

RESEARCH ARTICLE

Underexpression of Specific Interferon Genes Is Associated with Poor Prognosis of Melanoma

Aamir Zainulabadeen^{1,2}, Philip Yao^{1,3}, Habil Zare^{1*}

1 Department of Computer Science, Texas State University, San Marcos, Texas, United States of America, **2** Department of Computer Science, Princeton University, Princeton, New Jersey, United States of America, **3** Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, United States of America

☞ These authors contributed equally to this work.

* zare@txstate.edu



OPEN ACCESS

Citation: Zainulabadeen A, Yao P, Zare H (2017) Underexpression of Specific Interferon Genes Is Associated with Poor Prognosis of Melanoma. PLoS ONE 12(1): e0170025. doi:10.1371/journal.pone.0170025

Editor: Roger Chammas, Universidade de Sao Paulo, BRAZIL

Received: September 26, 2016

Accepted: December 26, 2016

Published: January 23, 2017

Copyright: © 2017 Zainulabadeen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Alternatively, data can also be accessed at the following link: https://tcga-data.nci.nih.gov/docs/publications/skcm_2015/ Alternatively, data can be accessed from TCGA Data Portal or from the following page (doi:10.1016/j.cell.2015.05.044): https://tcga-data.nci.nih.gov/docs/publications/skcm_2015/.

Funding: This study was supported by the National Science Foundation (NSF) through the Research Experiences for Undergraduates (REU) program

Abstract

Because the prognosis of melanoma is challenging and inaccurate when using current clinical approaches, clinicians are seeking more accurate molecular markers to improve risk models. Accordingly, we performed a survival analysis on 404 samples from The Cancer Genome Atlas (TCGA) cohort of skin cutaneous melanoma. Using our recently developed gene network model, we identified biological signatures that confidently predict the prognosis of melanoma (p -value $< 10^{-5}$). Our model predicted 38 cases as low-risk and 54 cases as high-risk. The probability of surviving at least 5 years was 64% for low-risk and 14% for high-risk cases. In particular, we found that the overexpression of specific genes in the *mitotic cell cycle pathway* and the underexpression of specific genes in the *interferon pathway* are both associated with poor prognosis. We show that our predictive model assesses the risk more accurately than the traditional Clark staging method. Therefore, our model can help clinicians design treatment strategies more effectively. Furthermore, our findings shed light on the biology of melanoma and its prognosis. This is the first *in vivo* study that demonstrates the association between the *interferon pathway* and the prognosis of melanoma.

Introduction

Cutaneous melanoma is a malignancy of melanocytes. It is the most common type of skin cancer. The American Cancer Society estimates that over 73,000 new cases were diagnosed in 2015 in the United States and about 10,000 deaths are caused by melanoma each year [1]. The prognosis of melanoma is highly variable [2]. For instance, the 5-year overall survival rate can be as high as 97% for stage I and as low as 3% for stage IV [3, 4]. Almost all common treatment options for melanoma, including surgery, chemotherapy, and radiation therapy, have harmful and severe side effects. Therefore, it is critical to identify patients who are not at a significant risk of metastasis and death due to the disease. The predictive power of clinical factors is limited [3, 5, 6] (e.g., staging based on the tumor size and the number of metastatic sentinel

(CNS1358939), and also through the infrastructure grant (CRI 1305302).

Competing Interests: The authors have declared that no competing interests exist.

lymph nodes [7]), therefore clinicians are seeking more accurate molecular markers to improve risk models and to avoid unnecessary treatment of low-risk patients [8–10].

Gene expression profile signatures have useful information on the molecular status of cells and they can predict the prognosis of many cancers [11–14], including melanoma [10, 15–18]. For example, Onken *et al.* discovered a gene expression prognostic signature that significantly improved the classification of uveal melanoma compared to traditional staging [15]. That is, they showed that the deregulation of the *LEDA*, *FZD6*, and *ENPP2* genes predicts metastatic death (p-value < 10⁻⁴). In the follow-up studies, they extended their test to include 15 genes [19] and their extended test correctly classified 446 (97%) of the 459 studied cases into low-risk (i.e., at least 95% chance of 5-year metastasis-free survival) and high-risk (i.e., not more than 20% chance of 5-year metastasis-free survival) groups [20].

Recently, Gerami *et al.* performed a meta-analysis on several published genomic analyses of cutaneous melanoma tumors [15, 21–27]. Based on the gene ontology [28] of the frequently reported genes, they identified 28 discriminant genes including *BAP1*, *MGP*, *SPPI*, *CXCL14*, *CLCA2*, *S100A8*, *BTG1*, *SAP130*, *ARG1*, *KRT6B*, *GJA1*, *ID2*, *EIF1B*, *S100A9*, *CRABP2*, *KRT14*, *ROBO1*, *RBM23*, *TACSTD2*, *DSC1*, *SPRR1B*, *TRIM29*, *AQP3*, *TYRP1*, *PPL*, *LTA4H*, and *CST6* [4]. They used the expression of these genes to train a generalized linear mixed model with a radial basis function kernel. The resulting signature classified 268 primary cutaneous melanoma tumors into low-risk and high-risk groups, with 5-year disease-free survival rates of 97% and 31%, respectively [4]. This test is commercially available as a diagnostic tool called DecisionDx-Melanoma [29, 30], but it is not yet recommended by the National Comprehensive Cancer Network [10]. Its benefits to the patients must be confirmed using prospective clinical trials that include significantly larger cohorts with more representative metastatic characteristics [18, 31].

We hypothesized that there is room for significant improvement in the melanoma prognostic tests through rigid, unbiased, and comprehensive analysis of gene expression profiles [10, 32]. Accordingly, we applied a robust large-scale network analysis to the gene expression data of 404 samples from The Cancer Genome Atlas (TCGA) cohort of skin cutaneous melanoma [33]. The aim of our study was to identify the molecular signatures that predict the prognosis of melanoma and to segregate patients into low-, medium-, and high-risk groups. Our approach is based on co-expression network analysis, and we use *eigengenes* as informative prognostic signatures [34].

Materials and Methods

The TCGA Dataset

Several mRNA expression profiling datasets have been produced to study melanoma prognosis [10, 23, 27, 33, 35–40]. We used the TCGA2STAT package to download gene expression data from The Cancer Genome Atlas (TCGA) repository [41]. Specifically, we downloaded RNA-Seq data from the skin cutaneous melanoma cohort of 473 patients [33] and used RPKM values (i.e., reads per kilobase of transcript per million mapped reads) [42] as a measure of gene expression. We manually downloaded the corresponding clinical data, including: a) information regarding the last status of each case (i.e. whether the event of disease recurrence or progression occurred), b) the length of the disease-free time period (i.e., the time from the initial melanoma diagnosis until this event or until the last follow-up date if the event did not occur), and c) the Clark scale stage of the melanoma, which was determined using clinicopathological features such as the size, number, and location of metastases [43].

We computed the Spearman's rank correlation between the disease-free time and gene expression [44]. Spearman's rank correlation is more robust than Pearson correlation and it is

the recommended approach for skewed distributions [45]. For instance, Mukaka showed that after removal of the outliers, the change in the Spearman's correlation can be negligible unlike the Pearson correlation [46]. Consistent with the approach taken by other scholars [47], we used only the top third of genes (6,834) that were most correlated with the disease-free time in our network analysis. We considered any progressed or recurred tumor as *high risk* ($n = 263$). In the rest of the tumors, which were not reported to recur during the follow-up period, any case that had at least five years of follow-up data was considered *low-risk* ($n = 33$) [48–51]. All results presented here can be conveniently reproduced using our supplementary code (S4 File).

Validation datasets

To confirm the findings that we obtained using the TCGA dataset, we validated them using two independent datasets. Specifically, we used the Leeds melanoma gene expression set 1 [39], which is publicly available through European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL–EBI, accession number: E-MTAB-4725). For brevity, we refer to this dataset as LEED in this paper. This cohort comprises whole-genome mRNA expression of 204 primary melanoma tumors, which are measured using Illumina DASL HT12.4. Kolesnikov *et al.* normalized the gene expression values with quantile method after background correction. We manually downloaded the gene expression and clinical data from the EMBL–EBI ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4725/>) [52]. We used the “last follow up” (time) and “viability” (death event) columns from the clinical data.

We also used a similar cohort that was produced at the Lund University in 2015 [40]. This dataset is publicly available through EMBL–EBI (accession number: E-GEOD-65904) and also through Gene Expression Omnibus (GEO) (accession number: GSE65904) [53]. For brevity, we refer to this dataset as LUND in this paper. Cirenajwis *et al.* extracted total RNA from 214 fresh-frozen melanoma tumors and performed genome-wide expression profiling using Illumina Human HT-12V4.0 BeadChip arrays. We downloaded the gene expression data using GEOquery package (Version 2.40.0) [54]. We downloaded the corresponding clinical data from the EMBL–EBI database (<http://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-65904/>) and used the “disease specific survival” (time) and “disease specific death” (event) columns.

Gene network analysis

We applied the weighted gene co-expression network analysis (WGCNA) (Version 1.51) to all 473 available samples to build a gene network and to cluster the genes into gene modules (clusters) [55]. Specifically, we used the `calculate.beta` function with the default parameters to infer that the soft-thresholding power for network construction was 7. We identified 13 gene modules using the `wgcna.one.step` function from the Pigengene package [34], which is a wrapper for the `blockwiseModules` function, with `power = 7` and left the remaining arguments as defaults. WGCNA could not confidently assign 1,404 genes to any of the modules, because these genes had little correlation with the other genes. We call the set of these outlier genes Module 0.

Computing eigengenes

An eigengene of a module is a weighted average of the expression of all the genes in that module. These weights are adjusted so that the loss of biological information is minimized [56, 57]. We used principal component analysis (PCA) to compute eigengenes. First, we balanced the

number of high-risk and low-risk cases using oversampling so that both groups had comparable representatives in the analysis. Specifically, we repeated the data of each high-risk and low-risk case 6 times and 45 times, respectively. This approach provided us with 1,485 and 1,578 samples from each group, respectively. Oversampling was necessary for computing the eigengene of a module. Because an eigengene is the first principal component of the module, it would be biased towards the high-risk group, which has around eight times more samples than the low-risk group. Oversampling resolves this issue. Then, we applied the `moduleEigengenes()` function from the WGCNA package to the oversampled data. This function computed the first principal component of each module, which maximized the explained variance, thus ensuring a minimum loss of biological information. We used the `project.eigen` function from the Pigengene package (Version 0.99.23) to infer the values of eigengenes for all of the 473 samples in the TCGA, the 204 samples in the LEEDS, and the 214 samples in the LUNDS datasets (S1 File) [34].

Survival analysis

We used the 14 inferred eigengenes as covariates (prognostic features), and we included only the 404 samples for which the final status and the survival time were available. We used the `glmnet()` function from the `glmnet` package (Version 2.0-5) [58] to perform a penalized Cox regression analysis [59, 60]. We set $\alpha = 1$ to use the least absolute shrinkage and selection operator (Lasso) [61]. The Lasso, also known as L_1 regularization, enforces most of the coefficients of the covariates (eigengenes) in the Cox proportional hazards model to be zero. Thereby, it identifies the modules that are the most associated with survival.

To evaluate the significance of the selected modules in predicting the survival time, we fitted an accelerated failure time (AFT) model to the selected eigengenes [62]. We used the `survreg` function from the `survival` package (Version 2.39-4) [63], set the Weibull distribution with `scale = 1` as the baseline hazard function, and used the default values for the rest of the parameters. We used the fitted accelerated failure time model to predict the survival time of each sample. We chose two thresholds for the predicted values that maximized the precision of low- and high-risk predictions. The samples that had a predicted survival time between the two thresholds were considered medium-risk. We used the `survfit` function to obtain a Kaplan-Meier survival curve for each of the risk groups [64]. We used the `survdiff` function to test whether the survival curves that correspond to high-risk and low-risk groups differ significantly. This function computed the log-rank p-value of the corresponding Mantel-Haenszel test [65].

Cross-validation

We performed 5-fold cross-validation to confirm that the selection of the modules by the penalized Cox regression is robust with respect to choosing the samples. We used all of the 14 eigengenes corresponding to the modules that were identified by our gene network analysis. We did not recompute the modules or eigengenes, instead, we repeated the penalized Cox regression model as follows. There were 404 samples for which the final status and the survival time were available. We randomly divided these 404 samples into five divisions. These divisions had an almost equal size and equal number of high-risk samples. We set aside one division and performed a penalized Cox regression analysis on the rest of the samples using *all* 14 eigengenes. We recorded the selected modules and repeated this procedure five times. We ran this experiment 10 times with different seeds and counted the frequency of the selected modules in each run (S2 File). On average, the three most frequent modules were selected 4.9, 4.8,

and 4.2 times, respectively. In contrast, all other modules were selected 0.6 times or less, on average.

Results

Using coexpression network analysis, we identified 13 modules of highly coexpressed genes. The size of these modules ranges from 48 to 2,247 genes with a mean of 418, a median of 88, and a standard deviation of 629 (S1 Fig). We computed an eigengene for each module, which summarizes the biological information of the module into one value per sample. We used these eigengenes as biological signatures (features) to perform a survival analysis.

The penalized Cox regression consistently selected three modules as the most associated modules with disease-free survival (Methods). These three gene modules include the *outlier module* (with 1,404 genes, ME0), the ninth largest module (with 58 genes, ME9), and the twelfth largest module (with 52 genes, ME12) (S3 File). The outlier module consists of genes that have too small a correlation with other genes to be included in any of the modules. Hypergeometric tests revealed that the other two selected modules are associated with the *mitotic cell cycle* and the *interferon (IFN) pathway*, respectively [66] (S2 Fig). That is, 15 (29%) of the 58 genes in the larger module are members of the Reactome mitotic cell cycle pathway, which has 454 genes (adjusted p-value < 10^{-9}) [67]. These genes include *AURKB*, *CCNE1*, *CDCA8*, *CDK4*, *CENPO*, *GINS2*, *H2AFZ*, *LIG1*, *PKMYT1*, *PLK1*, *PTTG1*, *SKA1*, *TUBA1B*, *TUBA1C*, and *TYMS*. According to the Cox model, overexpression of these genes is associated with poor prognosis, which is expected [68, 69].

The smaller module (ME12) has 52 genes. Interestingly, 16 (31%) of these genes are members of the Reactome interferon signaling pathway, which has 158 genes (adjusted p-value < 10^{-21}). The genes in the overlap include *DDX58*, *EIF2AK2*, *GBP3*, *HERC5*, *IFIT1*, *IFIT2*, *IFIT3*, *OAS1*, *OAS2*, *OAS3*, *PSMB8*, *SP100*, *STAT1*, *UBE2L6*, *USP18*, and *XAF1*. Our model indicates that the relatively higher expression of these genes is associated with a good prognosis in melanoma. Our *in silico* overrepresentation analysis showed that *type I interferon signaling pathway* (Gene Ontology accession number: 60337 [28]), which includes 63 genes, is the biological processes that has a significant overlap with this gene set. Specifically, the overlap consists of 12 genes, which is 75 times more than expected (p-value = 10^{-15}) [70]. This module is very stable with respect to selection of samples. To confirm this, we reconstructed the coexpression network 10 times using only 426 (90%) randomly selected samples. All of the resulting networks had a similar module, i.e., mean and median of the Jaccard [71] (Tanimoto [72]) similarity were 0.92 and 0.93, respectively.

To further validate the association of these three selected modules with melanoma prognosis, we fitted an accelerated failure time (AFT) model to the corresponding eigengenes, and classified the patients into low-, medium-, and high-risk groups (Methods) [62]. We compared the Kaplan-Meier (KM) curves of these groups [64] (Fig 1). Our AFT model predicted that 38 cases were low-risk. These cases had a significantly higher survival rate than the 54 cases that were predicted to be high-risk (log-rank p-value < 8×10^{-6} [73]). For instance, the probability of surviving for at least five years was 0.64 for low-risk and 0.14 for high-risk cases. Excluding the interferon module has a negative impact on the statistical power and the p-value (Fig 1). Specifically, while the number of cases in the predicted low-risk group does not increase dramatically (i.e., two cases, only 5% improvement), the number of predicted high-risk cases decreases considerably to 35 (i.e., 35% decline). That is, using the interferon pathway module, the model can identify more high-risk cases without sacrificing the accuracy.

As expected, the cases that were predicted to be high-risk had generally more advanced disease according to the traditional Clark scale stage of melanoma (Table 1). In particular, the

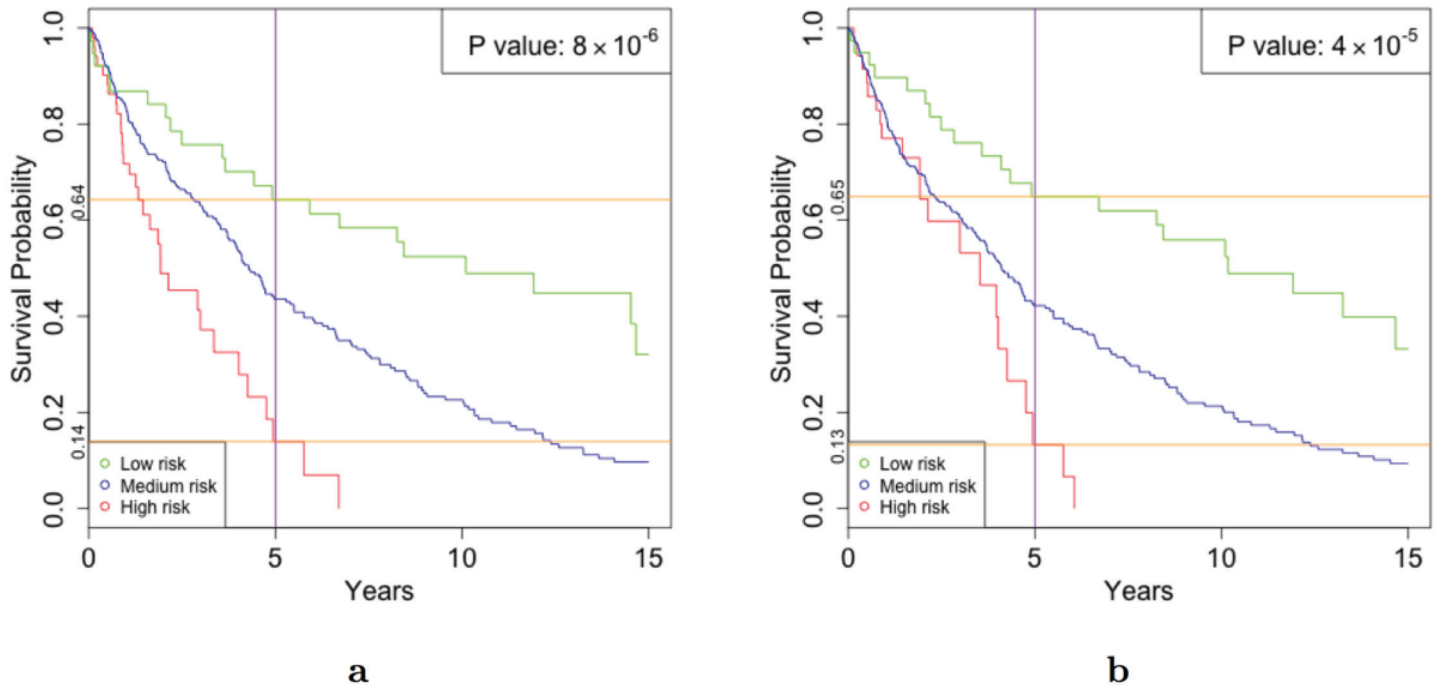


Fig 1. Kaplan–Meier survival curves. The p-values indicate that the difference between the low–risk group (green) and the high–risk group (red) is statistically significant. Using all the three modules, which are associated with the interferon pathway, mitotic cell cycle, and outliers; results in a better p-value (a) compared to a model without the interferon pathway (b). The orange horizontal lines indicate that both models have similar accuracies. However, including the interferon pathway improves the p-value, because more samples are classified in total (i.e., 38 low–risk plus 54 high–risk cases in (a), compared to 40 low–risk plus 35 high–risk cases in (b)).

doi:10.1371/journal.pone.0170025.g001

majority of the high–risk cases were in stage IV or V (36 cases, 66%). In contrast, only 13 (35%) of the low–risk cases were in stage IV or V. Interestingly, 6 (46%) of these 13 cases were disease–free for more than eight years, which indicates that, for these cases, the Clark scale staging is less accurate than our predictions. Also, 7 cases had stage I, II, or III melanoma, but they were classified as high–risk by our model. Only 2 (29%) of these cases survived more than 2 years.

To further validate the association between the interferon module and melanoma prognosis, we used the LEEDS and LUND gene expression datasets, which include 204 and 214 melanoma samples, respectively. We inferred the eigengenes and fitted an AFT model to the eigengenes to classify the samples into low–, medium–, and high–risk groups in each dataset

Table 1. The distribution of melanoma Clark stages in each risk group. Also, the percentage of each stage class in each risk group is shown. The low–risk group is enriched in patients at stages III and IV. The high–risk group is enriched in patients at stages IV and V.

Stage	I (%)	II (%)	III (%)	IV (%)	V (%)	Unknown (%)	Total
Low–risk	0 (0)	1 (3)	11 (29)	12 (32)	1 (3)	13 (34)	38
Medium–risk	4 (1)	16 (5)	54 (17)	105 (34)	37 (12)	96 (31)	312
High–risk	1 (2)	0 (0)	6 (11)	24 (44)	12 (22)	11 (20)	54
Total	5	17	71	141	50	120	404

doi:10.1371/journal.pone.0170025.t001

(Methods). Similar to the analysis on TCGA dataset, we used the three eigengenes corresponding to the outlier, *mitotic cell cycle* and *interferon pathway* modules.

Our AFT model predicted 96 cases in the LUND dataset to be low-risk and 45 cases to be high-risk. The predicted low-risk group had a significantly higher survival rate than the high-risk group (log-rank p-value 2×10^{-3} , S3a Fig). Excluding the *interferon pathway* module from the AFT model resulted in reducing the number of predicted low-risk to 49 cases and predicted high-risk to 25 cases, and also, a less significant p-value (3×10^{-2}). That is, without the *interferon pathway*, the number of predicted low- and high-risk samples would be reduced by almost half leaving 67 (33%) more samples in the medium-risk group. This shows the necessity of the *interferon pathway* module in predicting the prognosis. The results in the LEEDS dataset followed a similar pattern. Specifically, excluding the *interferon pathway* module from the AFT model resulted in reducing the number of predicted low-risk samples from 34 cases to 31 cases while slightly decreasing the number of predicted high-risk samples from 29 to 28. Nevertheless, the log-rank p-value increased by almost two orders of magnitude, from 9×10^{-6} to 7×10^{-4} , indicating that the prediction of survival is less accurate without the *interferon pathway* module (S3c and S3d Fig).

Discussion

Predicting the prognosis of melanoma is clinically useful and important [10]. To date, most of the studies that aim at predicting melanoma survival based on gene expression have been limited in their number of genes, number of samples, or their follow-up time [4, 15–27, 30, 74]. We performed gene network analysis on 470 melanoma cases to extend the previous studies and to identify novel prognostic signatures.

This is the first study to show that the underexpression of specific genes from the interferon pathway in melanoma tissues is a sign of poor prognosis. The role of the interferon pathway in other cancers were studied by others [75–80]. In general, defects in interferon signaling results in dysfunction of the immune system [81]. However, its association with melanoma was previously shown only *in vitro* [81–84].

Interestingly, our interferon module has 17 genes in common with the 274 genes that Hoek *et al.* reported to be downregulated in melanoma cell lines [82]. This is a significant overlap (p-value of the hypergeometric test $< 10^{-19}$). Similarly, the list of the top 25 genes that Critchley *et al.* reported to be differentially expressed in peripheral blood mononuclear cell (PBMC) samples of melanoma patients has a significant overlap with our interferon module (13 genes, p-value $< 10^{-27}$). Compared to *in vitro* experiments, our analysis provides much stronger evidence for the role of the interferon pathway in melanoma, because our study is based on the survival analysis of a relatively large cohort of patients with an extended follow-up time. The total number of patients classified as low-risk or high-risk increases from 75 to 92 (a 23% improvement) when we include the interferon module in our predictive model. This improvement, as well as the decrease in p-value, indicate that the information in the interferon pathway is essential for predicting the prognosis.

However, functional studies will be needed to determine the mechanism and impact of the *interferon pathway* on melanoma prognosis. One challenge in designing such a study is possibly the relatively low number of samples that could be associated with the genes in the interferon module. For example, in our study on the 404 TCGA cases, including these genes in the predictive model led to classification of only 17 (4%) more cases. Therefore, a follow-up functional study most likely needs to investigate at least hundreds of samples in order to include a few samples that are associated with the *interferon pathway*.

The probability of surviving for at least five years is 0.64 and 0.14 for our predicted low-risk and high-risk groups, respectively. The accuracy of our predictive model is comparable with previous studies on skin cutaneous melanoma that were based on gene expression [10]. For instance, Sivendran *et al.* developed a log-rank Mantel-Cox test based on a 21-gene expression signature and ulceration [74]. Their test is less specific than ours in predicting long-term poor prognosis. That is, they tested 48 patients and reported that the probability of surviving for at least five years was 60% and 35% for their predicted low-risk and high-risk groups, respectively.

Gerami *et al.* reported that the DecisionDx-Melanoma test [29, 30] classified 268 primary cutaneous melanoma tumors into low-risk and high-risk groups with at least 5-year disease-free survival rates of 97% and 31%, respectively [4]. Their low-risk specificity is more than ours, and their high-risk specificity is less than ours. However, the reported assessment of the sensitivity and specificity of this test is controversial, because their cohort was not representative of the general primary melanoma patient population [10, 18]. Our results are not comparable with the results reported by Onken *et al.* [20], because they studied uveal melanoma, which is different from skin cutaneous melanoma [85].

One reason underlying the difficulty in assessing the survival rate of melanoma is the relatively high heterogeneity of genetic mutations, which results in subpopulations (clones) that are resistant to therapy [86–88]. Recently, Tirosch *et al.* used single-cell sequencing to show that distinct clones within a tumor can have different gene expression profiles, and therefore, can interact with their microenvironment differently [89]. Investigation of the genes in our interferon module and the corresponding signaling proteins may reveal how the interactions between malignant cells and their microenvironments is modulated. For instance, the relatively high expression of *type I interferon signaling pathway* genes in low-risk melanoma cases can be associated with antitumor response of the immune system [90]. Future work in this direction can leverage available techniques for 1) clonal decomposition based on genetic mutations [91–97], and 2) deconvolution of gene expression profiles into signatures that are specific to a cell-type or tumor clone [98–101].

One limitation of our predictive model for clinical use is the relatively high number of cases that were classified as medium-risk (312, 77%). This can be addressed by improving the predictive model in a follow-up study or, alternatively, by using another prognostic test for the medium-risk cases. Compared to other tests, our predictive model is based on a relatively large number of genes. This is a double-edged sword. The inclusion of a large number of genes makes our model robust with respect to random changes in the expression of one or several genes, which are common due to technical or biological noise. On the other hand, the large number of genes makes our test difficult to apply in clinical settings. This can be addressed by excluding the genes that have a relatively smaller contribution to the eigengenes. Specifically, a greedy algorithm can be used to exclude the genes that have a smaller absolute weight (loading) [102]. Further follow-up experiments on different datasets will be needed to show that such a modification does not affect the accuracy of the model and to ensure that too many genes will not be excluded. Nevertheless, follow-up studies can examine the therapeutic value of the genes identified in our study that have a relatively high contribution to the model, as well as their upstream genes (S3 File).

Conclusion

We identified a specific set of genes in the interferon pathway that are underexpressed in high-risk melanoma. This biological signature, together with the overexpression of other genes in the mitotic cell cycle pathway, predicts the prognosis of melanoma with relatively high accuracy.

Supporting Information

S1 File. The eigengene values in the TCGA dataset useful for reproducing the results.

(XLS)

S2 File. The frequency of selected modules by 5-fold cross-validation of the penalized Cox model.

(XLS)

S3 File. The gene lists corresponding to the three selected modules. The gene symbol, Entrez ID, and the weight of each gene in the corresponding selected module is reported.

(XLS)

S4 File. Supplementary code. Our results can be conveniently reproduced using these R scripts. Uncompress the tarball file, install the packages mentioned in the `settings.R` script, and then source the `runall.R` script. This code works on Unix (Linux or Mac OS X). Data will be downloaded from TCGA and the results will be saved in the current directory. A desktop computer with a 2.8 GHz CPU and 8 GB of memory will reproduce all results in less than an hour.

(ZIP)

S1 Fig. The distribution of the size of modules.

(PNG)

S2 Fig. The overrepresentation analysis on the selected modules.

(PNG)

S3 Fig. The Kaplan–Meier survival curves for the validation datasets. Colors are similar to (Fig 1). In the LUND dataset, including the *interferon pathway* module results in better predictions of the survival time (a) with a more significant p-value of 2×10^{-3} compared to an AFT model that uses only two modules (b). Similarly, in the LEEDS dataset, the model predicts the survival rate better when the *interferon pathway* module is included (c) compared to a model that uses only two modules (d).

(PNG)

Acknowledgments

This study was supported by the National Science Foundation (NSF) through the Research Experiences for Undergraduates (REU) program (CNS1358939), and also through the infrastructure grant (CRI 1305302). We used gene expression data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

Author Contributions

Conceptualization: HZ.

Data curation: HZ.

Formal analysis: AZ PY.

Funding acquisition: HZ.

Investigation: AZ PY.

Methodology: HZ.

Project administration: HZ.

Resources: HZ.

Software: AZ PY.

Supervision: HZ.

Validation: HZ AZ PY.

Visualization: AZ PY.

Writing – original draft: HZ AZ PY.

Writing – review & editing: HZ AZ PY.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA: a cancer journal for clinicians*. 2015; 65(1):5–29.
2. Balch CM, Gershenwald JE, Soong Sj, Thompson JF, Ding S, Byrd DR, et al. Multivariate analysis of prognostic factors among 2,313 patients with stage III melanoma: comparison of nodal micrometastases versus macrometastases. *Journal of Clinical Oncology*. 2010; 28(14):2452–2459. doi: [10.1200/JCO.2009.27.1627](https://doi.org/10.1200/JCO.2009.27.1627) PMID: [20368546](https://pubmed.ncbi.nlm.nih.gov/20368546/)
3. Balch CM, Gershenwald JE, Soong Sj, Thompson JF, Atkins MB, Byrd DR, et al. Final version of 2009 AJCC melanoma staging and classification. *Journal of clinical oncology*. 2009; 27(36):6199–6206. doi: [10.1200/JCO.2009.23.4799](https://doi.org/10.1200/JCO.2009.23.4799) PMID: [19917835](https://pubmed.ncbi.nlm.nih.gov/19917835/)
4. Gerami P, Cook RW, Wilkinson J, Russell MC, Dhillon N, Amaria RN, et al. Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma. *Clinical Cancer Research*. 2015; 21(1):175–183. doi: [10.1158/1078-0432.CCR-13-3316](https://doi.org/10.1158/1078-0432.CCR-13-3316) PMID: [25564571](https://pubmed.ncbi.nlm.nih.gov/25564571/)
5. Sondak VK, Messina JL. Prediction is Difficult, Especially About the Future: Clinical Prognostic Tools in Melanoma. *Annals of surgical oncology*. 2016;p. 1–3.
6. Gimotty PA, Elder DE, Fraker DL, Botbyl J, Sellers K, Elenitsas R, et al. Identification of high-risk patients among those diagnosed with thin cutaneous melanomas. *Journal of Clinical Oncology*. 2007; 25(9):1129–1134. doi: [10.1200/JCO.2006.08.1463](https://doi.org/10.1200/JCO.2006.08.1463) PMID: [17369575](https://pubmed.ncbi.nlm.nih.gov/17369575/)
7. Morton DL, Thompson JF, Cochran AJ, Mozzillo N, Elashoff R, Essner R, et al. Sentinel-node biopsy or nodal observation in melanoma. *New England Journal of Medicine*. 2006; 355(13):1307–1317. doi: [10.1056/NEJMoa060992](https://doi.org/10.1056/NEJMoa060992) PMID: [17005948](https://pubmed.ncbi.nlm.nih.gov/17005948/)
8. Ekmekcioglu S, Davies MA, Tanese K, Roszik J, Shin-Sim M, Bassett RL, et al. Inflammatory Marker Testing Identifies CD74 Expression in Melanoma Tumor Cells, and its Expression Associates with Favorable Survival for Stage III Melanoma. *Clinical Cancer Research*. 2016;p. clincanres–2226. doi: [10.1158/1078-0432.CCR-15-2226](https://doi.org/10.1158/1078-0432.CCR-15-2226) PMID: [26783288](https://pubmed.ncbi.nlm.nih.gov/26783288/)
9. Kimbrough CW, Egger ME, McMasters KM, Stromberg AJ, Martin RC, Philips P, et al. Molecular Staging of Sentinel Lymph Nodes Identifies Melanoma Patients at Increased Risk of Nodal Recurrence. *Journal of the American College of Surgeons*. 2016; 222(4):357–363. doi: [10.1016/j.jamcollsurg.2015.12.042](https://doi.org/10.1016/j.jamcollsurg.2015.12.042) PMID: [26875070](https://pubmed.ncbi.nlm.nih.gov/26875070/)
10. Weiss SA, Hanniford D, Hernando E, Osman I. Revisiting determinants of prognosis in cutaneous melanoma. *Cancer*. 2015; 121(23):4108–4123. doi: [10.1002/cncr.29634](https://doi.org/10.1002/cncr.29634) PMID: [26308244](https://pubmed.ncbi.nlm.nih.gov/26308244/)
11. Francis P, Namløs HM, Müller C, Edén P, Fernebo J, Berner JM, et al. Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential. *BMC genomics*. 2007; 8(1):1. doi: [10.1186/1471-2164-8-73](https://doi.org/10.1186/1471-2164-8-73)
12. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine*. 2004; 351(27):2817–2826. doi: [10.1056/NEJMoa041588](https://doi.org/10.1056/NEJMoa041588) PMID: [15591335](https://pubmed.ncbi.nlm.nih.gov/15591335/)
13. Colman H, Zhang L, Sulman EP, McDonald JM, Shooshtari NL, Rivera A, et al. A multigene predictor of outcome in glioblastoma. *Neuro-oncology*. 2009;p. nop007.
14. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Richards WG, et al. Using gene expression ratios to predict outcome among patients with mesothelioma. *Journal of the National Cancer Institute*. 2003; 95(8):598–605. doi: [10.1093/jnci/95.8.598](https://doi.org/10.1093/jnci/95.8.598) PMID: [12697852](https://pubmed.ncbi.nlm.nih.gov/12697852/)

15. Onken MD, Worley LA, Ehlers JP, Harbour JW. Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death. *Cancer research*. 2004; 64(20):7205–7209. doi: [10.1158/0008-5472.CAN-04-1750](https://doi.org/10.1158/0008-5472.CAN-04-1750) PMID: [15492234](https://pubmed.ncbi.nlm.nih.gov/15492234/)
16. Lee CY, Gerami P. Molecular techniques for predicting behaviour in melanocytic neoplasms. *Pathology*. 2016; 48(2):142–146. doi: [10.1016/j.pathol.2015.12.004](https://doi.org/10.1016/j.pathol.2015.12.004) PMID: [27020386](https://pubmed.ncbi.nlm.nih.gov/27020386/)
17. March J, Hand M, Truong A, Grossman D. Practical application of new technologies for melanoma diagnosis: Part II. Molecular approaches. *Journal of the American Academy of Dermatology*. 2015; 72(6):943–958. doi: [10.1016/j.jaad.2015.02.1138](https://doi.org/10.1016/j.jaad.2015.02.1138) PMID: [25980999](https://pubmed.ncbi.nlm.nih.gov/25980999/)
18. Ji AL, Bichakjian CK, Swetter SM. Molecular Profiling in Cutaneous Melanoma. *Journal of the National Comprehensive Cancer Network*. 2016; 14(4):475–480. PMID: [27059194](https://pubmed.ncbi.nlm.nih.gov/27059194/)
19. Onken MD, Worley LA, Tuscan MD, Harbour JW. An accurate, clinically feasible multi-gene expression assay for predicting metastasis in uveal melanoma. *The Journal of Molecular Diagnostics*. 2010; 12(4):461–468. doi: [10.2353/jmoldx.2010.090220](https://doi.org/10.2353/jmoldx.2010.090220) PMID: [20413675](https://pubmed.ncbi.nlm.nih.gov/20413675/)
20. Onken MD, Worley LA, Char DH, Augsburger JJ, Correa ZM, Nudleman E, et al. Collaborative Ocular Oncology Group report number 1: prospective validation of a multi-gene prognostic assay in uveal melanoma. *Ophthalmology*. 2012; 119(8):1596–1603. doi: [10.1016/j.ophtha.2012.02.017](https://doi.org/10.1016/j.ophtha.2012.02.017) PMID: [22521086](https://pubmed.ncbi.nlm.nih.gov/22521086/)
21. Jaeger J, Koczan D, Thiesen HJ, Ibrahim SM, Gross G, Spang R, et al. Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. *Clinical cancer research*. 2007; 13(3):806–815. doi: [10.1158/1078-0432.CCR-06-1820](https://doi.org/10.1158/1078-0432.CCR-06-1820) PMID: [17289871](https://pubmed.ncbi.nlm.nih.gov/17289871/)
22. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*. 2000; 406(6795):536–540. doi: [10.1038/35020115](https://doi.org/10.1038/35020115) PMID: [10952317](https://pubmed.ncbi.nlm.nih.gov/10952317/)
23. Haqq C, Nosrati M, Sudilovsky D, Crothers J, Khodabakhsh D, Pulliam BL, et al. The gene expression signatures of melanoma progression. *Proceedings of the National Academy of Sciences*. 2005; 102(17):6092–6097. doi: [10.1073/pnas.0501564102](https://doi.org/10.1073/pnas.0501564102)
24. Scatolini M, Grand MM, Grosso E, Venesio T, Pisacane A, Balsamo A, et al. Altered molecular pathways in melanocytic lesions. *International Journal of Cancer*. 2010; 126(8):1869–1881. PMID: [19795447](https://pubmed.ncbi.nlm.nih.gov/19795447/)
25. Smith AP, Hoek K, Becker D. Whole-genome expression profiling of the melanoma progression pathway reveals marked molecular differences between nevi/melanoma in situ and advanced-stage melanomas. *Cancer biology & therapy*. 2005; 4(9):1018–1029. doi: [10.4161/cbt.4.9.2165](https://doi.org/10.4161/cbt.4.9.2165)
26. Weeraratna AT, Becker D, Carr KM, Duray PH, Rosenblatt KP, Yang S, et al. Generation and analysis of melanoma SAGE libraries: SAGE advice on the melanoma transcriptome. *Oncogene*. 2004; 23(12):2264–2274. doi: [10.1038/sj.onc.1207337](https://doi.org/10.1038/sj.onc.1207337) PMID: [14755246](https://pubmed.ncbi.nlm.nih.gov/14755246/)
27. Winnepenninckx V, Lazar V, Michiels S, Dessen P, Stas M, Alonso SR, et al. Gene expression profiling of primary cutaneous melanoma and clinical outcome. *Journal of the National Cancer Institute*. 2006; 98(7):472–482. doi: [10.1093/jnci/djj103](https://doi.org/10.1093/jnci/djj103) PMID: [16595783](https://pubmed.ncbi.nlm.nih.gov/16595783/)
28. Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genetics*. 2000; 25(1):25–9. doi: [10.1038/75556](https://doi.org/10.1038/75556) PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
29. Gerami P, Cook RW, Russell MC, Wilkinson J, Amaria RN, Gonzalez R, et al. Gene expression profiling for molecular staging of cutaneous melanoma in patients undergoing sentinel lymph node biopsy. *Journal of the American Academy of Dermatology*. 2015; 72(5):780–785. doi: [10.1016/j.jaad.2015.01.009](https://doi.org/10.1016/j.jaad.2015.01.009) PMID: [25748297](https://pubmed.ncbi.nlm.nih.gov/25748297/)
30. Berger AC, Davidson RS, Poitras JK, Chabra I, Hope R, Brackeen A, et al. Clinical impact of a 31-gene expression profile test for cutaneous melanoma in 156 prospectively and consecutively tested patients. *Current medical research and opinion*. 2016;(just-accepted):1–23.
31. Cassarino DS, Lewine N, Cole D, Wade B, Gustavsen G. Budget impact analysis of a novel gene expression assay for the diagnosis of malignant melanoma. *Journal of medical economics*. 2014; 17(11):782–791. doi: [10.3111/13696998.2014.950421](https://doi.org/10.3111/13696998.2014.950421) PMID: [25170544](https://pubmed.ncbi.nlm.nih.gov/25170544/)
32. Mahar AL, Compton C, Halabi S, Hess KR, Gershenwald JE, Scolyer RA, et al. Critical assessment of clinical prognostic tools in melanoma. *Annals of surgical oncology*. 2016;p. 1–9. doi: [10.1245/s10434-016-5212-5](https://doi.org/10.1245/s10434-016-5212-5) PMID: [27052645](https://pubmed.ncbi.nlm.nih.gov/27052645/)
33. Network CGA, et al. Genomic classification of cutaneous melanoma. *Cell*. 2015; 161(7):1681–1696. doi: [10.1016/j.cell.2015.05.044](https://doi.org/10.1016/j.cell.2015.05.044)
34. Zare H, et al. Pigengene: Computing and using eigengenes. *Bioconductor*; 2016. Available from: <https://bioconductor.org/packages/devel/bioc/html/Pigengene.html>.

35. Talantov D, Mazumder A, Jack XY, Briggs T, Jiang Y, Backus J, et al. Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clinical Cancer Research*. 2005; 11(20):7234–7242. doi: [10.1158/1078-0432.CCR-05-0683](https://doi.org/10.1158/1078-0432.CCR-05-0683) PMID: [16243793](https://pubmed.ncbi.nlm.nih.gov/16243793/)
36. Riker AI, Enkemann SA, Fodstad O, Liu S, Ren S, Morris C, et al. The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. *BMC medical genomics*. 2008; 1(1):1. doi: [10.1186/1755-8794-1-13](https://doi.org/10.1186/1755-8794-1-13)
37. Brunner G, Reitz M, Schwipper V, Tilkorn H, Lippold A, Biess B, et al. Increased expression of the tumor suppressor PLZF is a continuous predictor of long-term survival in malignant melanoma patients. *Cancer biotherapy & radiopharmaceuticals*. 2008; 23(4):451–460. doi: [10.1089/cbr.2008.0473](https://doi.org/10.1089/cbr.2008.0473)
38. Brunner G, Reitz M, Heinecke A, Lippold A, Berking C, Suter L, et al. A nine-gene signature predicting clinical outcome in cutaneous melanoma. *Journal of cancer research and clinical oncology*. 2013; 139(2):249–258. doi: [10.1007/s00432-012-1322-z](https://doi.org/10.1007/s00432-012-1322-z) PMID: [23052696](https://pubmed.ncbi.nlm.nih.gov/23052696/)
39. Nsengimana J, Laye J, Filia A, Walker C, Jewell R, Van den Oord JJ, et al. Independent replication of a melanoma subtype gene signature and evaluation of its prognostic value and biological correlates in a population cohort. *Oncotarget*. 2015; 6(13):11683. doi: [10.18632/oncotarget.3549](https://doi.org/10.18632/oncotarget.3549) PMID: [25871393](https://pubmed.ncbi.nlm.nih.gov/25871393/)
40. Cirenajwis H, Ekedahl H, Lauss M, Harbst K, Carneiro A, Enoksson J, et al. Molecular stratification of metastatic melanoma using gene expression profiling: prediction of survival outcome and benefit from molecular targeted therapy. *Oncotarget*. 2015;. doi: [10.18632/oncotarget.3655](https://doi.org/10.18632/oncotarget.3655)
41. Wan YW, Allen GI, Liu Z. TCGA2STAT: simple TCGA data access for integrated statistical analysis in R. *Bioinformatics*. 2015;p. btv677. PMID: [26568634](https://pubmed.ncbi.nlm.nih.gov/26568634/)
42. Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature biotechnology*. 2014; 32(5):462–464. doi: [10.1038/nbt.2862](https://doi.org/10.1038/nbt.2862) PMID: [24752080](https://pubmed.ncbi.nlm.nih.gov/24752080/)
43. Clark WH, Elder DE, Guerry D, Braitman LE, Trock BJ, Schultz D, et al. Model predicting survival in stage I melanoma based on tumor progression. *Journal of the National Cancer Institute*. 1989; 81(24):1893–1904. doi: [10.1093/jnci/81.24.1893](https://doi.org/10.1093/jnci/81.24.1893) PMID: [2593166](https://pubmed.ncbi.nlm.nih.gov/2593166/)
44. Daniel WW. Spearman rank correlation coefficient. *Applied nonparametric statistics*, 2nd ed PWS-Kent, Boston. 1990;p. 358–365.
45. Kumari S, Nie J, Chen HS, Ma H, Stewart R, Li X, et al. Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PloS one*. 2012; 7(11): e50411. doi: [10.1371/journal.pone.0050411](https://doi.org/10.1371/journal.pone.0050411) PMID: [23226279](https://pubmed.ncbi.nlm.nih.gov/23226279/)
46. Mukaka M. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*. 2012; 24(3):69–71. PMID: [23638278](https://pubmed.ncbi.nlm.nih.gov/23638278/)
47. Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezchnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013; 153(3):707–720. doi: [10.1016/j.cell.2013.03.030](https://doi.org/10.1016/j.cell.2013.03.030) PMID: [23622250](https://pubmed.ncbi.nlm.nih.gov/23622250/)
48. Bostick PJ, Morton DL, Turner RR, Huynh KT, Wang HJ, Elashoff R, et al. Prognostic significance of occult metastases detected by sentinel lymphadenectomy and reverse transcriptase–polymerase chain reaction in early-stage melanoma patients. *Journal of clinical oncology*. 1999; 17(10):3238–3244. PMID: [10506625](https://pubmed.ncbi.nlm.nih.gov/10506625/)
49. Essner R, Lee JH, Wanek LA, Itakura H, Morton DL. Contemporary surgical treatment of advanced-stage melanoma. *Archives of surgery*. 2004; 139(9):961–967. doi: [10.1001/archsurg.139.9.961](https://doi.org/10.1001/archsurg.139.9.961) PMID: [15381613](https://pubmed.ncbi.nlm.nih.gov/15381613/)
50. Ferrone CR, Panageas KS, Busam K, Brady MS, Coit DG. Multivariate prognostic model for patients with thick cutaneous melanoma: importance of sentinel lymph node status. *Annals of Surgical Oncology*. 2002; 9(7):637–645. doi: [10.1007/BF02574479](https://doi.org/10.1007/BF02574479) PMID: [12167577](https://pubmed.ncbi.nlm.nih.gov/12167577/)
51. Karjalainen JM, Kellokoski JK, Eskelinen MJ, Alhava EM, Kosma VM. Downregulation of transcription factor AP-2 predicts poor survival in stage I cutaneous malignant melanoma. *Journal of clinical oncology*. 1998; 16(11):3584–3591. PMID: [9817279](https://pubmed.ncbi.nlm.nih.gov/9817279/)
52. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, et al. ArrayExpress update x2014;simplifying data submissions. *Nucleic acids research*. 2014;p. gku1057. doi: [10.1093/nar/gku1057](https://doi.org/10.1093/nar/gku1057) PMID: [25361974](https://pubmed.ncbi.nlm.nih.gov/25361974/)
53. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*. 2013; 41(D1):D991–D995. doi: [10.1093/nar/gks1193](https://doi.org/10.1093/nar/gks1193) PMID: [23193258](https://pubmed.ncbi.nlm.nih.gov/23193258/)
54. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007; 23(14):1846–1847. doi: [10.1093/bioinformatics/btm254](https://doi.org/10.1093/bioinformatics/btm254) PMID: [17496320](https://pubmed.ncbi.nlm.nih.gov/17496320/)

55. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008; 9(1):559. doi: [10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559) PMID: [19114008](https://pubmed.ncbi.nlm.nih.gov/19114008/)
56. Jolliffe I. *Principal component analysis*. Wiley Online Library; 2002.
57. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proceedings of the National Academy of Sciences*. 2006; 103(47):17973–17978. doi: [10.1073/pnas.0605938103](https://doi.org/10.1073/pnas.0605938103)
58. Simon N, Friedman J, Hastie T, Tibshirani R, et al. Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of statistical software*. 2011; 39(5):1–13. doi: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05) PMID: [27065756](https://pubmed.ncbi.nlm.nih.gov/27065756/)
59. Cox DR. Regression models and life-tables. In: *Breakthroughs in statistics*. Springer; 1992. p. 527–541.
60. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*. 2005; 21(13):3001–3008. doi: [10.1093/bioinformatics/bti422](https://doi.org/10.1093/bioinformatics/bti422) PMID: [15814556](https://pubmed.ncbi.nlm.nih.gov/15814556/)
61. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)*. 1996;p. 267–288.
62. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. vol. 360. John Wiley & Sons; 2011.
63. Therneau TM, Grambsch PM. *Modeling survival data: extending the Cox model*. Springer Science & Business Media; 2000.
64. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*. 1958; 53(282):457–481. doi: [10.1080/01621459.1958.10501452](https://doi.org/10.1080/01621459.1958.10501452)
65. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies. *J natl cancer inst*. 1959; 22(4):719–748. PMID: [13655060](https://pubmed.ncbi.nlm.nih.gov/13655060/)
66. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic acids research*. 2012;p. gks1147. doi: [10.1093/nar/gks1147](https://doi.org/10.1093/nar/gks1147) PMID: [23180781](https://pubmed.ncbi.nlm.nih.gov/23180781/)
67. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. *Nucleic acids research*. 2014; 42(D1):D472–D477. doi: [10.1093/nar/gkt1102](https://doi.org/10.1093/nar/gkt1102) PMID: [24243840](https://pubmed.ncbi.nlm.nih.gov/24243840/)
68. Nakayama KI, Nakayama K. Ubiquitin ligases: cell-cycle control and cancer. *Nature Reviews Cancer*. 2006; 6(5):369–381. doi: [10.1038/nrc1881](https://doi.org/10.1038/nrc1881) PMID: [16633365](https://pubmed.ncbi.nlm.nih.gov/16633365/)
69. Wang L, Hurley DG, Watkins W, Araki H, Tamada Y, Muthukaruppan A, et al. Cell cycle gene networks are associated with melanoma prognosis. *PLoS one*. 2012; 7(4):e34247. doi: [10.1371/journal.pone.0034247](https://doi.org/10.1371/journal.pone.0034247) PMID: [22536322](https://pubmed.ncbi.nlm.nih.gov/22536322/)
70. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic acids research*. 2016; 44(D1):D336–D342. doi: [10.1093/nar/gkv1194](https://doi.org/10.1093/nar/gkv1194) PMID: [26578592](https://pubmed.ncbi.nlm.nih.gov/26578592/)
71. Jaccard P. *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Impr. Corbaz; 1901.
72. Tanimoto TT. *Elementary mathematical theory of classification and prediction*. 1958;.
73. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports Part 1*. 1966; 50(3):163–170. PMID: [5910392](https://pubmed.ncbi.nlm.nih.gov/5910392/)
74. Sivendran S, Chang R, Pham L, Phelps RG, Harcharik ST, Hall LD, et al. Dissection of Immune Gene Networks in Primary Melanoma Tumors Critical for Antitumor Surveillance of Patients with Stage II–III Resectable Disease. *Journal of Investigative Dermatology*. 2014; 134(8):2202–2211. doi: [10.1038/jid.2014.85](https://doi.org/10.1038/jid.2014.85) PMID: [24522433](https://pubmed.ncbi.nlm.nih.gov/24522433/)
75. Deivendran S, Marzook KH, Pillai MR. The role of inflammation in cervical cancer. In: *Inflammation and Cancer*. Springer; 2014. p. 377–399.
76. Scaruffi P, Morandi F, Gallo F, Stigliani S, Parodi S, Moretti S, et al. Bone marrow of neuroblastoma patients shows downregulation of CXCL12 expression and presence of IFN signature. *Pediatric blood & cancer*. 2012; 59(1):44–51. doi: [10.1002/pbc.23339](https://doi.org/10.1002/pbc.23339)
77. Rautela J, Baschuk N, Jayatilleke K, Hertzog P, Parker B. S-20: Exploiting the type-1 interferon pathway as a biomarker and therapeutic target for metastatic cancer. *Cytokine*. 2014; 70(1):25. doi: [10.1016/j.cyto.2014.07.240](https://doi.org/10.1016/j.cyto.2014.07.240)
78. Lu S, Pardini B, Cheng B, Naccarati A, Huhn S, Vymetalkova V, et al. Single nucleotide polymorphisms within interferon signaling pathway genes are associated with colorectal cancer susceptibility and survival. *PLoS one*. 2014; 9(10):e111061. doi: [10.1371/journal.pone.0111061](https://doi.org/10.1371/journal.pone.0111061) PMID: [25350395](https://pubmed.ncbi.nlm.nih.gov/25350395/)

79. Wrangle J, Wang W, Koch A, Easwaran H, Mohammad HP, Pan X, et al. Alterations of immune response of non-small cell lung cancer with azacytidine. *Oncotarget*. 2013; 4(11):2067–2079. doi: [10.18632/oncotarget.1542](https://doi.org/10.18632/oncotarget.1542) PMID: [24162015](https://pubmed.ncbi.nlm.nih.gov/24162015/)
80. Moerdyk-Schauwecker M, Shah NR, Murphy AM, Hastie E, Mukherjee P, Grdzlishvili VZ. Resistance of pancreatic cancer cells to oncolytic vesicular stomatitis virus: role of type I interferon signaling. *Virology*. 2013; 436(1):221–234. doi: [10.1016/j.virol.2012.11.014](https://doi.org/10.1016/j.virol.2012.11.014) PMID: [23246628](https://pubmed.ncbi.nlm.nih.gov/23246628/)
81. Critchley-Thorne RJ, Yan N, Nacu S, Weber J, Holmes SP, Lee PP. Down-regulation of the interferon signaling pathway in T lymphocytes from patients with metastatic melanoma. *PLoS Med*. 2007; 4(5):e176. doi: [10.1371/journal.pmed.0040176](https://doi.org/10.1371/journal.pmed.0040176) PMID: [17488182](https://pubmed.ncbi.nlm.nih.gov/17488182/)
82. Hoek K, Rimm DL, Williams KR, Zhao H, Ariyan S, Lin A, et al. Expression profiling reveals novel pathways in the transformation of melanocytes to melanomas. *Cancer research*. 2004; 64(15):5270–5282. doi: [10.1158/0008-5472.CAN-04-0731](https://doi.org/10.1158/0008-5472.CAN-04-0731) PMID: [15289333](https://pubmed.ncbi.nlm.nih.gov/15289333/)
83. Kim SH, Gunnery S, Choe JK, Mathews MB. Neoplastic progression in melanoma and colon cancer is associated with increased expression and activity of the interferon-inducible protein kinase, PKR. *Oncogene*. 2002; 21(57):8741–8748. doi: [10.1038/sj.onc.1205987](https://doi.org/10.1038/sj.onc.1205987) PMID: [12483527](https://pubmed.ncbi.nlm.nih.gov/12483527/)
84. Litvin O, Schwartz S, Wan Z, Schild T, Rocco M, Oh NL, et al. Interferon α/β enhances the cytotoxic response of MEK inhibition in melanoma. *Molecular cell*. 2015; 57(5):784–796. doi: [10.1016/j.molcel.2014.12.030](https://doi.org/10.1016/j.molcel.2014.12.030) PMID: [25684207](https://pubmed.ncbi.nlm.nih.gov/25684207/)
85. Belmar-Lopez C, Mancheno-Corvo P, Saornil MA, Baril P, Vassaux G, Quintanilla M, et al. Uveal vs. cutaneous melanoma. Origins and causes of the differences. *Clinical and Translational Oncology*. 2008; 10(3):137–142. doi: [10.1007/s12094-008-0170-4](https://doi.org/10.1007/s12094-008-0170-4) PMID: [18321815](https://pubmed.ncbi.nlm.nih.gov/18321815/)
86. Somasundaram R, Villanueva J, Herlyn M. Intratumoral heterogeneity as a therapy resistance mechanism: role of melanoma subpopulations. *Advances in pharmacology (San Diego, Calif)*. 2012; 65:335. doi: [10.1016/B978-0-12-397927-8.00011-7](https://doi.org/10.1016/B978-0-12-397927-8.00011-7)
87. Anaka M, Hudson C, Lo PH, Do H, Caballero OL, Davis ID, et al. Intratumoral genetic heterogeneity in metastatic melanoma is accompanied by variation in malignant behaviors. *BMC medical genomics*. 2013; 6(1):1. doi: [10.1186/1755-8794-6-40](https://doi.org/10.1186/1755-8794-6-40)
88. Sun Xx, Yu Q. Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacologica Sinica*. 2015; 36(10):1219–1227. doi: [10.1038/aps.2015.92](https://doi.org/10.1038/aps.2015.92) PMID: [26388155](https://pubmed.ncbi.nlm.nih.gov/26388155/)
89. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016; 352(6282):189–196. doi: [10.1126/science.aad0501](https://doi.org/10.1126/science.aad0501) PMID: [27124452](https://pubmed.ncbi.nlm.nih.gov/27124452/)
90. Zitvogel L, Galluzzi L, Kepp O, Smyth MJ, Kroemer G. Type I interferons in anticancer immunity. *Nature Reviews Immunology*. 2015; 15(7):405–414. doi: [10.1038/nri3845](https://doi.org/10.1038/nri3845) PMID: [26027717](https://pubmed.ncbi.nlm.nih.gov/26027717/)
91. Fischer A, Vázquez-García I, Illingworth CJ, Mustonen V. High-definition reconstruction of clonal composition in cancer. *Cell reports*. 2014; 7(5):1740–1752. doi: [10.1016/j.celrep.2014.04.055](https://doi.org/10.1016/j.celrep.2014.04.055) PMID: [24882004](https://pubmed.ncbi.nlm.nih.gov/24882004/)
92. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. *Nature methods*. 2014;.
93. Zare H, Wang J, Hu A, Weber K, Smith J, Nickerson D, et al. Inferring clonal composition from multiple sections of a breast cancer. *PLoS computational biology*. 2014; 10(7):e1003703. doi: [10.1371/journal.pcbi.1003703](https://doi.org/10.1371/journal.pcbi.1003703) PMID: [25010360](https://pubmed.ncbi.nlm.nih.gov/25010360/)
94. Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood*. 2012; 120(20):4191–4196. doi: [10.1182/blood-2012-05-433540](https://doi.org/10.1182/blood-2012-05-433540) PMID: [22915640](https://pubmed.ncbi.nlm.nih.gov/22915640/)
95. Marass F, Mouliere F, Yuan K, Rosenfeld N, Markowitz F. A phylogenetic latent feature model for clonal deconvolution. *arXiv preprint arXiv:160401715*. 2016;.
96. Matsui Y, Niida A, Uchi R, Mimori K, Miyano S, Shimamura T. phyC: Clustering cancer evolutionary trees. *bioRxiv*. 2016;p. 069302.
97. Jiang Y, Qiu Y, Minn AJ, Zhang NR. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proceedings of the National Academy of Sciences*. 2016; 113(37):E5528–E5537. doi: [10.1073/pnas.1522203113](https://doi.org/10.1073/pnas.1522203113)
98. Newman AM, Alizadeh AA. High-throughput genomic profiling of tumor-infiltrating leukocytes. *Current Opinion in Immunology*. 2016; 41:77–84. doi: [10.1016/j.coi.2016.06.006](https://doi.org/10.1016/j.coi.2016.06.006) PMID: [27372732](https://pubmed.ncbi.nlm.nih.gov/27372732/)
99. Gaujoux R, Seoighe C. CellMix: a comprehensive toolbox for gene expression deconvolution. *Bioinformatics*. 2013; 29(17):2211–2212. doi: [10.1093/bioinformatics/btt351](https://doi.org/10.1093/bioinformatics/btt351) PMID: [23825367](https://pubmed.ncbi.nlm.nih.gov/23825367/)
100. Chikina M, Zaslavsky E, Sealfon SC. CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. *Bioinformatics*. 2015;p. btv015. doi: [10.1093/bioinformatics/btv015](https://doi.org/10.1093/bioinformatics/btv015)

101. Houseman EA, Kile ML, Christiani DC, Ince TA, Kelsey KT, Marsit CJ. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *bioRxiv*. 2016;p. 037671.
102. Hastie T, Tibshirani R, Eisen MB, Alizadeh A, Levy R, Staudt L, et al. Gene shaving as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol*. 2000; 1(2):1–0003. doi: [10.1186/gb-2000-1-2-research0003](https://doi.org/10.1186/gb-2000-1-2-research0003)