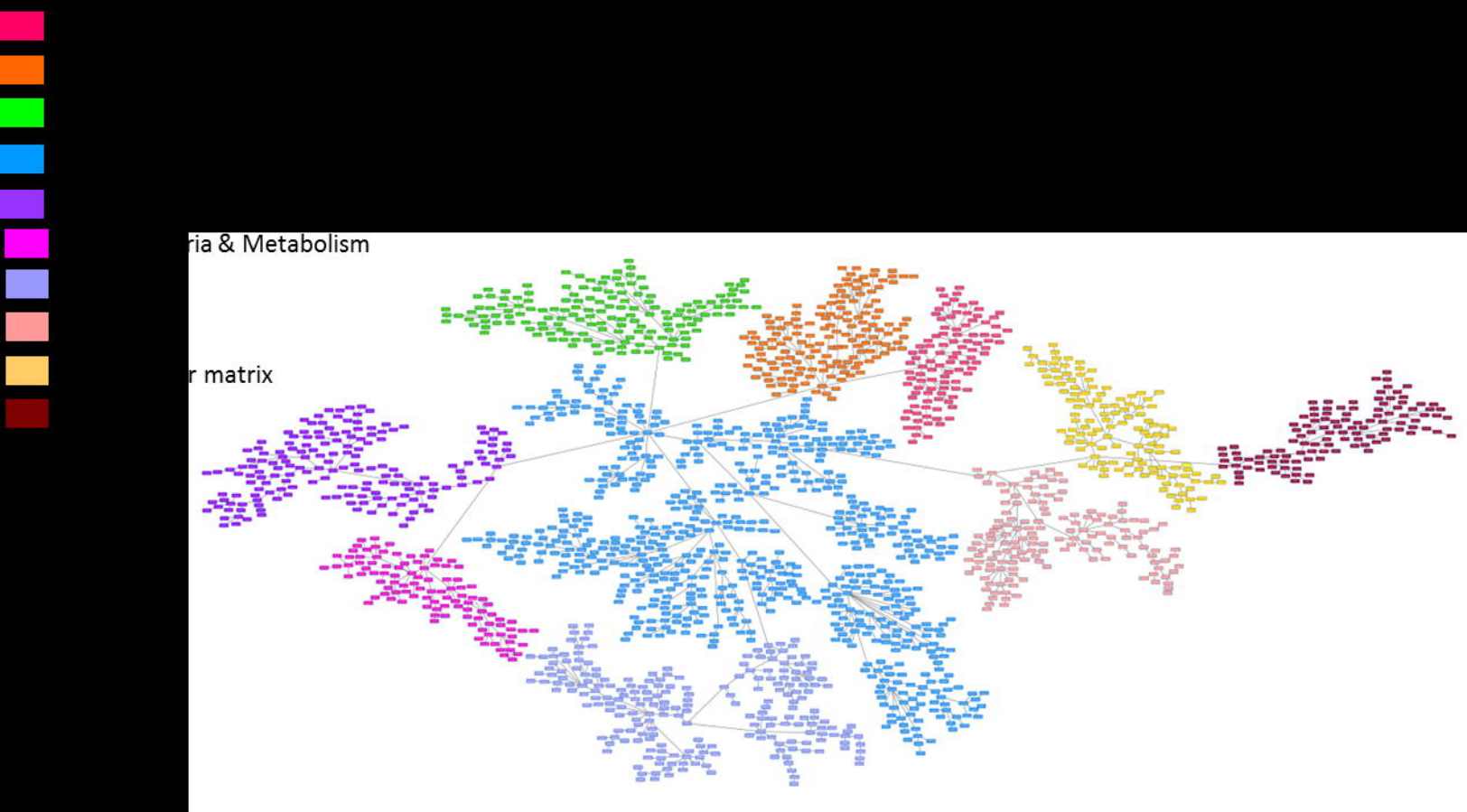
607 Sup Table 3: Clinical data distribution between the two molecular groups. p= p-values obtained
608 from a Chi-squared test comparing the two groups of patients.

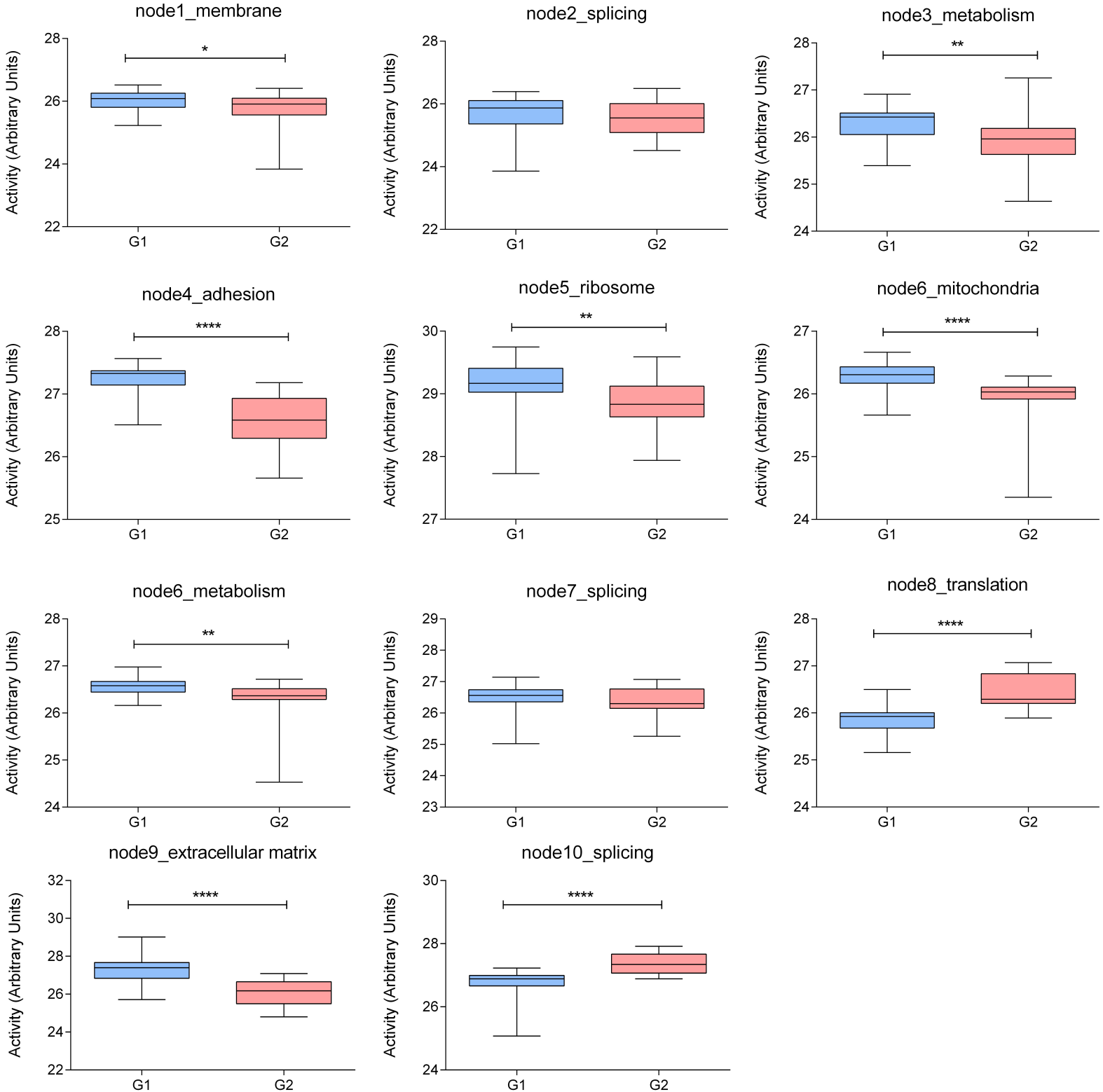
609


ATM
DNAH17
DMD
SYNE1
VWA3A
ANKHD1
DOCK10
MICAL3
TSPAN6
SLFN11
SYBU
DNAH1
Mutation rate
G1
G2


0.0
8.0
CAN01
CAN02
CAN03
CAN04
CAN05
CAN06
CAN10
CAN12
CAN16
CAN18
CAN19
CAN20
CAN21
CAN22
CAN23
CAN24
CAN25
CAN26
CAN27
CAN28
CAN29
CAN30
CAN31
CAN32
CAN34
CAN36
CAN37
CAN38
CAN39
CAN40
CAN50
CAN51
CAN52
CAN55
CAN60
CAN61
CAN64
CAN66
CAN67
CAN68
CAN70
CAN71
CAN74
CAN75
CAN76
TTN
BCLAF1
CS
PIK3CA
FBXW10
MUC4
MAGI1
MUC16
VTN
CTB-96E2.2
IFT122
LSMD1
MUC17
KMT2D
PCLO
ATM
SPEN
NCOR2
HUWE1
MACF1
NEB
OBSCN
USH2A
ADCK2
AHNAK
CSMD1
DST
KMT2C
MED12L
SRRM2


-1.284224
-0.642112
-0.0
-4.0
0.0
4.0
G2
G1


ria & Metabolism
r matrix


node1_membrane
Activity (Arbitrary Units)
28
26
24
22
*
G1
G2
node2_splicing
Activity (Arbitrary Units)
28
26
24
22
G1
G2
node3_metabolism
Activity (Arbitrary Units)
28
27
26
25
24
**
G1
G2
node4_adhesion
Activity (Arbitrary Units)
28
27
26
25
****
G1
G2
node5_ribosome
Activity (Arbitrary Units)
30
29
28
27
**
G1
G2
node6_mitochondria
Activity (Arbitrary Units)
27
26
25
24
****
G1
G2
node6_metabolism
Activity (Arbitrary Units)
28
27
26
25
24
**
G1
G2
node7_splicing
Activity (Arbitrary Units)
29
28
27
26
25
24
23
G1
G2
node8_translation
Activity (Arbitrary Units)
28
27
26
25
24
****
G1
G2
node9_extracellular matrix
Activity (Arbitrary Units)
32
30
28
26
24
22
****
G1
G2
node10_splicing
Activity (Arbitrary Units)
30
28
26
24
****
G1
G2