

Genome wide DNA methylation profiling identifies specific epigenetic features in high-risk cutaneous squamous cell carcinoma.

David Hervás-Marín^{1#}, Faatiemah Higgins^{2#}, Onofre Sanmartín^{3,4}, Jose Antonio López-Guerrero⁵, M.Carmen Bañó², J.Carlos Igual^{2*}, Inma Quilis^{2*}, Juan Sandoval^{6*}.

¹Department of Biostatistics, Instituto de Investigación Sanitaria La Fe, Valencia, Spain.

²Estructura de Recerca Interdisciplinar en Biotecnologia i Biomedicina (ERI BIOTECMED), Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot (Valencia), Spain.

³Dermatology Department, Instituto Valenciano de Oncología, Valencia, Spain.

⁴Facultad de Medicina, Universidad Católica de Valencia, Valencia, Spain.

⁵Laboratory of Molecular Biology, Instituto Valenciano de Oncología, Valencia, Spain.

⁶Biomarkers and Precision Medicine Unit (UByMP), Instituto de Investigación Sanitaria La Fe, Valencia, Spain.

These authors contributed equally to this work

* Corresponding authors:

J. Carlos Igual. Dept. Bioquímica i Biologia Molecular, Facultat de Ciències Biològiques, Universitat de València. C/ Dr. Moliner 50, 46100 Burjassot, Spain. Tel: (34)963543441. E-mail: jigual@uv.es

Inma Quilis. Dept. Bioquímica i Biologia Molecular, Facultat de Ciències Biològiques, Universitat de València. C/ Dr. Moliner 50, 46100 Burjassot, Spain. Tel: (34)963543795. E-mail: inmaculada.quilis@uv.es

Juan Sandoval. Biomarkers and Precision Medicine Unit (UByMP), Instituto de Investigación Sanitaria La Fe. Av. Fernando Abril Martorell, 106, 46026 Valencia, Spain. Tel: (34)961246709. E-mail: epigenomica@iislafe.es

Short title: Epigenetic features in high-risk cSCC

ABSTRACT

Cutaneous squamous cell carcinoma (cSCC) is the second most common skin cancer. Although most cSCCs have good prognosis, a subgroup of high-risk cSCC has a higher frequency of recurrence and mortality. Therefore, the identification of molecular risk factors associated with this aggressive subtype is of major interest. In this work we carried out a global-scale approach to investigate the DNA-methylation profile in patients at different stages, from premalignant actinic keratosis to low-risk invasive and high-risk non-metastatic and metastatic cSCC. The results showed massive non-sequential changes in DNA-methylome and identified a minimal methylation signature that discriminates between stages. Importantly, a direct comparison of low-risk and high-risk stages revealed epigenetic traits characteristic of high-risk tumours. Finally, a prognostic prediction model in cSCC patients identified a methylation signature able to predict the overall survival of patients. Thus, the analysis of DNA-methylation in cSCC revealed changes during the evolution of the disease through the different stages that can be of great value not only in the diagnosis but also in the prognosis of the disease.

Keywords: skin cancer, cSCC, epigenetic, DNA methylation, diagnosis signature, prognostic signature,

INTRODUCTION

Nonmelanoma skin cancer (NMSC) is the most common cancer in humans and its incidence has risen dramatically [1]. Approximately 20% of nonmelanoma skin cancers are cutaneous squamous cell carcinomas (cSCC). cSCC is by far the most common cancer with metastatic potential in white populations and there is a progressive increase of cSCC incidence with no tendency for levelling off [2-4]. Following the classical model of multistage carcinogenesis, cSCC is usually manifested as a spectrum of lesions of progressive malignancy. Actinic keratosis (AK), which consists in intraepidermal squamous neoplasms, is considered as cSCC precursor but its rate of malignant transformation is unknown. Once developed, the course of carcinoma consists of a local progression through the invasion of deep structures, regional extension by perineural or lymphatic invasion and the development of distant metastasis by vascular invasion. However, very few cases of cSCC reach advanced stages and the factors associated with rapid progression are unknown. In fact, most cSCCs have a good prognosis [5]. Nevertheless, there is a subgroup of SCC coined as high-risk cSCC associated with a higher frequency of local recurrence, lymph node metastasis and significant morbidity and mortality because of patient factors and features of the primary lesion. Risk stratification is critical to the management of cutaneous cSCC because it helps identify patients at high risk of poor outcomes. Although significant progress has been made in defining the clinic and pathologic factors associated with high-risk cSCC, identifying exactly which patients are at increased risk of recurrence, progression, metastasis, and death is not straightforward with these tools [6].

Genetic and genomic studies along the last decade have provided clues on the mutational landscape of cSCC [7-9]. However, the existence of a very high burden of mutation in cSCC, greater than in other solid tumours, makes difficult to identify the precise genetic events and number of mutations required for squamous cell carcinogenesis [10]. Alterations in the *p53* gene are the most common genetic abnormalities found in AK and invasive cSCC, and deregulation of *p53* pathway appears to be an early event in carcinogenesis of cSCC [11, 12]. Other molecular alterations in the cell cycle and proliferation regulatory circuit have been identified in cSCC such as loss-of-function

mutations in Notch [8, 9, 13] and TGF β receptors [14], upregulated MAPK [15], amplification of the *c-Myc* locus [16], activating mutations of *ras* [8, 9, 17], aberrant activation of tyrosine kinases EGFR [18] and Fyn [19], inactivation of inositol polyphosphate 5'-phosphatase, a negative regulator of PI3K/Akt signalling [20], as well as reduced expression of PKC δ [21] and p16^{INK4a} p14^{ARF} tumour suppressors [22]. Despite the notable number of genetic studies, the results are not determinant in explaining the complexity and evolution of the disease. In particular, there is a weak knowledge of molecular risk factors associated with the high-risk aggressive subtype of cSCC in immunocompetent patients before metastatic events..

Epigenetics, as an emerging field that provides a plausible link between the environment and alterations in gene expression, may be useful to unravel the molecular mechanisms involved in the pathophysiology of this disease [23]. DNA methylation is a particular mechanism of epigenetic regulation that implies a covalent epigenetic modification of cytosines within CpG dinucleotides established and maintained by DNA methyltransferases. The deregulation of the methylome is a hallmark of the development and progression of human cancers [24] and the altered DNA methylation patterns can be used as biomarkers for tumour detection, diagnosis and prognosis [23, 25]. In the case of cSCC, previous studies targeting specific genes have reported promoter hypermethylation associated with the disease in genes such as cell cycle regulator CDKN2A [22], cadherin CDH1 [26, 27] and CDH13 [28], transcription factor FOXE1 [29], modulators of Wnt signalling SFRPs [30] and FRZB [31], positive regulators of apoptosis ASC [32], G0S2 [33] and DAPK1 [34], and miRNA-204 [35], as well as hypomethylation of the DSS1 gene [36]. However, both DNA hypermethylation and hypomethylation associated with cancer is observed throughout the genome and these alterations, in particular the loss of methylation, may have functional consequences beyond regulation of proximal genes [24]. Due to this, a global vision of epigenetic regulation can only be obtained in large-scale approaches.

In the past decade, the advent of genome-scale analysis techniques, such as Infinium microarray systems, has revolutionized the field and has opened up the new era of epigenomics.

These large-scale approaches have been critical to significantly improve our molecular understanding of multiple diseases and enable potential key insights for developing epigenetic therapies [37, 38]. It has been described the existence of widespread blocks of hypomethylation associated with chronic exposure in older, sun-exposed nonmalignant skin that are maintained in cSCC samples [39]. More recently, it was reported that both AK and cSCC methylation patterns display classical features of cancer methylomes when compared to normal skin; however, no differences between AK and cSCC were detected [40]. Here, we use the last generation Infinium MethylationEPIC BeadChips 850K to interrogate DNA methylation changes in the progression from premalignant actinic keratosis to invasive cSCC, differentiating low-risk and high-risk tumours. The results allowed us to identify methylation signatures able to discriminate between progressive squamous cell carcinoma stages and to provide a prognostic prediction model for disease survival.

MATERIALS AND METHODS

Patient samples.

The patient samples comprised of 23 skin biopsies from 23 participants diagnosed at the four different stages of cutaneous squamous cell carcinoma (cSCC): actinic keratosis, early invasive carcinoma, high-risk non-metastatic carcinoma and high-risk carcinoma with nodal metastasis (S1 Table). The TNM classification of cSCC was evaluated according to the AJCC (American Joint Committee on Cancer) Cancer Staging Manual (8th edition) [41]. All the patients and samples were recruited from the Departament of Dermatology of the Instituto Valenciano de Oncología (IVO) in Valencia (Spain) after approval by the ethics committee of IVO. Written informed consent for the use of tissue samples was obtained from all included patients or their legal guardians. All research was performed in accordance with relevant guidelines and regulations. The samples were obtained from exposed skin areas from immunocompetent patients. The biopsies were obtained directly from the cSCC tumours at the time of diagnosis by surgical excision of 100-150 mg and stored as fresh frozen tissue, prepared and supplied by the Biobank at IVO. In the case of patients with actinic keratosis, a

shave biopsy of the entire lesion was obtained. The samples were divided into two parts, one part being sent to pathology for confirmation of the diagnosis and the other sent for freezing. To carry out the study we proceeded to separate the dermis and epidermis, avoiding the inclusion of dermal parts. This study included only samples of actinic keratosis where a sufficiently large area of epidermal layer could be dissected. In the case of invasive carcinomas, a punch biopsy of the most representative central portion of the tumour was obtained. All samples were evaluated visually by a trained dermatopathologist to validate tumour cellularity.

DNA extraction

Genomic DNA (gDNA) was extracted from the frozen tissues of the cSCC samples using the QIAcube from Qiagen® and the QIAamp® DNA Investigator Kit (Qiagen®, Germany), according to the manufacturer's protocol. Prior to performing the methylation studies, a DNA integrity quality control was performed to ensure that the DNA met the standard quality measurements that included a minimum requirement of 200 ng. All DNA samples were treated with RNaseA for 1 h at 45°C, quantified by the fluorometric method (Quant-iT PicoGreen dsDNA Assay, Life Technologies, CA, USA), and assessed for purity by NanoDrop (Thermo Scientific, MA, USA) 260/280 and 260/230 ratio measurements. DNA integrity of the fresh frozen samples was checked by electrophoresis on a 1.3% agarose gel.

Epigenomic studies

Epigenomic studies were performed with the Infinium MethylationEPIC BeadChips 850K array (Illumina Inc. USA) used for the interrogation of over 850,000 CpG sites (dinucleotides that are the main target for methylation), which has been previously established as a reliable technology to detect epigenetic alteration [42]. 600 ng of purified DNA were randomly distributed on a 96-well plate and processed using the EZ-96 DNA Methylation kit (Zymo Research Corp., CA, USA) following the manufacturer's recommendations for Infinium assays. Bisulfite-converted DNA (bs-DNA) was

processed as previously described [43]. DNA MethylationEPIC beadarray shares the Infinium HD chemistry Assay (Illumina Inc. USA) used to interrogate the cytosine markers with HumanMethylation450 beadchip. Thus, the applicable protocol for MethylationEPIC is the same as for HumanMethylation450, which is the Infinium HD Methylation Assay Protocol. 4 μ l of tissue bisulfite-DNA was processed following the Illumina Infinium HD Methylation Assay Protocol, as previously described [43].

The methylation score for each CpG was represented as a beta value according to the fluorescent intensity ratio. Beta values may take any value between 0 (non methylated) and 1 (completely methylated). The raw data (IDAT files) were normalized using functional normalization as implemented in the R-package minfi (version 1.22.1). CpG markers present on Methylation EPIC are classified based on its chromosome location and the feature category gene region as per UCSC annotation (TSS200, TSS1500, 5'UTR, 1st Exon, Body, 3'UTR). This array has a complete coverage of the genome including 99% of genes described. Additional criteria included the location of the marker relative to the CpG substructure (island, shore, shelf, open sea), fantom 5-associated enhancer regions and regulatory regions described on ENCODE project, such as transcription binding site sequences, open chromatin regions and DNase I hypersensitivity clusters.

Prefiltering

Every beta value in the EPIC array is accompanied by a detection p-value which represents the confidence of a given beta value. Probes and sample filtering involved a two-step process in which unreliable betas with high detection P value $P > 0.01$ (2220 CpGs) and 2932 CpGs associated with SNPs were removed. Previous analyses indicated that a threshold value of 0.01 allows a clear distinction to be made between reliable and unreliable beta values [44]. Sex chromosome probes were also removed (19681 CpGs). After the filtering, the remaining 841818 CpGs were considered valid for the study.

Statistical analysis

Two different approaches to assess differential methylation between the different studied groups were followed. First a multinomial elastic net penalized regression was used. Penalized regression methods consist in fitting a regression model subject to a specific restriction (a bound on the value of the coefficients). This method forces the shrinkage of the parameters to zero, potentially performing variable selection at the model-fitting step [45]. Penalization factor for the Elastic Net was selected using 500 repetitions of 10-fold cross-validation. From each repetition the highest lambda at one standard error from the minimum was selected (one-standard-error rule) and the median of the 500 lambda values was used as the final penalization factor. CpGs with coefficients different from zero after the penalization were considered as selected by the analysis. Additionally, a beta regression model was also adjusted for each CpG with each methylation value as response variable and the variable group as predictor [46]. P values of the beta regression analyses were adjusted for multiple comparisons by using False Discovery Rate. A prognosis model for survival was adjusted using an elastic net penalized cox regression model. Gene Ontology (GO) analysis was performed using *ShinnyGO* [47] with the results of the differential methylation analysis between low-risk initial invasive and high risk non-metastatic and metastatic cutaneous squamous cell carcinoma (cSCC) stages. All statistical analyses were performed using R (version 3.5.1) and R packages *glmnet* (version 2.0-13) and *betareg* (version 3.1-0).

RESULTS

Description and classification of multistage cutaneous squamous cell carcinoma patient samples

A critical point in this type of studies is the careful selection and classification of samples at different stages of the evolution of the pathology. In this work the classification is based on the Cancer Staging Manual 8th Edition of the American Joint Committee on Cancer (AJCC) [41], which distinguish cSCC stages based on the presence or absence of several high risk-features such as the presence of significant perineural invasion, invasion of the tumour beyond subcutaneous fat or depth

greater than 6mm and presence of bone involvement. The final cohort included patients with precursors of squamous cell carcinoma (actinic keratosis group) and patients with cSCC at different stages. Among the latter, we distinguish between carcinomas in the early stages of the disease (low-risk initially invasive cSCC group) defined by the absence of high risk features (T1 and T2 stages of AJCC-8) and cases at more advanced stages of the disease defined by the presence of high risk features (T3 and T4 stages of AJCC-8), differentiating those without nodal metastasis at the time of diagnosis (high-risk non metastatic group) from those with nodal metastasis (high-risk metastatic group). Long follow-up of all the cases is available to confirm the actual stage of the disease, and the overall survival. The patient characteristics are summarized in S1 Table.

Global DNA methylation profile in multistage cutaneous squamous cell carcinoma

In this work, we aimed to determine the overall genome DNA methylation profile in human skin from pre- and post-malignant multistage cSCC using the Infinium DNA Methylation EPIC BeadChip microarray 850K (Illumina), in order to investigate the existence of potential elements of distinction among specific stages of carcinoma development. In a first approach, an exploratory principal component analysis (PCA) revealed a separation among the clusters associated with the four analysed groups (Fig 1A). In particular, the results suggest that major transitions in methylation pattern occur between actinic keratosis, low-risk initial invasive carcinoma stage and high-risk carcinoma stages, with small differences associated with the presence or not of metastasis.

We wondered if DNA methylation alterations in cSCC occur sequentially throughout the different stages of skin tumour. To investigate this point, we performed a penalized ordinal regression including all CpGs as potential predictors. However, no ordinal association was found. This indicates that no sequential DNA methylation changes occur in the development of cSCC. Next, we performed a differential methylation study using a beta regression analysis for each CpG with the group as independent variable. Using an FDR-adjusted p-value cut-off of 0.05, we found 272,964 differentially methylated CpGs among the four groups that represent 32% of the CpGs investigated

(Fig 1B). A very high number of differentially methylated CpGs between the four groups were still detected when more stringent analysis (p-values from 10^{-2} to 10^{-10}) were carried out (S2 Table). An initial inspection of the heatmap suggests an interesting pattern in which the methylation state decreases from actinic keratosis to low-risk initial invasive carcinoma group but increases from low risk to high risk (non-metastatic and metastatic) groups. These results reveal a massive and complex pattern in the DNA methylation evolution during the development of cSCC affecting the global epigenomic structure.

Fig 1. Analysis of the global DNA-methylation profile in multistage cSCC. A) Representation of the Principal Component Analysis (PCA) on the methylation data. Centroids of each group of patients are represented by different shapes and colors (legend indicated in B). B) Heatmap with a random sample of 5000 CpGs from the results of the beta regression analysis for differential methylation among all cutaneous squamous cell carcinoma groups of patients. Z-score colour scale ranges from green for lower methylation to red for higher methylation levels.

Identification of a DNA-methylation signature characterizing the different stages of cSCC

In order to identify the minimum number of CpGs characterizing the distinct groups, we performed an elastic net penalized multinomial logistic regression with stage as response variable and all CpGs as potential predictors of epigenetic changes able to discriminate among groups. Interestingly, the results of this model showed that a profile consisting of only 94 CpG was enough to correctly classify 100% of the samples. The results of this supervised analysis are depicted in a clustered heatmap (Fig 2) and presented in S3 Table. Strikingly, the heatmap of these CpGs suggested that, on one hand, the initial invasive group showed lower methylation levels than premalignant actinic keratosis, and on the other hand, high-risk non-metastatic and metastatic groups displayed higher methylation levels. This confirms the existence of a non-sequential and complex pattern of DNA-methylation during cSCC evolution. Importantly, we have generated an epigenetic model based on genome wide profiling that may classify prospective samples based on the methylation level of 94 CpGs.

Fig 2. Identification of a DNA-methylation signature characterizing the different stages of cSCC.

Heatmap with the results of the elastic net analysis for discrimination among all cutaneous squamous cell carcinoma groups of patients. Rows (CpGs) and columns (individuals) are ordered according to the results of a hierarchical clustering algorithm. Z-score colour scale ranges from green for lower methylation to red for higher methylation levels.

Comparison of DNA-methylation pattern between low-risk and high-risk stages

To better differentiate between cSCC stages depending on the risk, a new elastic net penalized multinomial logistic regression analysis was performed to compare low-risk initial invasive cSCC patients to high-risk non-metastatic and metastatic cSCC patients. Given the small differences we observed in the PCA analysis between the two latter groups of samples (see Fig 1A) and previous reported analysis indicating that there is not widespread differences in the methylation patterns between non-metastatic versus metastatic cSCC [31], the high-risk non-metastatic and metastatic samples were considered as a single group in this analysis in order to optimize the output. The results revealed 637 differentially methylated CpGs (Fig 3A and S4 Table) between low-risk and high-risk cSCC. Most of them (602 CpGs) represent a gain of methylation in high-risk versus low-risk stages, whereas the minority of cases (35 CpGs) present a loss of methylation in high-risk versus low-risk stages. Thus, and consistent with the analysis showed above, the low-risk invasive group also showed lower median methylation levels than high-risk groups (Fig 3B). In conclusion, there is a risk-dependent change in DNA methylation patterns, mostly indicating a gain of methylation.

Fig 3. Comparison of DNA-methylation pattern between low-risk and high-risk stages. A) Heatmap with the results of the elastic net analysis for differential methylation between low-risk initial invasive and high risk non-metastatic and metastatic cutaneous squamous cell carcinoma (cSCC) stages. Z-score colour scale ranges from green for lower methylation to red for higher methylation levels. B) DNA methylation median β -values of the different epigenomic substructures in the different groups of cSCC. Box plots indicate highly significant ($p < 0.001$, two-sided t-test) hypermethylation of all substructures in both high-risk non-metastatic and metastatic groups compared to low-risk initial invasive cSCC samples. C) Distribution of selected differentially

methyated CpGs in high-risk compared to low-risk cSCC (hypomethyated in green and hypermethyated in red) and total CpG in the Illumina DNA methylation EPIC array in genes, promoters, gene body or intergenic regions. *D*) Distribution of differentially methyated CpGs in CpG islands, shore, shelf or open sea.

Next, the distribution of the identified CpGs with gain and loss of methylation between high-risk versus low-risk stages, was analysed in the Illumina EPIC array regarding functional genomic distribution (FGD) (Fig 3C) and CpG substructure context (Fig 3D). With regards to FGD, CpGs with gain of methylation were preferentially found in the gene body whereas CpGs with loss of methylation were enriched in the promoter regions (Fig 3C). Notably, based on CpG substructure context, the distribution of CpGs with gain of methylation revealed a significant enrichment in open sea (low density CpG regions), while CpGs with loss of methylation were mainly found in CpG shores (Fig 3D). Strikingly, CpGs located in CpG islands (high density CpG regions) were underrepresented in both cases.

Finally, the functional signature of the genes associated with DNA methylation changes was investigated with the ShinyGO data set [47]. 395 out of the 454 genes with altered CpG methylation were classified based on Biological Process categories. The detected enriched categories using a p-value cutoff of 0.01 are shown in Fig 4 and S5 Table. Interestingly, this analysis revealed that genes with changes in CpG methylation are mainly associated with categories that are directly related to cancer, such as signal transduction and cell motility, as well as phospholipid metabolic processes. Further analysis of KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways [48] showed an enrichment in key signalling pathways related to cell proliferation and cell fate like Notch, hedgehog, MAPK, TFG β and Wnt pathways (S6 Table).

Fig 4. Analysis of genes associated to differentially methyated CpG between low-risk and high-risk stages. Hierarchical clustering tree of the enriched gene sets obtained from the GO analysis of the genes selected in the comparison between low-risk initial invasive and high risk non-metastatic and metastatic cutaneous squamous cell carcinoma (cSCC) stages.

Prognosis signature

Finally, we attempted to produce a prognostic prediction model for overall survival in cSCC patients. To do this, a penalized cox regression model using all CpGs as potential predictors was adjusted. The adjusted elastic net model selected a specific signature of 32 CpGs (S7 Table) able to predict risk of death in these patients with a very high discriminant capacity (Area under the curve, AUC = 0.98). This discrimination capacity was represented using model-estimated survival curves showing that patients with a higher predicted risk had a markedly shorter survival time compared to patients with a lower predicted risk (Fig 5). Thus, the analysis of the methylation status of these 32 CpGs emerges as a good approach to predict the evolution of the disease.

Fig 5. Survival model based on a prognosis DNA-methylation signature. Representation of the predicted survival curves in patients with cSCC for different values of the risk estimated by the regression model using 32 CpGs. Patients with higher predicted risk levels show sizeably lower overall survival (OS) in comparison with patients with lower predicted risk levels.

DISCUSSION

In recent years, epigenetics has become a discipline capable of explaining aspects of different pathologies. In particular, it has shed light on the evolution of different types of cancer [24]. cSCC is a major type of skin cancer with an increasing incidence. However, it has been less studied than other skin cancers and there is not a definitive description of its genetic and epigenetic bases. In this context, this work tried to analyse the epigenetic phenomena underlying the evolution of the pathology.

cSCC is a particular type of cancer in which tumour cells accumulate a big amount of genetic alterations due to chronic exposure to sun radiation[10]. It has been hypothesized that ultraviolet radiation in natural sunlight may promote tumorigenesis not merely due to DNA damage, but also through induction of epigenetic dysregulation. Consistent with this, previous reports have described

that aberrant DNA-methylation patterns related to cSCC, when compared to healthy skin, seems to involve vast changes encompassing a large proportion of the genome [39, 40]. Thus, cSCC is associated with massive alterations of both genetic and epigenetic traits. We have now identified that an enormous amount of CpG also show significant changes between the different stages of the disease. This indicates that the massive epigenetic alterations associated with cSCC is not just a stationary terminal characteristic of tumour cells; rather, enormous changes are dynamically occurring along the evolution of the disease through their different stages. In addition, our results demonstrate that the different stages of cSCC present indeed specific DNA-methylation characteristics that allow to distinguish them. In fact, a minimal epigenetic signature that provides a powerful tool to discriminate between the different stages was obtained using DNA-methylation information from only 94 CpGs. This signature could be useful to classify prospective samples.

From a clinicopathological point of view, the most interesting aspect is the ability to distinguish between a low-risk and a high-risk stage. A direct comparison of these groups provides evidence for the existence of considerable changes in the DNA methylation pattern that are associated with a worse evolution of the tumour. In fact, 637 differentially methylated CpGs were identified. In 95% of cases, the change implies a hypermethylation of CpGs in the high-risk samples. A recent study suggested that UV-irradiation through chronic sun exposure gives rise to large hypomethylated blocks in healthy epidermis [39]. In view of our results, it is tempting to speculate that DNA-hypomethylation caused by sun exposure characterizes the premalignant and low-risk stages of cSCC and that the evolution to more advanced stages of worse prognosis occurs when the state of methylation is reversed changing from hypo to hypermethylated DNA.

Specific changes in gene function related to cell proliferation, cell fate and cell motility could be epigenetically controlled in the transition from low-risk to high-risk stage. DNA methylation could affect in different ways to gene function [24, 49]. The methylation state of a promoter affects gene expression, DNA hypermethylation resulting in gene silencing. Also, DNA methylation increases the rate of mutation and tumours tend to be globally hypomethylated, which generally results in

genomic instability. In our analysis, CpGs with gain of methylation are preferentially found in the gene body whereas CpGs with loss of methylation were enriched in the promoter regions. It can be speculated that, overall, the loss of methylation in promoters could lead to increased expression of key genes and that hypermethylation in body regions could increase the rate of mutation. With regards to the CpG substructure, it is surprising that CpG islands were underrepresented in the group of differentially methylated CpGs between low-risk and high-risk stages. Nevertheless, other authors have also pointed out the relevance of methylation changes in open seas and shores in aged sun-exposed skin and cSCC [39, 40], as occurs in other cancers [50]. Thus, the major changes in the methylome during the development of cSCC does not seem to occur in CpG islands.

Interestingly, the GO analysis of the genes associated with differentially methylated CpG between low-risk and high-risk stages, indicates an enrichment of genes in categories directly involved in cancer development like signal transduction and cell motility. In particular, it has to be noted that the number of genes related to signal transduction that manifested an alteration in their DNA-methylaton pattern is of 119 (S5 Table); this suggest that the transition to a high-risk state of the disease could be caused by a collection of dysfunctions in several signalling processes controlling cell behaviour. In fact, the KEGG analysis identifies important signalling pathways like MAPK, Notch, Wnt, TFG β and hedgehog pathways, as potential pathways missregulated in the transition to a high-risk stage. It is important to remark that epigenetic alterations in other genes related to the Wnt-signalling pathway had been previously reported in cSCC samples [30, 31], as well as genetic mutation in Notch receptors [8, 9, 13]; our results supports a more broad extension of epigenetic alterations in the Wnt pathway and point to a combination of genetic and epigenetic alterations in the deregulation of the Notch pathway in cSCC.

Very recently, other authors reported a study of the DNA-methylome in cSCC [39, 40]. They report that both AK and cSCC methylation patterns display classical features of cancer methylomes. Importantly however, and in contrast to our results, no differences in the methylation profile between these stages of the pathology were identified using linear regression methods. A big

difference between Rodriguez-Paredes et al. analysis and our analysis is that Rodriguez-Paredes et al. only considered two groups of samples, premalignant AK and tumour samples, independently of the stage of evolution of cSCC. As commented above, we have detected a complex non-linear evolution of the DNA-methylation pattern during the development of cSCC, so it is possible that a low-stringent classification of samples could mask the detection of changes. To address the potential conflict between both studies, we performed an analysis on our dataset including our three cSCC subgroups (low-risk initial invasive, high risk non metastatic and metastatic) as one single group and comparing this group against the actinic keratosis group. Results of the analysis completely agreed with those of Rodriguez-Paredes et al, finding no significant differences between the actinic keratosis patients and all-grouped cSCC patients. This result confirmed that disregarding the non-sequential epigenetic pattern among the different cSCC subgroups results in an inability to detect epigenetic alterations, since DNA methylation gain/loss among cSCC subgroups could be compensated and cancelled.

Finally, a prognostic prediction model was developed. Histopathology and immunohistochemistry, the standard methods for cancer diagnosis, have important limitations since they are time consuming and the result depends on a pathologist's subjective interpretation. Due to this, there is a need for molecular biomarkers that could be useful in tumour diagnosis and grading and with a potential use in prognosis of the disease. In this sense, we have identified a signature based on the methylation state of 32 CpGs that provides a risk-score, which is associated with the overall survival of the patients. Notably, one of these CpGs is located in the gene coding for p73, a functional homolog of p53 that has been related to cancer [51-53]. p53 deregulation is an early event in carcinogenesis of cSCC, so it is tempting to speculate that alteration of its functional homolog p73 could be critical for bad prognosis. Although this 32 CpG signature is a preliminary proposal, it supports the general conclusion that the analysis of the degree of methylation for DNA samples from cSCC patients can be of great value, not only in the diagnosis but also in the prognosis. The development of future tools following this basis and its validation with larger cohorts could be of important help in clinical practice.

CONCLUSIONS

Our results showed massive non-sequential changes in DNA-methylome along the different cSCC stages, identified a minimal methylation signature that discriminates between stages and revealed epigenetic traits characteristic of high-risk tumours. In addition, a prognostic prediction model in cSCC patients identified a methylation signature able to predict the overall survival of patients. Thus, we conclude that DNA-methylation information can be of great value in the diagnosis and the prognosis of cSCC.

ACKNOWLEDGEMENTS

We thank Diana García from IIS La Fe and Marta Ramírez-Calvo and Arantxa Rodríguez-Hernández from IVO for their technical support.

REFERENCES

1. Donaldson MR, Coldiron BM. No end in sight: the skin cancer epidemic continues. *Semin Cutan Med Surg.* 2011;30(1):3-5. doi: 10.1016/j.sder.2011.01.002. PubMed PMID: 21540015.
2. Alam M, Ratner D. Cutaneous squamous-cell carcinoma. *N Engl J Med.* 2001;344(13):975-83. doi: 10.1056/NEJM200103293441306. PubMed PMID: 11274625.
3. Green AC, Olsen CM. Cutaneous squamous cell carcinoma: an epidemiological review. *Br J Dermatol.* 2017;177(2):373-81. doi: 10.1111/bjd.15324. PubMed PMID: 28211039.
4. Leiter U, Keim U, Eigentler T, Katalinic A, Holleczek B, Martus P, et al. Incidence, Mortality, and Trends of Nonmelanoma Skin Cancer in Germany. *J Invest Dermatol.* 2017;137(9):1860-7. doi: 10.1016/j.jid.2017.04.020. PubMed PMID: 28487088.
5. Schmults CD, Karia PS, Carter JB, Han J, Qureshi AA. Factors predictive of recurrence and death from cutaneous squamous cell carcinoma: a 10-year, single-institution cohort study. *JAMA Dermatol.* 2013;149(5):541-7. doi: 10.1001/jamadermatol.2013.2139. PubMed PMID: 23677079.
6. Ruiz ES, Karia PS, Besaw R, Schmults CD. Performance of the American Joint Committee on Cancer Staging Manual, 8th Edition vs the Brigham and Women's Hospital Tumor Classification System for Cutaneous Squamous Cell Carcinoma. *JAMA Dermatol.* 2019. doi: 10.1001/jamadermatol.2019.0032. PubMed PMID: 30969315.
7. Inman GJ, Wang J, Nagano A, Alexandrov LB, Purdie KJ, Taylor RG, et al. The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. *Nat Commun.* 2018;9(1):3667. doi: 10.1038/s41467-018-06027-1. PubMed PMID: 30202019; PubMed Central PMCID: PMC6131170.
8. Pickering CR, Zhou JH, Lee JJ, Drummond JA, Peng SA, Saade RE, et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin Cancer Res.* 2014;20(24):6582-92. doi: 10.1158/1078-0432.CCR-14-1768. PubMed PMID: 25303977; PubMed Central PMCID: PMC4367811.
9. South AP, Purdie KJ, Watt SA, Haldenby S, den Breems N, Dimon M, et al. NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. *J Invest Dermatol.* 2014;134(10):2630-8. doi: 10.1038/jid.2014.154. PubMed PMID: 24662767; PubMed Central PMCID: PMC4753672.
10. Harwood CA, Proby CM, Inman GJ, Leigh IM. The Promise of Genomics and the Development of Targeted Therapies for Cutaneous Squamous Cell Carcinoma. *Acta Derm Venereol.* 2016;96(1):3-16. doi: 10.2340/00015555-2181. PubMed PMID: 26084328.
11. Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, et al. Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov.* 2011;1(2):137-43. doi: 10.1158/2159-8290.CD-11-0028. PubMed PMID: 21984974; PubMed Central PMCID: PMC3187561.
12. Giglia-Mari G, Sarasin A. TP53 mutations in human skin cancers. *Hum Mutat.* 2003;21(3):217-28. doi: 10.1002/humu.10179. PubMed PMID: 12619107.
13. Wang NJ, Sanborn Z, Arnett KL, Bayston LJ, Liao W, Proby CM, et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proc Natl Acad Sci U S A.* 2011;108(43):17761-6. doi: 10.1073/pnas.1114669108. PubMed PMID: 22006338; PubMed Central PMCID: PMC3203814.
14. Cammareri P, Rose AM, Vincent DF, Wang J, Nagano A, Libertini S, et al. Inactivation of TGFbeta receptors in stem cells drives cutaneous squamous cell carcinoma. *Nat Commun.* 2016;7:12493. doi: 10.1038/ncomms12493. PubMed PMID: 27558455; PubMed Central PMCID: PMC5007296.
15. Lambert SR, Mladkova N, Gulati A, Hamoudi R, Purdie K, Cerio R, et al. Key differences identified between actinic keratosis and cutaneous squamous cell carcinoma by transcriptome profiling. *Br J Cancer.* 2014;110(2):520-9. doi: 10.1038/bjc.2013.760. PubMed PMID: 24335922;

PubMed Central PMCID: PMCPMC3899778.

16. Masferrer E, Ferrandiz-Pulido C, Lloveras B, Masferrer-Niubo M, Espinet B, Salido M, et al. MYC copy number gains are associated with poor outcome in penile squamous cell carcinoma. *J Urol*. 2012;188(5):1965-71. doi: 10.1016/j.juro.2012.07.003. PubMed PMID: 22999547.
17. Su F, Viros A, Milagre C, Trunzer K, Bollag G, Spleiss O, et al. RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. *N Engl J Med*. 2012;366(3):207-15. doi: 10.1056/NEJMoa1105358. PubMed PMID: 22256804; PubMed Central PMCID: PMCPMC3724537.
18. Toll A, Salgado R, Yebenes M, Martin-Ezquerria G, Gilaberte M, Baro T, et al. Epidermal growth factor receptor gene numerical aberrations are frequent events in actinic keratoses and invasive cutaneous squamous cell carcinomas. *Exp Dermatol*. 2010;19(2):151-3. doi: 10.1111/j.1600-0625.2009.01028.x. PubMed PMID: 20156290.
19. Yadav V, Denning MF. Fyn is induced by Ras/PI3K/Akt signaling and is required for enhanced invasion/migration. *Mol Carcinog*. 2011;50(5):346-52. doi: 10.1002/mc.20716. PubMed PMID: 21480388; PubMed Central PMCID: PMCPMC3080437.
20. Sekulic A, Kim SY, Hostetter G, Savage S, Einspahr JG, Prasad A, et al. Loss of inositol polyphosphate 5-phosphatase is an early event in development of cutaneous squamous cell carcinoma. *Cancer Prev Res (Phila)*. 2010;3(10):1277-83. doi: 10.1158/1940-6207.CAPR-10-0058. PubMed PMID: 20876729; PubMed Central PMCID: PMCPMC2955780.
21. Yadav V, Yanez NC, Fenton SE, Denning MF. Loss of protein kinase C delta gene expression in human squamous cell carcinomas: a laser capture microdissection study. *Am J Pathol*. 2010;176(3):1091-6. doi: 10.2353/ajpath.2010.090816. PubMed PMID: 20093486; PubMed Central PMCID: PMCPMC2832131.
22. Brown VL, Harwood CA, Crook T, Cronin JG, Kelsell DP, Proby CM. p16INK4a and p14ARF tumor suppressor genes are commonly inactivated in cutaneous squamous cell carcinoma. *J Invest Dermatol*. 2004;122(5):1284-92. doi: 10.1111/j.0022-202X.2004.22501.x. PubMed PMID: 15140233.
23. Rodriguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med*. 2011;17(3):330-9. doi: 10.1038/nm.2305. PubMed PMID: 21386836.
24. Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. *Adv Exp Med Biol*. 2013;754:31-56. doi: 10.1007/978-1-4419-9967-2_2. PubMed PMID: 22956495.
25. Sandoval J, Esteller M. Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev*. 2012;22(1):50-5. doi: 10.1016/j.gde.2012.02.008. PubMed PMID: 22402447.
26. Chiles MC, Ai L, Zuo C, Fan CY, Smoller BR. E-cadherin promoter hypermethylation in preneoplastic and neoplastic skin lesions. *Mod Pathol*. 2003;16(10):1014-8. doi: 10.1097/01.MP.0000089779.35435.9D. PubMed PMID: 14559984.
27. Murao K, Kubo Y, Ohtani N, Hara E, Arase S. Epigenetic abnormalities in cutaneous squamous cell carcinomas: frequent inactivation of the RB1/p16 and p53 pathways. *Br J Dermatol*. 2006;155(5):999-1005. doi: 10.1111/j.1365-2133.2006.07487.x. PubMed PMID: 17034532.
28. Takeuchi T, Liang SB, Matsuyoshi N, Zhou S, Miyachi Y, Sonobe H, et al. Loss of T-cadherin (CDH13, H-cadherin) expression in cutaneous squamous cell carcinoma. *Lab Invest*. 2002;82(8):1023-9. PubMed PMID: 12177241.
29. Venza I, Visalli M, Tripodo B, De Grazia G, Loddo S, Teti D, et al. FOXE1 is a target for aberrant methylation in cutaneous squamous cell carcinoma. *Br J Dermatol*. 2010;162(5):1093-7. doi: 10.1111/j.1365-2133.2009.09560.x. PubMed PMID: 19845668.
30. Liang J, Kang X, Halifu Y, Zeng X, Jin T, Zhang M, et al. Secreted frizzled-related protein promoters are hypermethylated in cutaneous squamous carcinoma compared with normal epidermis. *BMC Cancer*. 2015;15:641. doi: 10.1186/s12885-015-1650-x. PubMed PMID: 26394929; PubMed Central PMCID: PMCPMC4579584.

31. Darr OA, Colacino JA, Tang AL, McHugh JB, Bellile EL, Bradford CR, et al. Epigenetic alterations in metastatic cutaneous carcinoma. *Head Neck*. 2015;37(7):994-1001. doi: 10.1002/hed.23701. PubMed PMID: 24700717; PubMed Central PMCID: PMC4183742.
32. Meier K, Drexler SK, Eberle FC, Lefort K, Yazdi AS. Silencing of ASC in Cutaneous Squamous Cell Carcinoma. *PLoS One*. 2016;11(10):e0164742. doi: 10.1371/journal.pone.0164742. PubMed PMID: 27768771; PubMed Central PMCID: PMC45074456.
33. Nobeyama Y, Watanabe Y, Nakagawa H. Silencing of G0/G1 switch gene 2 in cutaneous squamous cell carcinoma. *PLoS One*. 2017;12(10):e0187047. doi: 10.1371/journal.pone.0187047. PubMed PMID: 29073263; PubMed Central PMCID: PMC5658152.
34. Li L, Jiang M, Feng Q, Kiviat NB, Stern JE, Hawes S, et al. Aberrant Methylation Changes Detected in Cutaneous Squamous Cell Carcinoma of Immunocompetent Individuals. *Cell Biochem Biophys*. 2015;72(2):599-604. doi: 10.1007/s12013-014-0507-2. PubMed PMID: 25575897.
35. Toll A, Salgado R, Espinet B, Diaz-Lagares A, Hernandez-Ruiz E, Andrades E, et al. MiR-204 silencing in intraepithelial to invasive cutaneous squamous cell carcinoma progression. *Mol Cancer*. 2016;15(1):53. doi: 10.1186/s12943-016-0537-z. PubMed PMID: 27457246; PubMed Central PMCID: PMC4960761.
36. Venza M, Visalli M, Catalano T, Beninati C, Teti D, Venza I. DSS1 promoter hypomethylation and overexpression predict poor prognosis in melanoma and squamous cell carcinoma patients. *Hum Pathol*. 2017;60:137-46. doi: 10.1016/j.humpath.2016.10.018. PubMed PMID: 27825810.
37. Ahuja N, Sharma AR, Baylin SB. Epigenetic Therapeutics: A New Weapon in the War Against Cancer. *Annu Rev Med*. 2016;67:73-89. doi: 10.1146/annurev-med-111314-035900. PubMed PMID: 26768237; PubMed Central PMCID: PMC4937439.
38. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, et al. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods*. 2013;10(10):949-55. doi: 10.1038/nmeth.2632. PubMed PMID: 24076989.
39. Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X, et al. Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol*. 2015;16:80. doi: 10.1186/s13059-015-0644-y. PubMed PMID: 25886480; PubMed Central PMCID: PMC4423110.
40. Rodriguez-Paredes M, Bormann F, Raddatz G, Gutekunst J, Lucena-Porcel C, Kohler F, et al. Methylation profiling identifies two subclasses of squamous cell carcinoma related to distinct cells of origin. *Nat Commun*. 2018;9(1):577. doi: 10.1038/s41467-018-03025-1. PubMed PMID: 29422656; PubMed Central PMCID: PMC5805678.
41. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, et al. *AJCC Cancer Staging Manual*. 8 ed. New York: Springer International Publishing; 2007.
42. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8(3):389-99. doi: 10.2217/epi.15.114. PubMed PMID: 26673039; PubMed Central PMCID: PMC4864062.
43. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*. 2011;6(6):692-702. PubMed PMID: 21593595.
44. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J Clin Oncol*. 2013;31(32):4140-7. doi: 10.1200/JCO.2012.48.5516. PubMed PMID: 24081945.
45. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67(2):301-20. doi: doi:10.1111/j.1467-9868.2005.00503.x.
46. Ferrari S, Cribari-Neto F. Beta Regression for Modelling Rates and Proportions. *Journal of*

Applied Statistics. 2004;31(7):799-815. doi: 10.1080/0266476042000214501.

47. Ge S, Jung D. ShinyGO: a graphical enrichment tool for animals and plants. bioRxiv. 2018:315150. doi: 10.1101/315150.

48. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019;47(D1):D590-D5. doi: 10.1093/nar/gky962. PubMed PMID: 30321428; PubMed Central PMCID: PMC6324070.

49. van Doorn R, Gruis NA, Willemze R, van der Velden PA, Tensen CP. Aberrant DNA methylation in cutaneous malignancies. Semin Oncol. 2005;32(5):479-87. doi: 10.1053/j.seminoncol.2005.07.001. PubMed PMID: 16210089.

50. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009;41(2):178-86. doi: 10.1038/ng.298. PubMed PMID: 19151715; PubMed Central PMCID: PMC2729128.

51. Inoue K, Fry EA. Alterations of p53 and p73 in Human Cancers. In: Deb SP, Deb S, editors. Mutant p53 and MDM2 in Cancer. Dordrecht: Springer Netherlands; 2014. p. 17-40.

52. Pflaum J, Schlosser S, Muller M. p53 Family and Cellular Stress Responses in Cancer. Front Oncol. 2014;4:285. doi: 10.3389/fonc.2014.00285. PubMed PMID: 25374842; PubMed Central PMCID: PMC4204435.

53. Rodriguez N, Pelaez A, Barderas R, Dominguez G. Clinical implications of the deregulated TP73 isoforms expression in cancer. Clin Transl Oncol. 2018;20(7):827-36. doi: 10.1007/s12094-017-1802-3. PubMed PMID: 29230693.

SUPPORTING INFORMATION

S1 Table: List of samples.

S2 Table: Number of differentially methylated CpGs between groups.

S3 Table: DNA-methylation signature: 94 CpGs discriminating the four groups.

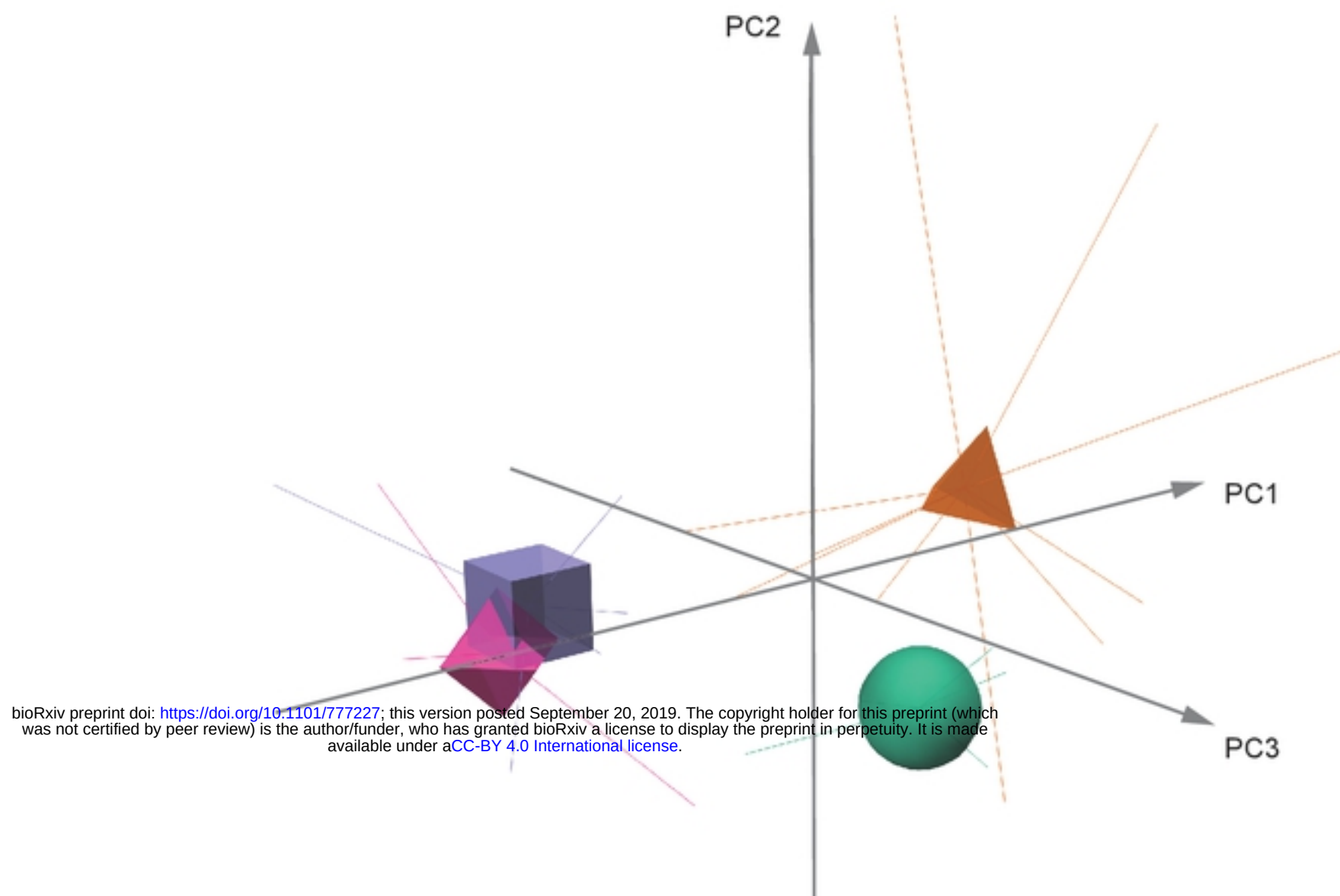
S4 Table: Differentially methylated CpGs between high-risk and low-risk carcinomas.

S5 Table: GO categories enriched in genes associated to DM CpGs between low-risk and high-risk carcinomas.

S6 Table: Enriched KEGG pathways

S7 Table: DNA-methylation prognosis signature: 32 CpGs predicting bad prognosis

A



B

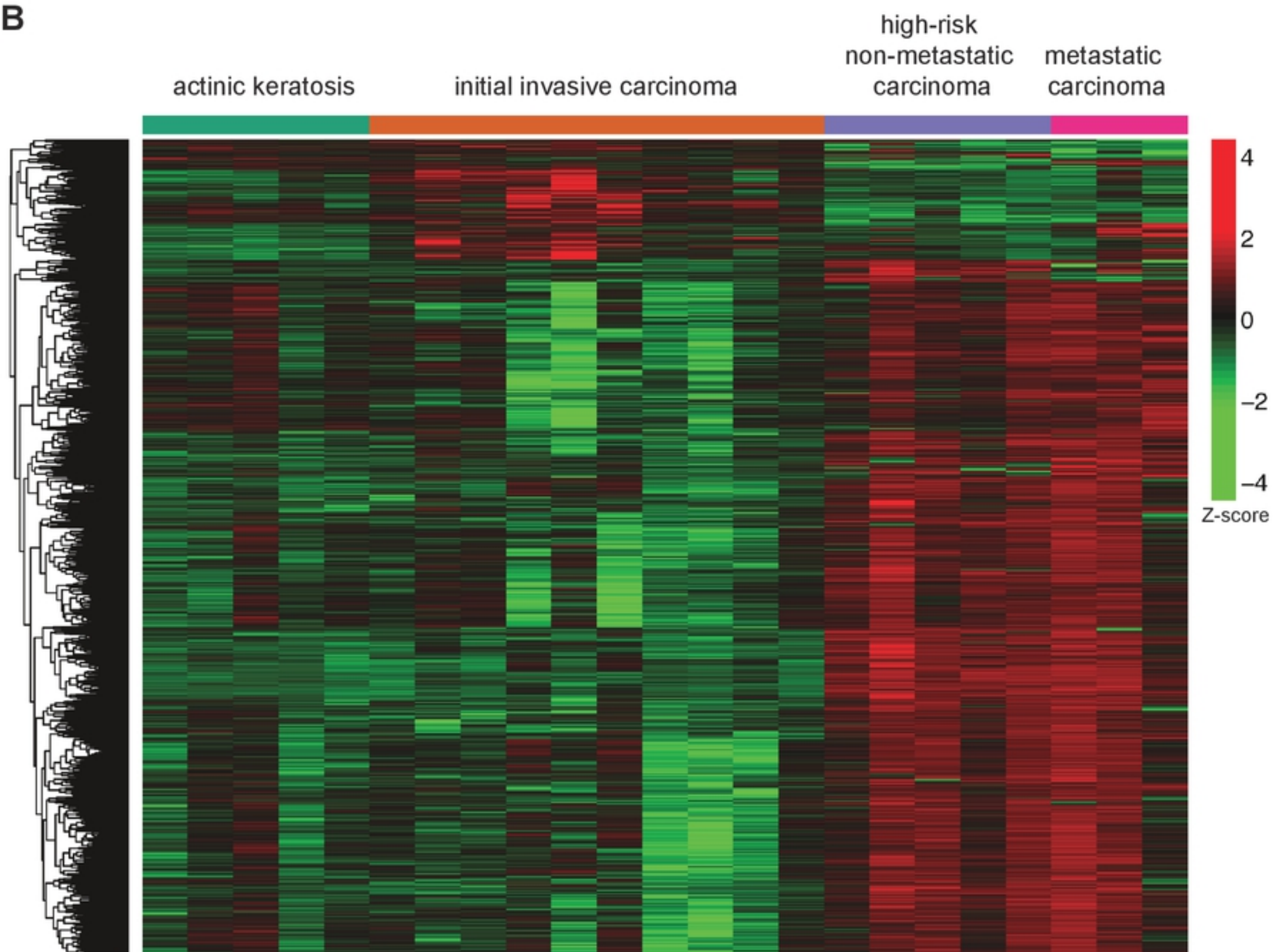


Figure 1

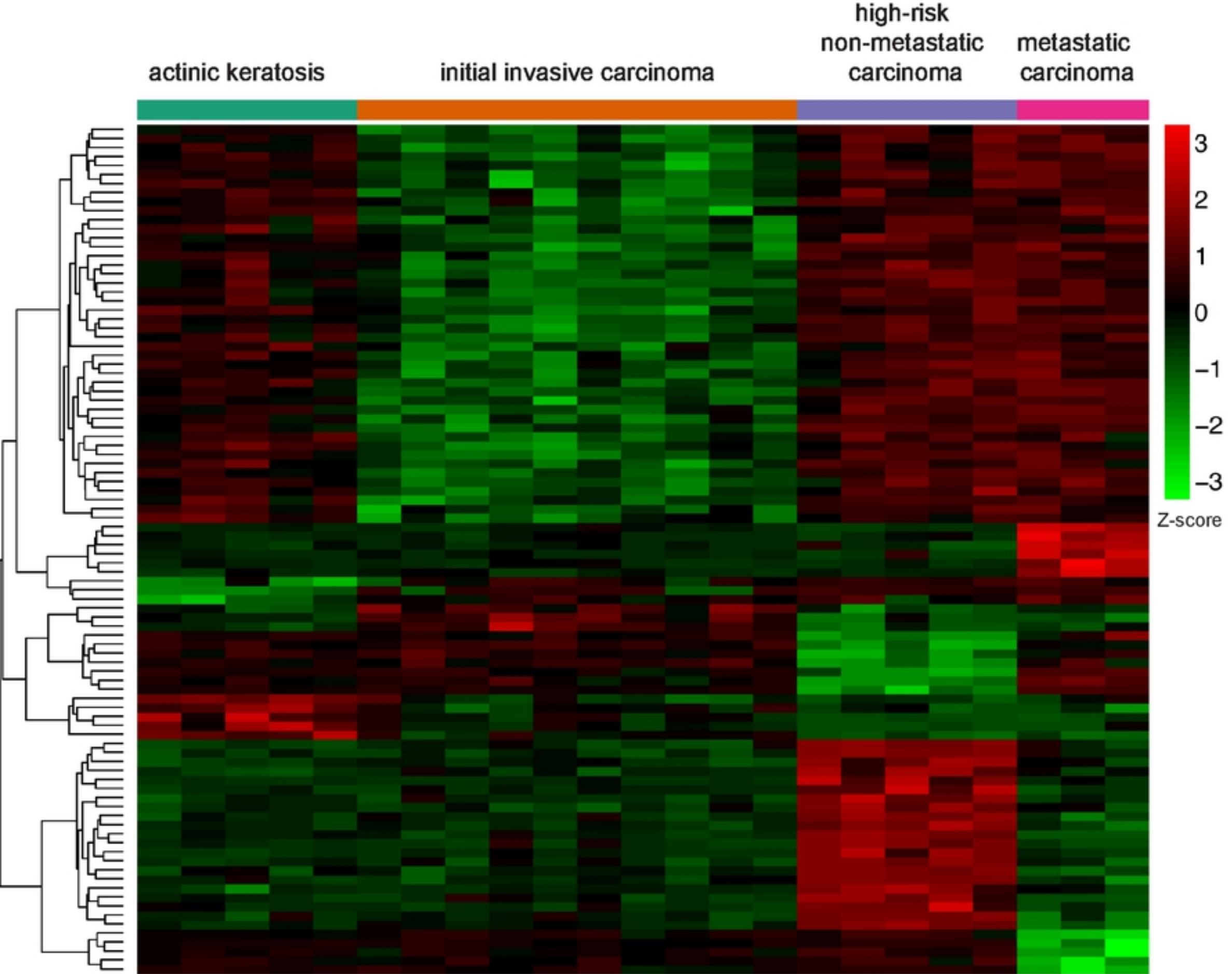
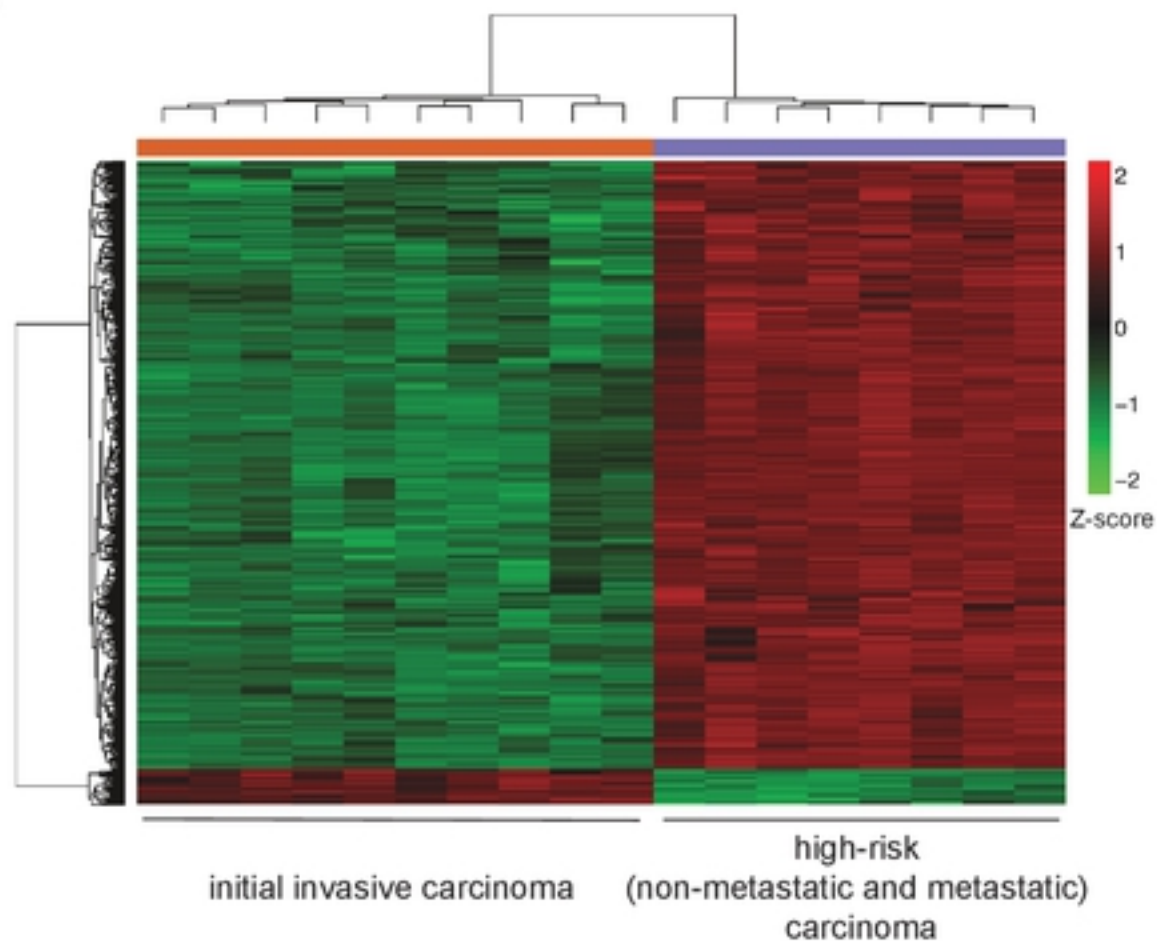
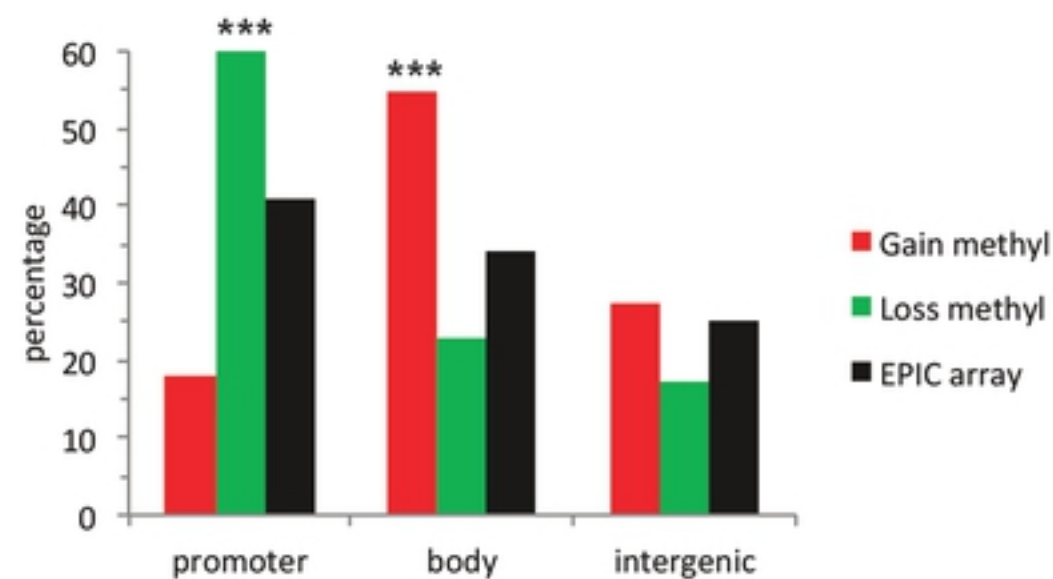


Figure 2

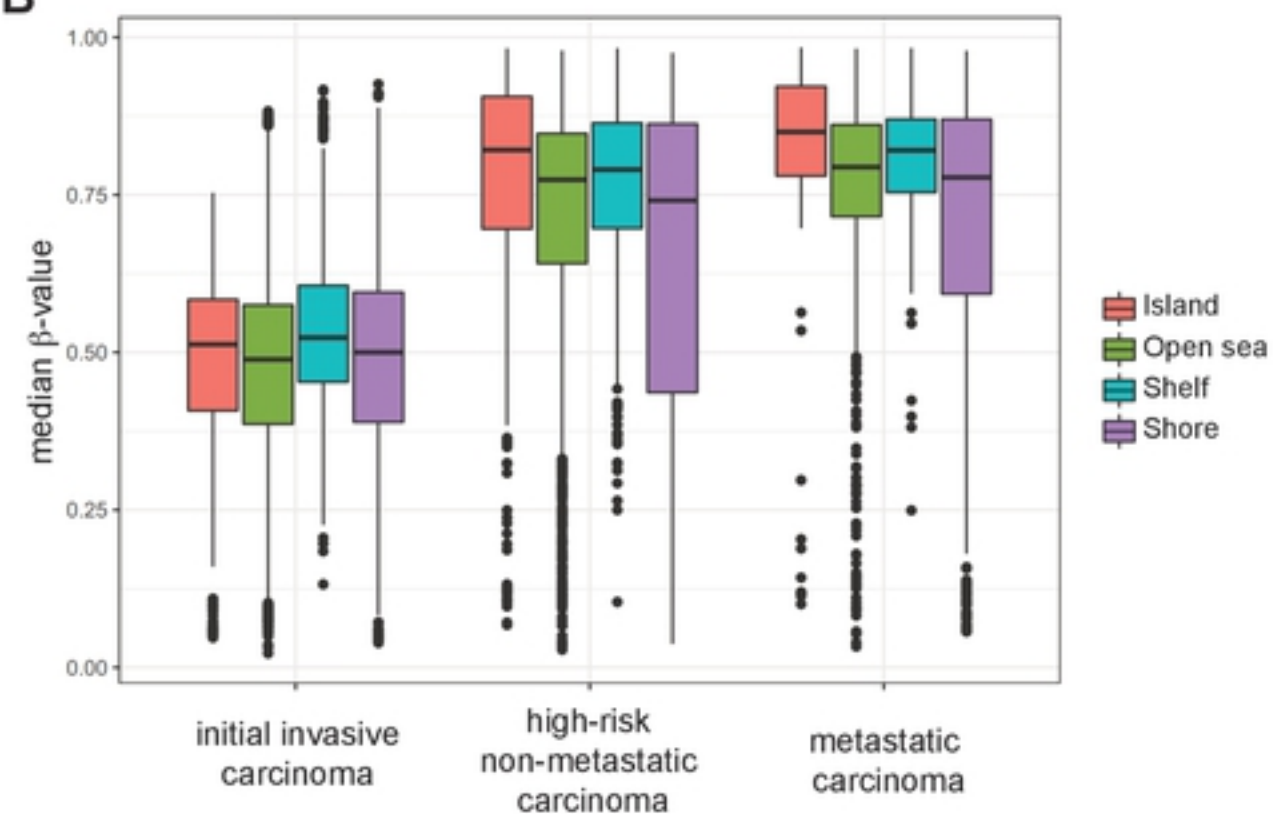
A



C



B



D

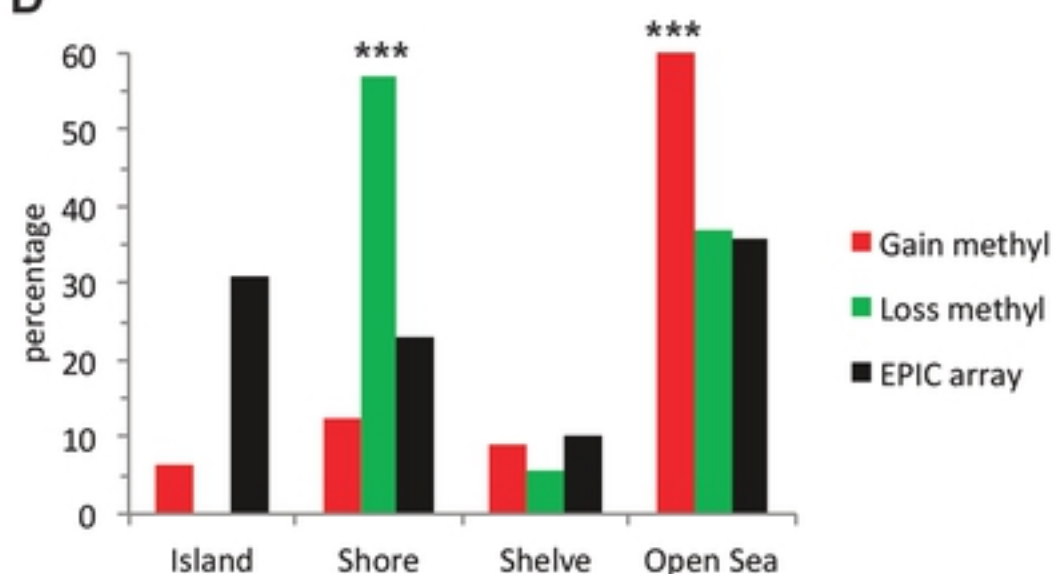


Figure 3

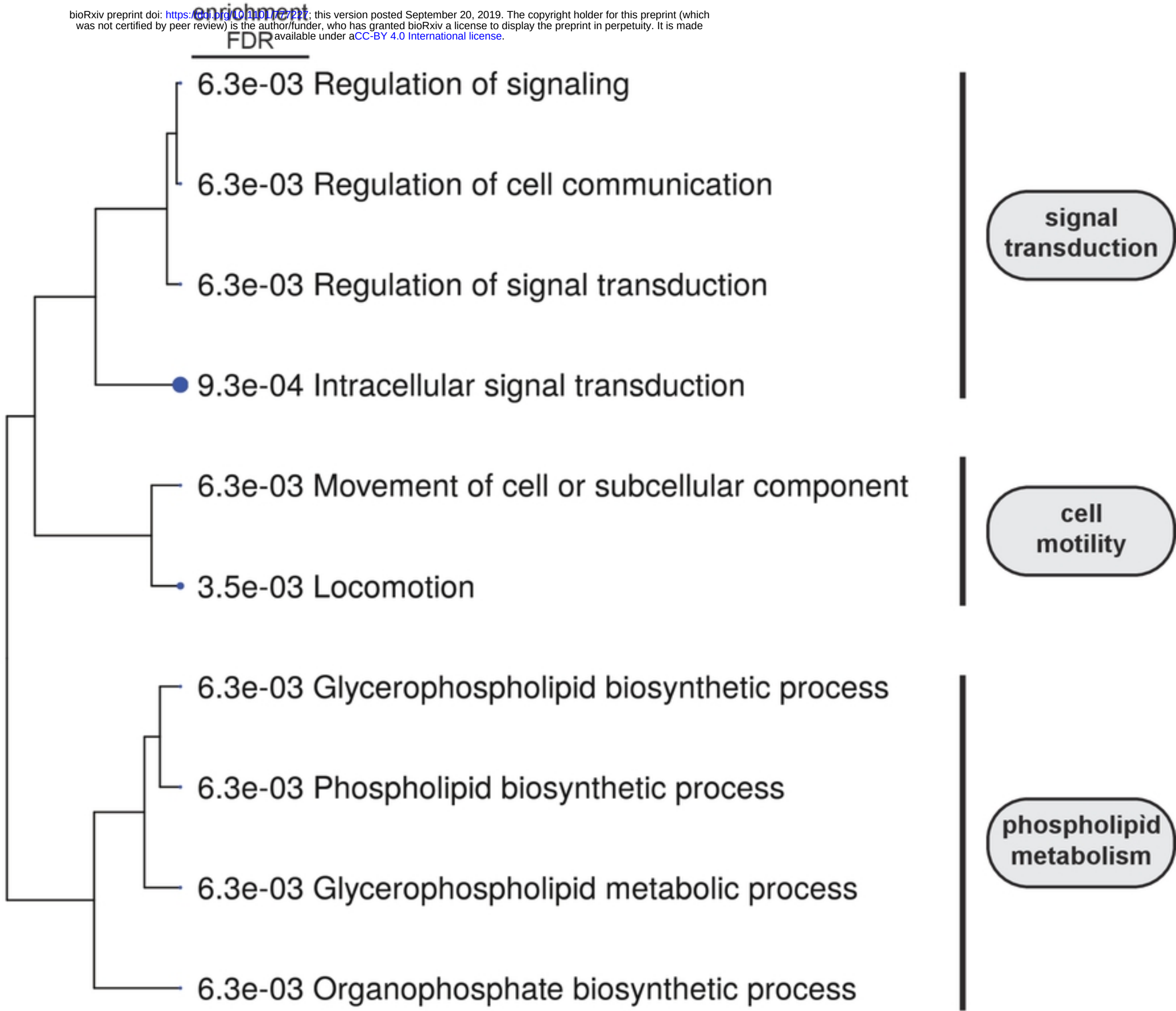


Figure 4

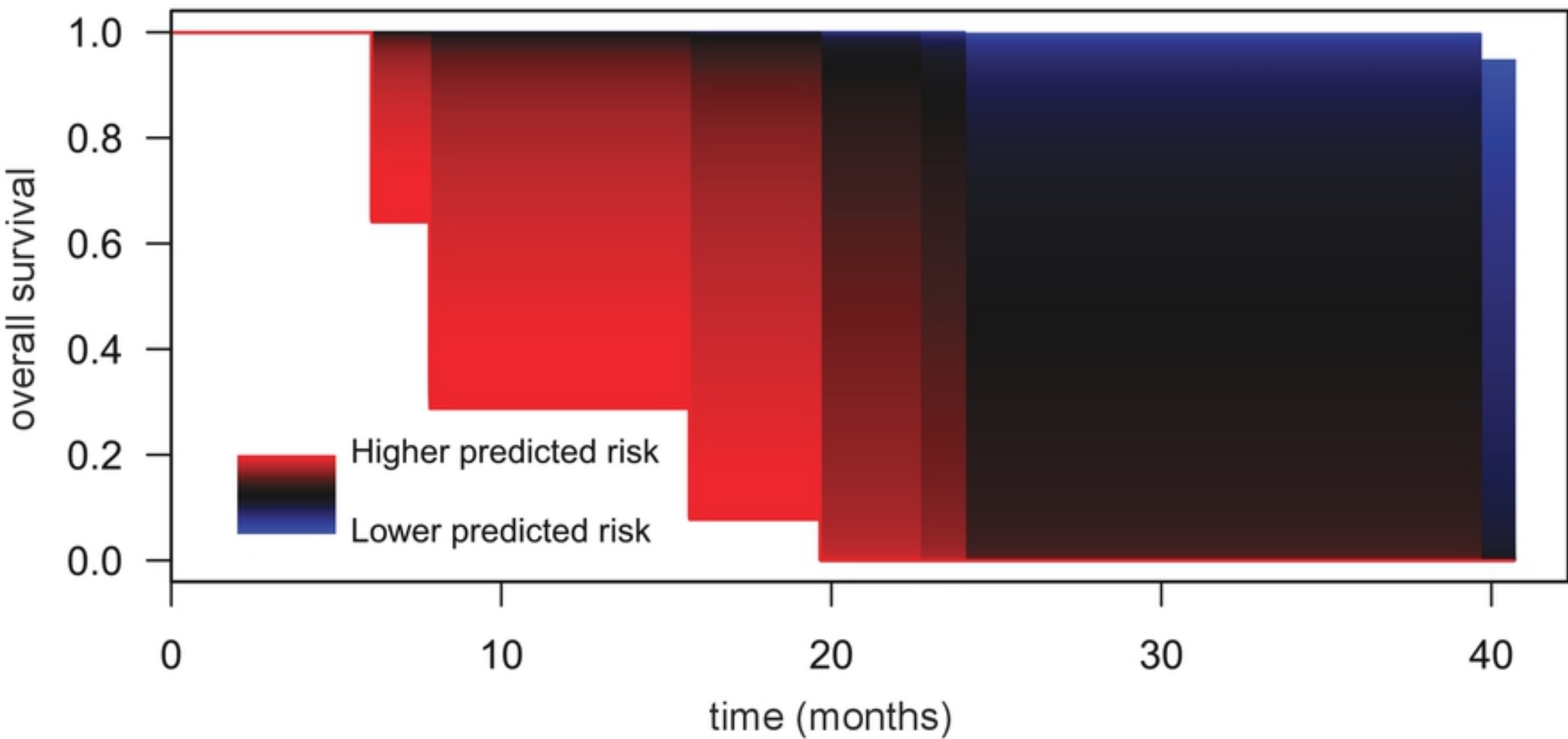


Figure 5