



Article

# Machine Learning Methods for Gene Selection in Uveal Melanoma

Francesco Reggiani <sup>1</sup>, Zeinab El Rashed <sup>1</sup>, Mariangela Petito <sup>1,2</sup>, Max Pfeffer <sup>3</sup>, Anna Morabito <sup>1</sup>,  
Enrica Teresa Tanda <sup>4,5</sup>, Francesco Spagnolo <sup>4,6</sup>, Michela Croce <sup>7</sup>, Ulrich Pfeffer <sup>1,\*</sup> and Adriana Amaro <sup>1</sup>

<sup>1</sup> Laboratory of Gene Expression Regulation, IRCCS Ospedale Policlinico San Martino, 16132 Genova, Italy; francesco.reggiani@hsanmartino.it (F.R.); mariangela.petito@hsanmartino.it (M.P.); anna.morabito@hsanmartino.it (A.M.)

<sup>2</sup> Department of Experimental Medicine (DIMES), University of Genova, Via Leon Battista Alberti, 16132 Genova, Italy

<sup>3</sup> Institute of Numerical and Applied Mathematics, University of Göttingen, 37083 Göttingen, Germany; m.pfeffer@math.uni-goettingen.de

<sup>4</sup> Skin Cancer Unit, IRCCS Ospedale Policlinico San Martino, 16132 Genova, Italy; enricateresa.tanda@hsanmartino.it (E.T.T.); francesco.spagnolo@hsanmartino.it (F.S.)

<sup>5</sup> Department of Internal Medicine and Medical Specialties, University of Genova, Viale Benedetto XV, 16132 Genova, Italy

<sup>6</sup> Department of Surgical Sciences and Integrated Diagnostics (DISC), University of Genova, 16132 Genova, Italy

<sup>7</sup> Biotherapies, IRCCS Ospedale Policlinico San Martino, 16132 Genova, Italy; michela.croce@hsanmartino.it

\* Correspondence: ulrich.pfeffer@hsanmartino.it



**Citation:** Reggiani, F.; El Rashed, Z.; Petito, M.; Pfeffer, M.; Morabito, A.; Tanda, E.T.; Spagnolo, F.; Croce, M.; Pfeffer, U.; Amaro, A. Machine Learning Methods for Gene Selection in Uveal Melanoma. *Int. J. Mol. Sci.* **2024**, *25*, 1796. <https://doi.org/10.3390/ijms25031796>

Academic Editor: Robert M. Verdjik

Received: 27 December 2023

Revised: 25 January 2024

Accepted: 30 January 2024

Published: 1 February 2024

**Correction Statement:** This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Uveal melanoma (UM) is the most common primary intraocular malignancy with a limited five-year survival for metastatic patients. Limited therapeutic treatments are currently available for metastatic disease, even if the genomics of this tumor has been deeply studied using next-generation sequencing (NGS) and functional experiments. The profound knowledge of the molecular features that characterize this tumor has not led to the development of efficacious therapies, and the survival of metastatic patients has not changed for decades. Several bioinformatics methods have been applied to mine NGS tumor data in order to unveil tumor biology and detect possible molecular targets for new therapies. Each application can be single domain based while others are more focused on data integration from multiple genomics domains (as gene expression and methylation data). Examples of single domain approaches include differentially expressed gene (DEG) analysis on gene expression data with statistical methods such as SAM (significance analysis of microarray) or gene prioritization with complex algorithms such as deep learning. Data fusion or integration methods merge multiple domains of information to define new clusters of patients or to detect relevant genes, according to multiple NGS data. In this work, we compare different strategies to detect relevant genes for metastatic disease prediction in the TCGA uveal melanoma (UVM) dataset. Detected targets are validated with multi-gene score analysis on a larger UM microarray dataset.

**Keywords:** uveal melanoma; multi-domain data; data fusion

## 1. Introduction

In the last two decades, next-generation sequencing data have unveiled the genomics behind cancer and genetic diseases at a high level of detail. A consistent amount of this data has been made available in public databases related to projects such as The Cancer Genome Atlas (TCGA), the Personal Genome Project or repositories such as the European Genome-phenome Archive [1–3].

Public repositories of NGS data store processed or raw datasets, while in the first case data are directly available for bioinformatic analysis, as an expression matrix with samples in columns and genes in rows. In the other case, data should be preprocessed and prepared

for further analysis. Several ad hoc public pipelines to obtain human readable files (as expression matrices or variant call format) from raw data files as FASTQ files are now available. They can be used by the researcher to detect genetic variants or genes related to disease severity or progression [4,5].

Genomic data such as methylation, gene expression or copy number alteration (CNA) matrices, from the same samples, can be analyzed individually or by integrating multiple domains at the same time. Examples of the first kind of approach include the application of deep learning methods to extract informative genes from gene expression profiles [6]. Among the approaches that process multiple domains at the same time, examples include the analysis of the association between gene expression and the presence of CNA events [7], or the integration of multiple domains by data fusion to cluster patients into groups with different survival [8]. Data fusion (DF) methods were developed to merge different data domains in an unsupervised way for feature selection and sample clustering [8–10]. DF applications such as joined singular value decomposition (jSVD) integrate the information from multiple domains to produce a single matrix; therefore, each sample is projected in a  $k$  dimensional space that can be used to define new clusters with methods such as  $k$ -means [8,9,11]. Other methods such as similarity network fusion (SNF) directly perform sample clustering based on the integration on one network of single-sample correlation matrices computed on each domain [12].

Uveal melanoma, a rare cancer of the eye that affects two to eight people a year per million people, has been molecularly characterized in great detail. Genomic analyses have shown that it is driven by a very limited number of driver events, and the analysis of gene expression, chromosome copy number alterations and DNA methylation concordantly reveals the existence of four major risk related subtypes that are clearly distinct from each other and tightly linked to the development of metastases and disease-free and overall survival after diagnosis [13]. UM has a very low mutational burden of 17–30 somatic mutations that affect protein coding sequences and might have functional consequences. Apparently, a single initiator mutation in one of four genes (GNAQ, GNA11, PCLB2 and CYSLTR2) is sufficient to yield a tumor that, upon acquisition of an additional mutation in the genes BAP1, a tumor suppressor gene, or SF3B1, a splice factor gene, and cytogenetic alterations, will progress to metastasis [13–15]. The four molecular classes are characterized by disomy of chromosome 3 without (class A) or with (class B) a hotspot SF3B1 mutation or monosomy of chromosome 3 and BAP1 mutation without (class C) or with (class D) amplification of chromosome 8q and an inflammatory infiltrate. Metastatic risk is low in class A, intermediate in class B and high in classes C and D [16]. These molecular classes are reflected by cytogenetic alterations and differential gene expression as well as differential DNA methylation [16,17].

Known molecular drivers and clearly distinct risk classes make UM especially suited for the development of data fusion approaches since it is straightforward to test classifications as to whether they improve classification over known classification systems based on single domains. However, it should be considered that genomics data are always characterized by some degree of noise from biological or technical factors (e.g., sample preparation, quality, etc.) and size limitations that prohibit perfect classification, which could instead be observed in an artificial training set [18].

Previously, we applied and adapted data fusion approaches to prognostic classification of UM. We performed joined singular value decomposition (jSVD) known in chemometrics as simultaneous component analysis, a simultaneous principal component analysis (PCA), and we developed joined constrained matrix factorization (jCMF) based on a form of coupled matrix factorization, also known as the  $k$ -table method, with a generalization of this factorization by allowing different constraints on the factor matrices [8,9].

Here, we report on the analysis of the uveal melanoma dataset with algorithms based on open-source code, i.e., R and Python implementations.

Information from multiple domains such as expression, methylation and copy number alteration from TCGA patients affected by UM were merged using data fusion or integration

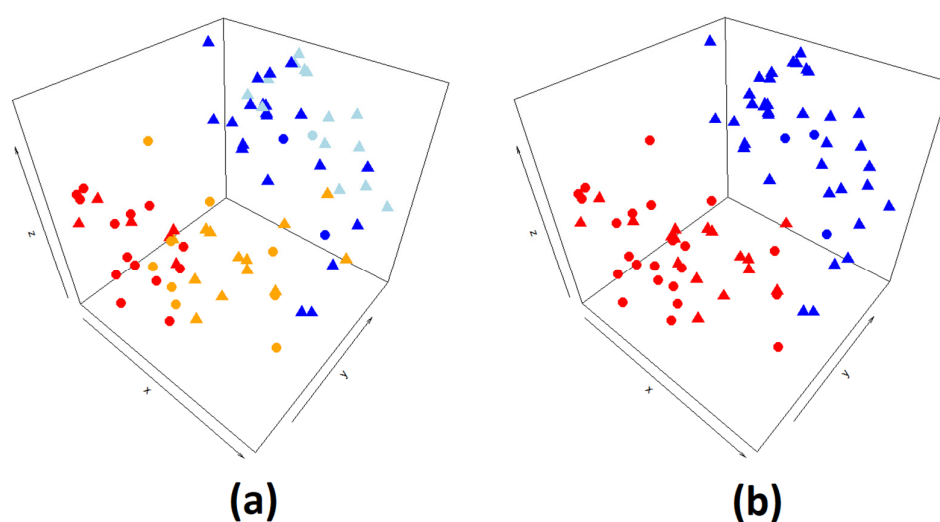
methods and applied to distribute samples in different clusters to perform feature selection, as we previously applied to the skin cutaneous melanoma TCGA dataset [9]. Different methods of data integration based on genomic domains are compared to evaluate which features (genes) are most relevant for UM subtypes and risk class detection and in which domains their effect is detectable (CNA level, gene expression or methylation). Selected methods will analyze genomic data that are relevant for UM subtyping in the high or low risk classes: CNA-based methods, which prioritize genes with expression levels altered by variation in copy number, data fusion or integration approaches will integrate expression and methylation data for patient clustering and feature selection as well as to identify which ones are transcriptionally predictive (i.e., genes with an association between methylation and expression levels).

## 2. Results

### 2.1. Gene Prioritization Methods

#### 2.1.1. Data Fusion

jSVD was used to integrate RNA-seq and methylation TCGA UVM data in order to produce a matrix  $U$ , with each patient defined by a three-dimensional space (Figure 1a,b). High- and low-risk classes are clearly separated in space (Figure 1a); the two clusters defined by k-means on the  $U$  matrix classified two patients as high risk who did not develop metastasis during follow-up in the blue cluster, which contains all low-risk samples (Figure 1b). The number of k-means clusters (2) was defined by balancing low-connectivity values while maximizing the silhouette score (Table 1).

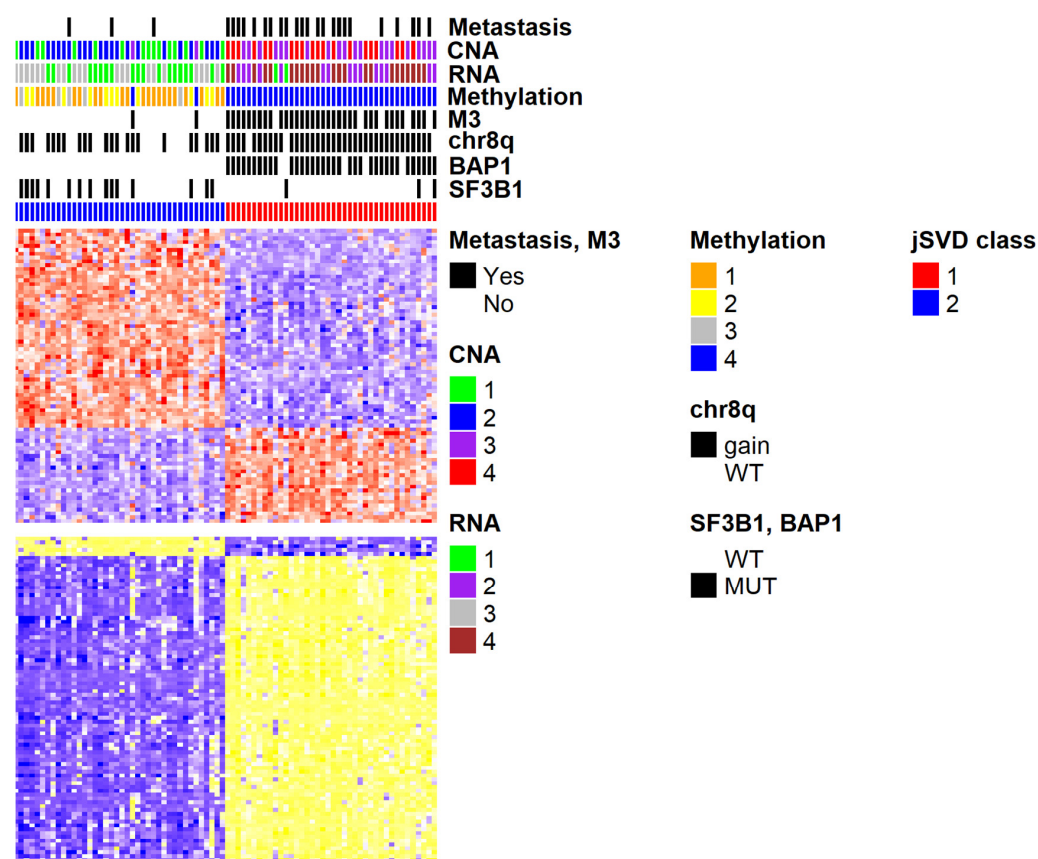


**Figure 1.** (a) Scatterplot of the 80 patients from the UVM TCGA dataset in the three-dimensional ( $k = 3$ )  $U$  matrix produced using jSVD data integration of RNA-seq and methylation array data. Points are colored according to the metastatic risk classes [16], from high (4) to low (1): 4 in red, 3 in orange, 2 in blue, 1 in azure. Patients that developed metastasis are reported as circles. (b) Scatterplot of the 80 patients from the UVM TCGA dataset in the three-dimensional ( $k = 3$ )  $U$  matrix produced using jSVD data integration of RNA-seq and methylation array data. Points are colored according to the two clusters defined by k-means on the jSVD  $U$  matrix.

**Table 1.** Performance measures used to define the optimal number of clusters ( $k$ ). Connectivity is 0 if one sample has no neighbors from different clusters, while silhouette score represents sample fit in its cluster.

| Number of $k$      | 2    | 3    | 4     |
|--------------------|------|------|-------|
| Connectivity score | 2.25 | 6.27 | 15.66 |
| Silhouette score   | 0.45 | 0.51 | 0.52  |

At this point, we applied bootstrap analyses (significance analysis of microarrays, SAM [19]) to detect differentially expressed genes among the two classes of UM samples as defined by k-means. Samples of the two clusters are characterized by a set of differentially expressed genes and methylated probes (Figure 2). The two high-risk samples (class 3) that were clustered among the lower-risk cases by k-means show a methylation and expression profile that is more similar to their neighbors than the ones of the other group (in red, Figure 2). Generally, the classification as “high risk” of patients who did not develop metastases can be misclassification but must not be so, since they might develop metastases later on or have responded to therapy. The latter does not apply to UM for the absence of adjuvant therapy.



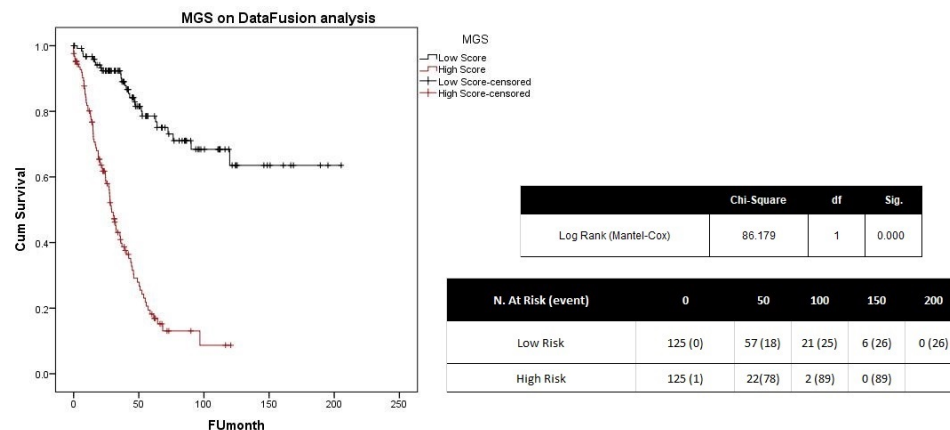
**Figure 2.** Heatmap of the differentially expressed genes and methylated probes of the 80 patients from the UVM TCGA dataset, considering the two clusters detected on the jSVD U matrix. At the top of the figure, each row represents different sample features, respectively: presence of metastasis, CNA metastatic risk, RNA and methylation cluster class [16], loss of chromosome 3 (M3), gain of chromosome 8q, mutations on BAP1 and SF3B1.

We tested the differentially expressed genes on a dataset of 253 UVM patients [17], using a multi-gene score (MGS); this produced two groups with a significant difference in terms of survival (Figure 3). Differentially methylated probes were tested only on the TCGA dataset; only one probe passed the multivariate testing (cg05522415, Figure S1).

### 2.1.2. CNA Analysis Methods

The IGC R package (v 1.22) [7] was used to detect genes with expression values associated with CNA gain or loss. Considering a false discovery rate (FDR) below 0.05, 2036 genes were detected: 502 associated with CNA gain, the remaining 1534 with loss events. The CNAPE R package (<https://github.com/WangLabHKUST/CNAPE>, accessed on 24 January 2024) detects relevant features for CNA detection from RNA-seq data [20];

we used this package to develop a model able to distinguish between monosomic and disomic samples in chromosome 3, using TCGA UM data. A total of 299 genes were used to make the prediction and were considered for further analysis. These genes can distinguish disomic (low risk) from monosomic (high risk) patients.



**Figure 3.** KM curves of patients with a high or low MGS considering the differentially expressed genes detected using SAM analysis on the two jSVD-related clusters. Low score curve is in black, while the high one is reported in red.

### 2.1.3. Methylation Analysis Methods

The MethylMix R package (v 2.32) [21] was used to detect genes with methylation levels associated with expression. This method uses a control group of samples to remove genes that are not differentially methylated compared to cancer samples and detect which ones are transcriptionally predictive (e.g., genes for which there is a significant inverse relationship of expression and methylation) [21]. Patients classified in class 2 by data fusion were considered as controls and those in class 1 were treated as tumor samples. Methylmix detected 90 genes as transcriptionally predictive.

### 2.2. Integration of Results

Several genes were detected using each method (Figure 4). In particular, data fusion selected 28 features that were also detected with CNAPE and IGC. CNAPE and IGC shared more genes compared to data fusion; this was expected since both methods are expected to detect genes with expression levels associated to CNAs, while data fusion analysis is based on RNAseq and methylation data. Among the seventeen DF selected genes that have passed survival analysis, two were detected using all methods (ROBO1, ROPN1, Table 2), while nine were shared with at least one R package based on CNA data analysis (IGC or CNAPE), and one by MethylMix.

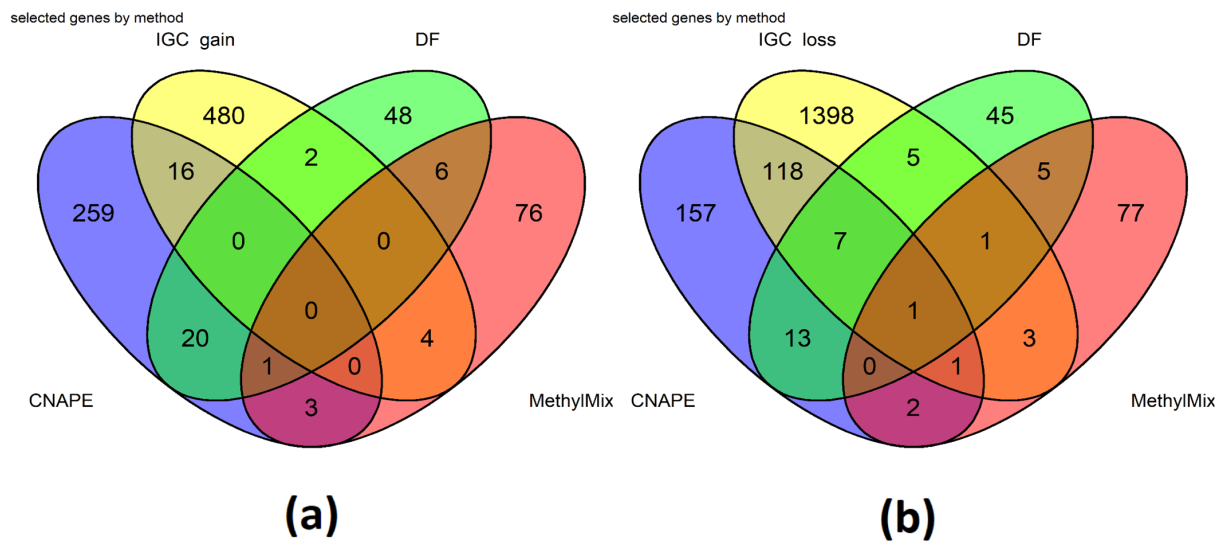
ROBO1, ROPN1, BCHE, and CHL1 present lower gene expression in patients with a CNA loss at their locus (Figure 5): while ROBO1 and CHL1 gene expression reduced the score of the MGS signature, the opposite effect is produced by ROPN1 and BCHE (overexpressed in some samples with bad prognosis, as shown in Figures S1 and S2). ROBO1 and CHL1 map on chromosome 3p; the other two are located on 3q. CHL1 was found to be one of the most downregulated genes in UM that metastasized to the liver compared to non-metastatic tumors [22]. ROPN1 has previously been described as related to good prognosis when overexpressed in the UM TCGA dataset [23]; however, in the Piaggio et al. dataset [17], several metastatic patients have high expression levels of this gene (Figures S2 and S3). Among the genes selected by CNAPE and DF, we can find several genes related to worse prognosis. CADM1 and other genes involved in the production of cell adhesion molecules were found to be overexpressed in UM cells with BAP1 inactivation: experimental evidence supports a role of this gene in the metastasization process [24]. ITPR2 was previously described as mutated in the TCGA dataset; it is involved



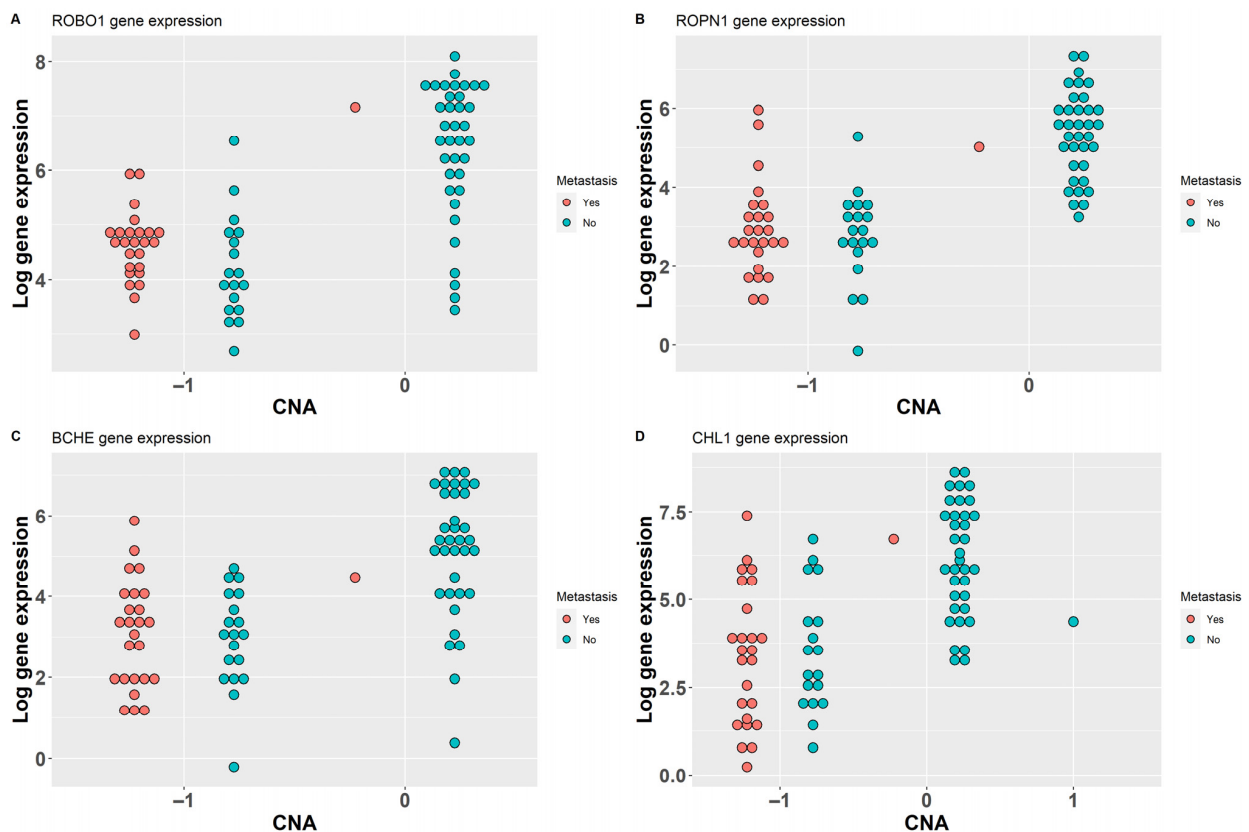
in G-protein-related pathways [15] and has been selected as part of a signature for tumor immune infiltration [25]. ISM1 was selected as a negative prognosis factor in a previously published 21-gene signature related to the UM tumor microenvironment, while MTUS1 and IL12RB2 were considered as indicators of favorable prognosis [26]. PDE4B was previously found as a protective factor in a prognostic signature based on inflammatory-related genes in UM [27]. ACSF2 was found to be among ferroptosis regulators in a signature, used to distinguish UM patients with different overall survival, that defined two clusters of patients with differences in prognosis and tumor-microenvironment-infiltrating cells [28,29]. CTF1 has been part of a previously defined UM-immune-related three-gene signature on TCGA data [30]. CARD11 was detected as a prognostic marker, with high expression associated to poor OS in the TCGA UVM dataset; in particular, metastatic patients had higher expression of this gene [31]; however, the MGS based on a larger dataset [17] assigned a protective effect to this gene, probably due to a set of patients with limited survival, metastatic disease and low CARD11 expression (Figures S2 and S3). HTR2B, TNFRSF19 and PTGER4 were previously found to be overexpressed in class 2 tumors (metastatic) [32]; in particular, TCGA UVM patients with high PTGER4 expression had worse survival [33]. Gene set enrichment analysis of MGS elements (Table 2) shows that these genes are involved in inflammatory (CARD11, PDE4B, TNFRSF19, HTR2B) and cell-motility-related biological processes (MTUS1, ROBO1, PTGER4, CHL1, HTR2B, PDE4B, ROPN1, Figure 6, Table S2).

**Table 2.** Data fusion genes selected using multi-gene score (MGS) procedure. Genes are ordered considering the number of different methods that detected the gene (n overlap column). Column 1 to 3 report presence (1) or absence (0) of the gene using each method: CNAPE and IGC are CNA loss and data fusion, respectively. The cytoband and the multi-gene score (MGS score) of each gene are reported in the last columns. Of all genes in Table 2, only the CTF1 gene was also detected using MethylMix.

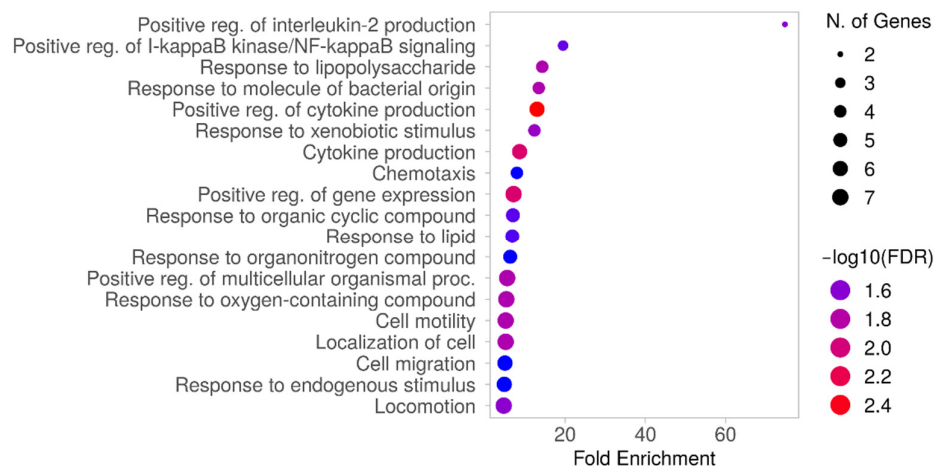
| GENE     | CNAPE | IGC Loss | Data Fusion | n Overlap | Cytoband | MGS Score |
|----------|-------|----------|-------------|-----------|----------|-----------|
| ROBO1    | 1     | 1        | 1           | 3         | 3p12.3   | −0.241    |
| ROPN1    | 1     | 1        | 1           | 3         | 3q21.1   | 0.312     |
| CADM1    | 1     | 0        | 1           | 2         | 11q23.3  | 0.233     |
| ITPR2    | 1     | 0        | 1           | 2         | 12p12.1  | −0.323    |
| ISM1     | 1     | 0        | 1           | 2         | 20p12.1  | 0.213     |
| PDE4B    | 1     | 0        | 1           | 2         | 1p31.3   | −0.291    |
| ACSF2    | 1     | 0        | 1           | 2         | 17q21.33 | 0.302     |
| BCHE     | 0     | 1        | 1           | 2         | 3q26.1   | 0.274     |
| CHL1     | 0     | 1        | 1           | 2         | 3p26.3   | −0.152    |
| IL12RB2  | 0     | 0        | 1           | 1         | 1p31.3   | −0.225    |
| MTUS1    | 0     | 0        | 1           | 1         | 8p22     | −0.276    |
| CTF1     | 0     | 0        | 1           | 1         | 16p11.2  | −0.301    |
| CPS1     | 0     | 0        | 1           | 1         | 2q34     | 0.177     |
| HTR2B    | 0     | 0        | 1           | 1         | 2q37.1   | 0.21      |
| CARD11   | 0     | 0        | 1           | 1         | 7p22.2   | −0.254    |
| TNFRSF19 | 0     | 0        | 1           | 1         | 13q12.12 | 0.125     |
| PTGER4   | 0     | 0        | 1           | 1         | 5p13.1   | 0.12      |



**Figure 4.** (a,b) Genes detected using different data analysis and integration methods: most of the data fusion genes were not detected using any other methods (48, 43 also considering IGC gain and MethylMix), 13 are shared with CNAPE (b), and 8 were detected using CNAPE and IGC as low expression driven by CNA loss; 2 were also detected as CNA gain by IGC; 5 are only shared with IGC loss (b). Only 7 genes are shared between DF and Methylmix (a,b).



**Figure 5.** Expression level and CNA state of genes detected using data fusion and IGC (TCGA UVM dataset): low level of expression is associated with CNA loss. The Y axis reports the log expression values, while the x axis reports the CNA state with  $-1$  as loss,  $1$  as gain and  $0$  as neutral (e.g., no CNA). Metastatic samples have generally low expression values compared to copy neutral ( $0$ ) samples. Each panel represents the expression levels of one gene in the TCGA dataset: ROBO1 (A), ROPN1 (B), BCHE (C), CHL1 (D).



**Figure 6.** Gene set enrichment analysis on MGS features (Table 2, FDR cutoff of 0.05). Enriched GO BP terms are reported ordered on the basis of the number of genes of the signature in Table 2 and fold enrichment. Dot size is based on the number of genes (in Table 2) involved in the process; color is based on FDR.

Eventually, we considered whether there was any overlap between the 90 genes detected using MethylMix with a consistent correlation between expression and methylation levels in the high-risk data fusion class (1) and not in the other class (2, Figure 4). SLC25A38 was detected using all methods; it maps chromosome 3p and is downregulated in metastatic UM patients; inactivation of this gene has been shown to promote distant metastasis in mouse models [34]. Other genes such as PLXNB1 and HLA-A were part of an immune gene signature used to define two risk classes, one of which had higher immune cell infiltration and lower survival in the TCGA UVM dataset [35]. CTF1 and RAPGEF3 were previously reported to be parts of gene signatures related to tumor microenvironment and immune system [26,30], with the first seen downregulated and methylated in BAP1 mutated samples [36]; PALMD was found to have low expression in metastatic UM tissues [37], and GSTA3 in low-survival patients [38].

### 3. Discussion

Integration of multiple genomics and phenotype data is gradually unveiling the complex molecular biology behind genetic diseases and cancers [39–41]. Data fusion has been previously applied as a tool to cluster patients or how to extract relevant features for disease prognosis by integrating data of several NGS, imaging and other clinically related datasets from the same group of samples [42–44]. The main limitations to the application of these approaches are batch effects, the curse of dimensionality that arises with genomic data and missing information or heterogeneity (data incompleteness) [43]. Regarding the first point, in each sequencing experiment, technical differences among replicates could mask or mimic biological variation; for example, different sequencing coverage among two groups of samples sequenced with RNAseq could potentially lead to the discovery of several false positives, as differentially expressed genes, if samples are not properly normalized [45]. The curse of dimensionality resides in the fact that in an NGS experiment, the number of features greatly exceeds the number of samples [46], which can easily result in model overfitting [47] and the inability to extract any relevant biological features or perform meaningful classification using the same model in a different dataset. Data fusion or integration methods can work on a full dataset or on a limited subset of genes, i.e., the most variable features [8,10,12,48]. In this way, most of the non-informative features are removed, reducing the required computational resources and the noise inside the dataset. In this work, we have shown that data fusion can potentially improve patient classification, as two patients previously classified by single domain analysis as high risk, but that had not developed metastasis during follow up, were classified with low- and



intermediate-risk patients (Figure 2, on the left). However, it is not clear whether this classification could be efficient in a larger dataset since, to date, TCGA UVM is the only publicly available multidomain uveal melanoma dataset. However, promising results were obtained by applying DF on UM samples with expression data only and on TCGA samples for which mean gene methylation data were also available [49]. Interestingly, 9 out of the 17 DF detected genes that passed MGS were also detected using CNAPE or IGC; 2 of them were associated with a CNA loss (ROBO1, ROPN1, BCHE, CHL1). Interestingly, ROPN1 and BCHE, both mapping on chromosome 3q, have generally low expression levels in TCGA patients that developed metastasis during follow-up but not so in other UM datasets (Figures S2 and S3). One explanation could be that several patients from datasets other than the TCGA dataset could have a partial deletion on chromosome 3, not involving these two genes (Figure S3). Unfortunately, no CNA data are available for these patients. The use of multiple datasets to evaluate the method is essential to obtain an accurate estimate of the reliability of a classification method. Limited training set size, in the past, had determined the development of overfitted bioinformatic models that were not superior to a random predictor in the classification of new samples [50]. In this work, we could only test the performance of the genes selected by data fusion applications with a multi-gene score on a larger dataset. Some of these genes were also described in different works regarding UM [51], while two of them (CHL1 and IL12RB2) were also found to be hypermethylated with low expression in invasive malignant melanoma cells [52]; in particular, CHL1 is in an hypermethylated region on 3p in TCGA class 2 UMs [53].

Data fusion research should focus on new methods of data integration from multiple domains. Some genes could be affected by multiple genomic events that inactivate their expression (as from mutation, CNA and methylation domain). Single domain analysis failed to detect these genes as significantly altered in tumors, while the analysis of multiple domains could be a strong basis to distinguish between genes with a functional role in pathogenesis and those not causally involved markers.

#### 4. Materials and Methods

##### 4.1. *jSVD Data Preparation and Analysis*

TCGA methylation and RNA-seq data were downloaded from Broad GDAC Firehose (<https://gdac.broadinstitute.org/>, accessed on 31 January 2023). RSEM gene expression counts were filtered from outliers by removing genes with less than 100 or more than  $10^6$  counts over all samples. RNAseq data normalization was based on the blind vst normalization function, as implemented in the DESeq2 R package (v 1.32.0) [54]. Feature reduction was performed by selecting the 1500 genes with the highest MAD for RNA-seq and the 1% most variable methylation probes; these data were used as input for the *jSVD* python script, as previously applied by Amaro and coauthors [9], setting the number of columns produced by the U matrix to 3. Patient clustering on the U matrix, produced by the *jSVD*, was based on the k-means method (complete agglomeration, Euclidean distance) from the R package ConsensusClusterPlus [55]; the number of cluster k was selected by minimizing connectivity and maximizing silhouette score, as computed by the *clvalid* R package [56]. Differentially expressed genes and methylated features, among patient clusters, were extracted with the significance analysis of microarray as implemented in R (Samr) [19]. Resulting DEGs and differentially methylated probes were analyzed with SPSS Statistics 20; in particular, multivariate Cox regression and multi-gene score analysis was computed on Piaggio et al.'s dataset [17], and the same analysis was conducted on the methylation probes of the TCGA UVM dataset [16].

##### 4.2. *CNAPE and IGC*

RNA-seq and CNA data analyzed using CNAPE and the IGC R package [7,20] were downloaded from cBioPortal (<https://www.cbioportal.org/>, accessed on 24 January 2024) [57,58]; only genes with a CNA in at least 4 samples were considered for further analysis. These pieces of software work on expression and CNA matrices with the same

genes, as rows, and patients, as columns. IGC tests whether the expression of one gene is associated to CNA events overlapping the locus: detected relations could be “loss” if a decrease in RNA expression is associated with deletion events, “gain” if increased expression is associated with augmented copies of one gene, or “both” when the two events (gain and loss) are observed in the same gene [7]. In a first step, samples with CNA on one gene are classified as CNA-gain (“gain”, with an increase in CNA), CNA-loss (“loss”, with a decrease in CNA) or CNA-neutral (no CNA detected). At this point, a gene can be classified as gain or loss on the proportion of samples that have the CNA event (e.g., if more than 20% of samples have a CNA gain on that gene, it is classified as “gain”). As a final step, Student’s *t*-test with unequal variance is computed on the expression values. For each gene, a false discovery rate (FDR) and p value is reported; in this work, only “gain” and “loss” elements with an FDR below 0.05 were considered (as obtained with the `find_cna_driven_gene` function with standard parameters: `gain`, `loss_prop` = 0.2). CNAPE uses RNA-seq data to develop a model able to distinguish between samples with or without a large CNA event [20]. In this work, the model was trained on NGS data in order to distinguish between chromosome 3 gain or loss; the genes selected by the model to make a prediction were considered for further analysis and reported in Supplementary Table S1 and Figure 4. The model was trained with the `cv.glmnet` function with default parameters, except the number of cross-validation folds, which was set to 20 to have stable results (`md = cv.glmnet(x = as.matrix(dtx), y = dty, family = “binomial”, nfolds = 20, alpha = 0.1)`)

#### 4.3. MethylMix

RNA-seq UVM data were downloaded from cBioPortal [57,58], and mean gene methylation levels were obtained from <https://gdac.broadinstitute.org/> (accessed on 31 January 2023). The table of mean gene methylation was split in two, the first composed of samples classified in class 1 by data fusion and considered as cancer samples (*METcancer*), the second comprising class 2 patients, treated as control samples (*METnormal*). RNA-seq data of class 1 patients were retrieved from cBioportal normalized expression data and treated as a cancer gene expression profile (*MAcancer*). The Methylmix R package [21] was used to detect transcriptionally predictive genes with the *MethylMix* function *MethylMix(METcancer, GEcancer, METnormal)*. Briefly, genes with different methylation levels in cancer and control data were tested to assess whether they had a significant relationship with expression data.

#### 4.4. Joint Singular Value Decomposition

Joint singular value decomposition, described in [8,59], was developed with the Python package Pymanopt (v 0.2.5) [60]. jSVD factorizes each genomic data matrix *A* as (1):

$$A \approx U \Sigma_i V_i^T \quad (1)$$

$\Sigma_i$  is a singular value diagonal matrix; the others are orthonormal. The *U* matrix is shared among each matrix decomposition; therefore, it represents the fused information from *A* datasets and is used for patient clustering. A Riemannian Trust scheme has been used to obtain a minimum on the product of Stiefel manifolds (set as `Product([Stiefel(I,k), Stiefel(N1,k), Stiefel(N2,k)]`: *N1* and *N2* represent the number of genes or methylation probes of the RNA-seq, Methylation matrix, respectively). The minimization was stopped when the norm of the projected gradient was lower than  $1^{-12}$  (`mingradnorm` =  $1 \times 10^{-12}$ ).

#### 4.5. Gene Signature Performance Evaluation

Features selected by all methods presented in this paper (CNAPE, IGC, MethylMix and Data Fusion) were assessed as gene signatures to predict chromosome 3 monosomy and metastatic disease development compared to chromosome 3 monosomy on the Piaggio et al. 2022 dataset [17] in terms of AUC, as previously applied for signature and phenotype prediction validation [34,61–63]. Gene signature scores were computed with the *simpleScore()* function of the *signscore* R package [64,65]. The computed TotalScore of each signature and overall score computed on available genes reported in Table S1 were

converted to a value between 0 and 1 by subtracting to each value the minimum value of the signature and dividing it by the difference of the maximum and minimum value of the signature. The ROCR package (v 1.0-11) [66] was used to compute the ROC curves and relative AUC of each signature, by comparing the difference between 1 and the signature score (except for IGC gain) and the M3 or metastasis classes.

## 5. Conclusions

In this work, different data integration and single domain gene prioritization methods were applied to the UM TCGA dataset. Most of the genes detected using IGC and CNAPE are located on chromosomal positions where relevant CNAs used for clinical assessment of UM metastatic risk are present (1p, 16q, 3p loss and 6p, 8q gain, Table 3) [67].

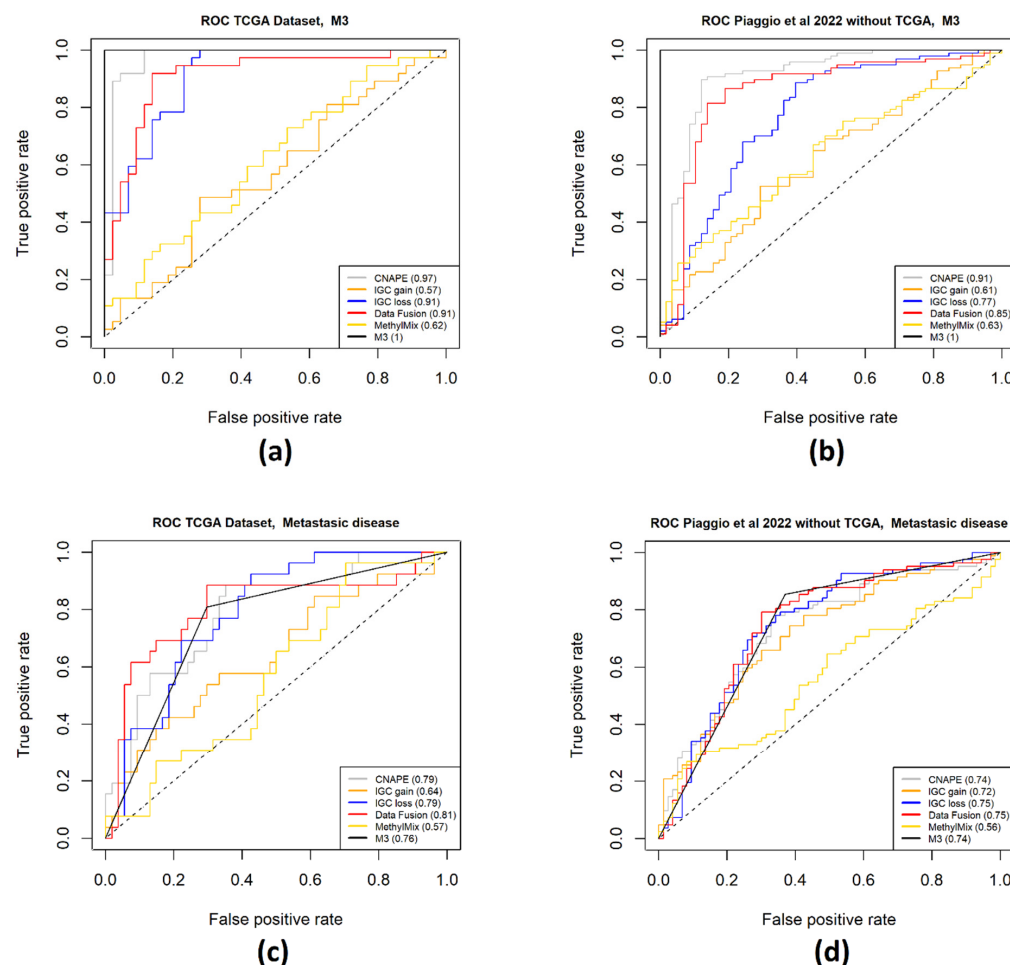
**Table 3.** Chromosome localization of genes reported in Table S1. For each method, the number of genes mapping on chromosomes relevant for cytogenetic characterization of uveal melanoma are reported.

| Chr   | CNAPE | IGC Gain | IGC Loss | Data Fusion | MethylMix |
|-------|-------|----------|----------|-------------|-----------|
| 1p    | 9     | 0        | 606      | 5           | 6         |
| 1q    | 8     | 0        | 0        | 1           | 4         |
| 3p    | 71    | 0        | 301      | 7           | 4         |
| 3q    | 53    | 0        | 285      | 5           | 0         |
| 6p    | 11    | 265      | 0        | 2           | 5         |
| 6q    | 3     | 0        | 244      | 1           | 2         |
| 8p    | 5     | 1        | 0        | 2           | 1         |
| 8q    | 11    | 236      | 0        | 3           | 3         |
| 16p   | 3     | 0        | 0        | 2           | 2         |
| 16q   | 2     | 0        | 98       | 1           | 0         |
| other | 123   | 0        | 0        | 48          | 63        |
| total | 299   | 502      | 1534     | 77          | 90        |

Chromosome 16q and 1p deletion were found to increase metastatic risk in patients with M3 and chromosome 8 amplification [67]. IGC prioritizes CNA associated genes on the basis of related RNA expression. Therefore, genes that are not strictly regulated by deletion or gain events will not be detected. CNAPE selected a set of genes able to discriminate between chromosome 3 monosomic and disomic patients of the TCGA UVM dataset: the genes detected were not only localized on 3p or 3q, since features in other genomics locations were used for M3 prediction. Data-fusion- and MethylMix-prioritized feature localization was more dispersed on the whole genome compared to genes detected using IGC: the integration of RNAseq and methylation array data also prioritizes genes that are not strictly regulated by CNAs. Therefore, the integration of results from different gene selection methods can detect features that are relevant for UM prognosis but are not detectable in a single genomic domain. The prediction performance of the signature detected using all methods presented in this work (Table S1) is reported in Figure 7. In general, gene expression signatures predictive of chromosome 3 monosomy obtained higher AUC values compared to metastatic disease onset estimation. This reflects the fact that chr3 monosomy is certain at the time of analysis whereas metastases can also develop after the end of follow-up. High-risk cases that did not develop metastases during follow-up might do so afterwards.

Interestingly, CNAPE outperformed all methods on M3 prediction, obtaining high AUC in the TCGA and the remaining part of the Piaggio dataset [17] (Figure 7a,b). If we compare the performance of IGC loss on M3 prediction, we can observe a consistent de-

crease in the AUC score by comparing the dataset where the gene signature was computed and a different test set (AUC from 0.91 to 0.77). It should be taken into account that IGC simply detects genes with expression level associated to CNA of the gene, without performing any supervised feature selection for M3 prediction. Therefore, worse performance in different datasets could be expected. Regarding the prediction of metastatic disease from gene signatures, a general decrease in the performance is observable comparing the dataset where features were extracted with the validation datasets (Figure 7c,d). Considering only the TCGA dataset, some methods show a performance superior to chromosome 3 classification for metastatic risk prediction, while all curves are near M3 classification in all other samples (Figure 7d).



**Figure 7.** ROC curves of gene-signature-based prediction of M3 and metastatic disease development during follow up. Each ROC curve was computed on the TCGA UVM (a,c) and the Piaggio dataset [17] (b,d). The four panels report the performance of gene signatures on M3 (a,b) and metastasis prediction (c,d). AUC scores are reported in each panel legend; chromosome 3 monosomy ROC curve is reported as a black line and can be used as a reference for comparison.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25031796/s1>.

**Author Contributions:** Conceptualization, F.R., U.P. and A.A.; methodology, F.R., U.P. and A.A.; software, F.R.; validation, F.R., U.P. and A.A.; formal analysis, F.R., U.P., A.A., M.P. (Mariangela Petito), M.C., E.T.T., F.S., Z.E.R., M.P. (Max Pfeffer) and A.M.; investigation, F.R., U.P., A.A., M.C., E.T.T., F.S., Z.E.R., M.P. (Mariangela Petito), M.P. (Max Pfeffer) and A.M.; resources, U.P. and A.A.; data curation, F.R. and A.A.; writing—original draft preparation, F.R., U.P., A.A., M.P. (Mariangela Petito), M.C., E.T.T., F.S., Z.E.R., M.P. (Max Pfeffer) and A.M.; writing—review and editing, F.R., U.P., A.A.,

M.P. (Mariangela Petito), M.C., E.T.T., F.S., Z.E.R., M.P. (Max Pfeffer) and A.M.; visualization, F.R.; supervision, U.P. and A.A.; project administration, U.P.; funding acquisition, U.P. and A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Italian Ministry of Health, Ricerca Corrente 2022 to U.P. and the Italian Ministry of Health  $5 \times 1000$  2018/19 to A.A. Additionally, M.P. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 448293816.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data used in this work for data integration can be downloaded from <http://cancergenome.nih.gov/> and cBioportal, as described in the Material and Methods section. The UVM gene expression test set, described in [17], is composed of public datasets (GSE27831, GSE51880, TCGA-UVM) and the Leiden dataset, which is available upon request to authors [68].

**Acknowledgments:** The results shown here are in part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/> (URL accessed on 11 December 2023). The minfi and data.table R packages were used for methylation data analysis [69,70]. Plot3D R package, Complex Heatmap and Circlize [71–73] R packages for Figures 1 and 2, respectively. IlluminaHumanMethylation450kanno.ilmn12.Hg19 and Org.Hs.Eg.Db R annotation packages were used for methylation array analysis [74,75]. Cowplot R package was used to aggregate panels in Figures 1 and 4 [76]. The Venn diagram shown in Figure 4 was produced with the ggVenn R package [77]. Enrichment analysis was performed with ShinyGO [78]. Further details on how to develop a prediction model with CNAPE are available on the relative github page: [https://github.com/WangLabHKUST/CNAPE/blob/master/example/Example\\_copy\\_number\\_alteration\\_in\\_glioma.md](https://github.com/WangLabHKUST/CNAPE/blob/master/example/Example_copy_number_alteration_in_glioma.md) (accessed on 24 January 2024). A tutorial on how to use the IGC package is available here: <https://www.bioconductor.org/packages/devel/bioc/vignettes/iGC/inst/doc/Introduction.html> (accessed on 24 January 2024). jSVD-related scripts are available here: [https://github.com/FranzReg91/Amaro\\_et\\_al\\_2022\\_SKCM](https://github.com/FranzReg91/Amaro_et_al_2022_SKCM) (accessed on 24 January 2024).

**Conflicts of Interest:** Enrica Teresa Tanda has received honoraria for Bristol Myers Squibb; MSD; and Pierre Fabre. Francesco Spagnolo had lecture fees from BMS, MSD, Pierre Fabre, Novartis, Sun Pharma, Sanofi, Merck, and advisory boards of MSD, Novartis, Pierre Fabre, Philogen. The other authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. The Cancer Genome Atlas Research Network; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef] [PubMed]
2. Freeberg, M.A.; Fromont, L.A.; D’Altri, T.; Romero, A.F.; Ciges, J.I.; Jene, A.; Kerry, G.; Moldes, M.; Ariosa, R.; Bahena, S.; et al. The European Genome-Phenome Archive in 2021. *Nucleic Acids Res.* **2022**, *50*, D980–D987. [CrossRef] [PubMed]
3. Church, G.M. The Personal Genome Project. *Mol. Syst. Biol.* **2005**, *1*, 2005.0030. [CrossRef] [PubMed]
4. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow Enables Reproducible Computational Workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [CrossRef] [PubMed]
5. Dotolo, S.; Esposito Abate, R.; Roma, C.; Guido, D.; Preziosi, A.; Tropea, B.; Palluzzi, F.; Giacob, L.; Normanno, N. Bioinformatics: From NGS Data to Biological Complexity in Variant Detection and Oncological Clinical Practice. *Biomedicines* **2022**, *10*, 2074. [CrossRef] [PubMed]
6. Morabito, F.; Adornetto, C.; Monti, P.; Amaro, A.; Reggiani, F.; Colombo, M.; Rodriguez-Aldana, Y.; Tripepi, G.; D’Arrigo, G.; Vener, C.; et al. Genes Selection Using Deep Learning and Explainable Artificial Intelligence for Chronic Lymphocytic Leukemia Predicting the Need and Time to Therapy. *Front. Oncol.* **2023**, *13*, 1198992. [CrossRef] [PubMed]
7. Lai, Y.-P.; Wang, L.-B.; Wang, W.-A.; Lai, L.-C.; Tsai, M.-H.; Lu, T.-P.; Chuang, E.Y. iGC-an Integrated Analysis Package of Gene Expression and Copy Number Alteration. *BMC Bioinform.* **2017**, *18*, 35. [CrossRef]
8. Pfeffer, M.; Uschmajew, A.; Amaro, A.; Pfeffer, U. Data Fusion Techniques for the Integration of Multi-Domain Genomic Data from Uveal Melanoma. *Cancers* **2019**, *11*, 1434. [CrossRef]
9. Amaro, A.; Pfeffer, M.; Pfeffer, U.; Reggiani, F. Evaluation and Comparison of Multi-Omics Data Integration Methods for Subtyping of Cutaneous Melanoma. *Biomedicines* **2022**, *10*, 3240. [CrossRef]



10. Duan, R.; Gao, L.; Gao, Y.; Hu, Y.; Xu, H.; Huang, M.; Song, K.; Wang, H.; Dong, Y.; Jiang, C.; et al. Evaluation and Comparison of Multi-Omics Data Integration Methods for Cancer Subtyping. *PLoS Comput. Biol.* **2021**, *17*, e1009224. [\[CrossRef\]](#)
11. Leng, D.; Zheng, L.; Wen, Y.; Zhang, Y.; Wu, L.; Wang, J.; Wang, M.; Zhang, Z.; He, S.; Bo, X. A Benchmark Study of Deep Learning-Based Multi-Omics Data Fusion Methods for Cancer. *Genome Biol.* **2022**, *23*, 171. [\[CrossRef\]](#)
12. Wang, B.; Mezlini, A.M.; Demir, F.; Fiume, M.; Tu, Z.; Brudno, M.; Haibe-Kains, B.; Goldenberg, A. Similarity Network Fusion for Aggregating Data Types on a Genomic Scale. *Nat. Methods* **2014**, *11*, 333–337. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Rossi, E.; Croce, M.; Reggiani, F.; Schinzari, G.; Ambrosio, M.; Gangemi, R.; Tortora, G.; Pfeffer, U.; Amaro, A. Uveal Melanoma Metastasis. *Cancers* **2021**, *13*, 5684. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Amaro, A.; Gangemi, R.; Piaggio, F.; Angelini, G.; Barisione, G.; Ferrini, S.; Pfeffer, U. The Biology of Uveal Melanoma. *Cancer Metastasis Rev.* **2017**, *36*, 109–140. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Piaggio, F.; Tozzo, V.; Bernardi, C.; Croce, M.; Puzone, R.; Viaggi, S.; Patrone, S.; Barla, A.; Coviello, D.; Jager, M.J.; et al. Secondary Somatic Mutations in G-Protein-Related Pathways and Mutation Signatures in Uveal Melanoma. *Cancers* **2019**, *11*, 1688. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Robertson, A.G.; Shih, J.; Yau, C.; Gibb, E.A.; Oba, J.; Mungall, K.L.; Hess, J.M.; Uzunangelov, V.; Walter, V.; Danilova, L.; et al. Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. *Cancer Cell* **2017**, *32*, 204–220. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Piaggio, F.; Croce, M.; Reggiani, F.; Monti, P.; Bernardi, C.; Ambrosio, M.; Banelli, B.; Dogrusöz, M.; Jockers, R.; Bordo, D.; et al. In Uveal Melanoma Gα-Protein GNA11 Mutations Convey a Shorter Disease-Specific Survival and Are More Strongly Associated with Loss of BAP1 and Chromosomal Alterations than Gα-Protein GNAQ Mutations. *Eur. J. Cancer* **2022**, *170*, 27–41. [\[CrossRef\]](#)
18. Pfeffer, U.; Romeo, F.; Noonan, D.M.; Albini, A. Prediction of Breast Cancer Metastasis by Genomic Profiling: Where Do We Stand? *Clin. Exp. Metastasis* **2009**, *26*, 547–558. [\[CrossRef\]](#)
19. Tibishirani, R.; Michael, J.; Seo, G.C.; Balasubramanian, N.; Jun, L. SAM: Significance Analysis of Microarrays R Package Version 2018, 3.0.
20. Mu, Q.; Wang, J. CNAPE: A Machine Learning Method for Copy Number Alteration Prediction from Gene Expression. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 306–311. [\[CrossRef\]](#)
21. Gevaert, O. MethylMix: An R Package for Identifying DNA Methylation-Driven Genes. *Bioinformatics* **2015**, *31*, 1839–1841. [\[CrossRef\]](#)
22. Zhang, Y.; Yang, Y.; Chen, L.; Zhang, J. Expression Analysis of Genes and Pathways Associated with Liver Metastases of the Uveal Melanoma. *BMC Med. Genet.* **2014**, *15*, 29. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Gao, G.; Yu, Z.; Zhao, X.; Fu, X.; Liu, S.; Liang, S.; Deng, A. Immune Classification and Identification of Prognostic Genes for Uveal Melanoma Based on Six Immune Cell Signatures. *Sci. Rep.* **2021**, *11*, 22244. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Baqai, U.; Purwin, T.J.; Bechtel, N.; Chua, V.; Han, A.; Hartsough, E.J.; Kuznetsoff, J.N.; Harbour, J.W.; Aplin, A.E. Multi-Omics Profiling Shows BAP1 Loss Is Associated with Upregulated Cell Adhesion Molecules in Uveal Melanoma. *Mol. Cancer Res. MCR* **2022**, *20*, 1260–1271. [\[CrossRef\]](#)
25. Geng, Y.; Geng, Y.; Liu, X.; Chai, Q.; Li, X.; Ren, T.; Shang, Q. PI3K/AKT/mTOR Pathway-Derived Risk Score Exhibits Correlation with Immune Infiltration in Uveal Melanoma Patients. *Front. Oncol.* **2023**, *13*, 1167930. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Luo, H.; Ma, C. Identification of Prognostic Genes in Uveal Melanoma Microenvironment. *PLoS ONE* **2020**, *15*, e0242263. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Zhang, F.; Deng, Y.; Wang, D.; Wang, S. Construction and Verification of the Molecular Subtype and a Novel Prognostic Signature Based on Inflammatory Response-Related Genes in Uveal Melanoma. *J. Clin. Med.* **2023**, *12*, 861. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Jin, Y.; Wang, Z.; He, D.; Zhu, Y.; Gong, L.; Xiao, M.; Chen, X.; Cao, K. Analysis of Ferroptosis-Mediated Modification Patterns and Tumor Immune Microenvironment Characterization in Uveal Melanoma. *Front. Cell Dev. Biol.* **2021**, *9*, 685120. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Barbagallo, C.; Stella, M.; Broggi, G.; Russo, A.; Caltabiano, R.; Ragusa, M. Genetics and RNA Regulation of Uveal Melanoma. *Cancers* **2023**, *15*, 775. [\[CrossRef\]](#)
30. Wang, W.; Wang, S. The Prognostic Value of Immune-Related Genes AZGP1, SLCO5A1, and CTF1 in Uveal Melanoma. *Front. Oncol.* **2022**, *12*, 918230. [\[CrossRef\]](#)
31. Shi, X.; Xia, S.; Chu, Y.; Yang, N.; Zheng, J.; Chen, Q.; Fen, Z.; Jiang, Y.; Fang, S.; Lin, J. CARD11 Is a Prognostic Biomarker and Correlated with Immune Infiltrates in Uveal Melanoma. *PLoS ONE* **2021**, *16*, e0255293. [\[CrossRef\]](#)
32. Van Gils, W.; Lodder, E.M.; Mensink, H.W.; Kiliç, E.; Naus, N.C.; Brüggewirth, H.T.; van Ijcken, W.; Paridaens, D.; Luyten, G.P.; de Klein, A. Gene Expression Profiling in Uveal Melanoma: Two Regions on 3p Related to Prognosis. *Investig. Ophthalmol. Vis. Sci.* **2008**, *49*, 4254–4262. [\[CrossRef\]](#)
33. Yang, B.; Fan, Y.; Liang, R.; Wu, Y.; Gu, A. Identification of a Prognostic Six-Immune-Gene Signature and a Nomogram Model for Uveal Melanoma. *BMC Ophthalmol.* **2023**, *23*, 2. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Fan, Z.; Duan, J.; Luo, P.; Shao, L.; Chen, Q.; Tan, X.; Zhang, L.; Xu, X. SLC25A38 as a Novel Biomarker for Metastasis and Clinical Outcome in Uveal Melanoma. *Cell Death Dis.* **2022**, *13*, 330. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Zhang, Z.; Su, J.; Li, L.; Du, W. Identification of Precise Therapeutic Targets and Characteristic Prognostic Genes Based on Immune Gene Characteristics in Uveal Melanoma. *Front. Cell Dev. Biol.* **2021**, *9*, 666462. [\[CrossRef\]](#) [\[PubMed\]](#)

36. Smit, K.N.; Boers, R.; Vaarwater, J.; Boers, J.; Brands, T.; Mensink, H.; Verdijk, R.M.; van IJcken, W.F.J.; Gribnau, J.; de Klein, A.; et al. Genome-Wide Aberrant Methylation in Primary Metastatic UM and Their Matched Metastases. *Sci. Rep.* **2022**, *12*, 42. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Cai, M.-Y.; Xu, Y.-L.; Rong, H.; Yang, H. Low Level of PALMD Contributes to the Metastasis of Uveal Melanoma. *Front. Oncol.* **2022**, *12*, 802941. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Lei, S.; Zhang, Y. Integrative Analysis Identifies Key Genes Related to Metastasis and a Robust Gene-Based Prognostic Signature in Uveal Melanoma. *BMC Med. Genom.* **2022**, *15*, 61. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Steyaert, S.; Qiu, Y.L.; Zheng, Y.; Mukherjee, P.; Vogel, H.; Gevaert, O. Multimodal Deep Learning to Predict Prognosis in Adult and Pediatric Brain Tumors. *Commun. Med.* **2023**, *3*, 44. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Wang, S.; Zheng, K.; Kong, W.; Huang, R.; Liu, L.; Wen, G.; Yu, Y. Multimodal Data Fusion Based on IGERNNC Algorithm for Detecting Pathogenic Brain Regions and Genes in Alzheimer's Disease. *Brief. Bioinform.* **2023**, *24*, bbac515. [\[CrossRef\]](#)
41. Yu, G.; Yang, Y.; Yan, Y.; Guo, M.; Zhang, X.; Wang, J. DeepIDA: Predicting Isoform-Disease Associations by Data Fusion and Deep Neural Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 2166–2176. [\[CrossRef\]](#)
42. Cai, Z.; Poulos, R.C.; Liu, J.; Zhong, Q. Machine Learning for Multi-Omics Data Integration in Cancer. *iScience* **2022**, *25*, 103798. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Turrise, R.; Squillario, M.; Abate, G.; Uberti, D.; Barla, A. An Overview of Data Integration in Neuroscience with Focus on Alzheimer's Disease. *IEEE J. Biomed. Health Inform.* **2023**, 1–12. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Rappoport, N.; Shamir, R. NEMO: Cancer Subtyping by Integration of Partial Multi-Omic Data. *Bioinformatics* **2019**, *35*, 3348–3356. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Zhang, Y.; Parmigiani, G.; Johnson, W.E. ComBat-Seq: Batch Effect Adjustment for RNA-Seq Count Data. *NAR Genom. Bioinform.* **2020**, *2*, lqaa078. [\[CrossRef\]](#) [\[PubMed\]](#)
46. De los Campos, G.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A.; Crossa, J. Semi-Parametric Genomic-Enabled Prediction of Genetic Values Using Reproducing Kernel Hilbert Spaces Methods. *Genet. Res.* **2010**, *92*, 295–308. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Berisha, V.; Krantsevich, C.; Hahn, P.R.; Hahn, S.; Dasarathy, G.; Turaga, P.; Liss, J. Digital Medicine and the Curse of Dimensionality. *NPJ Digit. Med.* **2021**, *4*, 153. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Ciaramella, A.; Nardone, D.; Staiano, A. Data Integration by Fuzzy Similarity-Based Hierarchical Clustering. *BMC Bioinform.* **2020**, *21*, 350. [\[CrossRef\]](#)
49. Reggiani, F.; Ambrosio, M.; Croce, M.; Tanda, E.T.; Spagnolo, F.; Raposio, E.; Petito, M.; El Rashed, Z.; Forlani, A.; Pfeffer, U.; et al. Interdependence of Molecular Lesions That Drive Uveal Melanoma Metastasis. *Int. J. Mol. Sci.* **2023**, *24*, 15602. [\[CrossRef\]](#)
50. Piovesan, D.; Hatos, A.; Minervini, G.; Quaglia, F.; Monzon, A.M.; Tosatto, S.C.E. Assessing Predictors for New Post Translational Modification Sites: A Case Study on Hydroxylation. *PLoS Comput. Biol.* **2020**, *16*, e1007967. [\[CrossRef\]](#)
51. Ferrier, S.T.; Burnier, J.V. Novel Methylation Patterns Predict Outcome in Uveal Melanoma. *Life* **2020**, *10*, 248. [\[CrossRef\]](#)
52. Koroknai, V.; Szász, I.; Hernandez-Vargas, H.; Fernandez-Jimenez, N.; Cuenin, C.; Herceg, Z.; Vízkeleti, L.; Ádány, R.; Ecsedi, S.; Balázs, M. DNA Hypermethylation Is Associated with Invasive Phenotype of Malignant Melanoma. *Exp. Dermatol.* **2020**, *29*, 39–50. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Field, M.G.; Kuznetsov, J.N.; Bussies, P.L.; Cai, L.Z.; Alawa, K.A.; Decatur, C.L.; Kurtenbach, S.; Harbour, J.W. BAP1 Loss Is Associated with DNA Methylomic Repatterning in Highly Aggressive Class 2 Uveal Melanomas. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2019**, *25*, 5663–5673. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [\[CrossRef\]](#)
55. Wilkerson, M.D.; Hayes, D.N. ConsensusClusterPlus: A Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* **2010**, *26*, 1572–1573. [\[CrossRef\]](#)
56. Brock, G.; Pihur, V.; Datta, S.; Datta, S. clValid: An R Package for Cluster Validation. *J. Stat. Softw.* **2008**, *25*, 1–22. [\[CrossRef\]](#)
57. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2012**, *2*, 401–404. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [\[CrossRef\]](#)
59. Sato, H. Joint Singular Value Decomposition Algorithm Based on the Riemannian Trust-Region Method. *JSIAM Lett.* **2015**, *7*, 13–16. [\[CrossRef\]](#)
60. Townsend, J.; Koep, N.; Weinchwald, S. Pymanopt: A Python Toolbox for Optimization on Manifolds Using Automatic Differentiation. *J. Mach. Learn. Res.* **2016**, *17*, 1–5.
61. Mao, Y.; Gide, T.N.; Adegoke, N.A.; Quek, C.; Maher, N.; Potter, A.; Patrick, E.; Saw, R.P.M.; Thompson, J.F.; Spillane, A.J.; et al. Cross-Platform Comparison of Immune Signatures in Immunotherapy-Treated Patients with Advanced Melanoma Using a Rank-Based Scoring Approach. *J. Transl. Med.* **2023**, *21*, 257. [\[CrossRef\]](#)
62. Vihinen, M. How to Evaluate Performance of Prediction Methods? Measures and Their Interpretation in Variation Effect Analysis. *BMC Genom.* **2012**, *13* (Suppl. S4), S2. [\[CrossRef\]](#) [\[PubMed\]](#)

63. Carraro, M.; Monzon, A.M.; Chiricosta, L.; Reggiani, F.; Aspromonte, M.C.; Bellini, M.; Pagel, K.; Jiang, Y.; Radivojac, P.; Kundu, K.; et al. Assessment of Patient Clinical Descriptions and Pathogenic Variants from Gene Panel Sequences in the CAGI-5 Intellectual Disability Challenge. *Hum. Mutat.* **2019**, *40*, 1330–1345. [[CrossRef](#)] [[PubMed](#)]
64. Foroutan, M.; Bhuva, D.D.; Lyu, R.; Horan, K.; Cursons, J.; Davis, M.J. Single Sample Scoring of Molecular Phenotypes. *BMC Bioinform.* **2018**, *19*, 404. [[CrossRef](#)] [[PubMed](#)]
65. Bhuva, D.D.; Cursons, J.; Davis, M.J. Stable Gene Expression for Normalisation and Single-Sample Scoring. *Nucleic Acids Res.* **2020**, *48*, e113. [[CrossRef](#)] [[PubMed](#)]
66. Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCr: Visualizing Classifier Performance in R. *Bioinformatics* **2005**, *21*, 3940–3941. [[CrossRef](#)] [[PubMed](#)]
67. Lalonde, E.; Ewens, K.; Richards-Yutz, J.; Ebrahimzadeh, J.; Terai, M.; Gonsalves, C.F.; Sato, T.; Shields, C.L.; Ganguly, A. Improved Uveal Melanoma Copy Number Subtypes Including an Ultra-High-Risk Group. *Ophthalmol. Sci.* **2022**, *2*, 100121. [[CrossRef](#)] [[PubMed](#)]
68. Dogrusöz, M.; Ruschel Trasel, A.; Cao, J.; Çolak, S.; Van Pelt, S.I.; Kroes, W.G.M.; Teunisse, A.F.A.S.; Alsafadi, S.; Van Duinen, S.G.; Luyten, G.P.M.; et al. Differential Expression of DNA Repair Genes in Prognostically-Favorable versus Unfavorable Uveal Melanoma. *Cancers* **2019**, *11*, 1104. [[CrossRef](#)]
69. Aryee, M.J.; Jaffe, A.E.; Corrada-Bravo, H.; Ladd-Acosta, C.; Feinberg, A.P.; Hansen, K.D.; Irizarry, R.A. Minfi: A Flexible and Comprehensive Bioconductor Package for the Analysis of Infinium DNA Methylation Microarrays. *Bioinformatics* **2014**, *30*, 1363–1369. [[CrossRef](#)]
70. Dowle, M.; Srinivasan, A. Data.Table: Extension of “Data.Frame”. *R Package Version* **2021**, 1.14.2.
71. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. Circlize Implements and Enhances Circular Visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812. [[CrossRef](#)]
72. Gu, Z.; Eils, R.; Schlesner, M. Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics* **2016**, *32*, 2847–2849. [[CrossRef](#)]
73. Soetaert, K. plot3D: Plotting Multi-Dimensional Data. 2021. *R Package Version* **2022**, 1.4.
74. Hansen, K. IlluminaHumanMethylation450kanno.ilmn12.Hg19: Annotation for Illumina’s 450k Methylation Arrays. *R Package Version* **2016**, 0.6.0.
75. Carlson, M. Org.Hs.Eg.Db: Genome Wide Annotation for Human. *R Package Version* **2019**, 3.15.0.
76. Wilke, O.C. Cowplot: Streamlined Plot Theme and Plot Annotations for “Ggplot2”. *R package version* **2020**, 1.1.1.
77. Yan, L. Ggvenn: Draw Venn Diagram by “Ggplot2”. *R Package Version* **2023**, 0.1.10.
78. Ge, S.X.; Jung, D.; Yao, R. ShinyGO: A Graphical Gene-Set Enrichment Tool for Animals and Plants. *Bioinformatics* **2020**, *36*, 2628–2629. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.