

RESEARCH

Open Access



# Transcriptional patterns of cancer-related genes in primary and metastatic tumours revealed by machine learning

Faeze Keshavarz-Rahaghi<sup>1,2</sup>, Erin Pleasance<sup>1</sup> and Steven J. M. Jones<sup>1,3,4\*</sup>

## Abstract

**Background** A key to understanding cancer is to determine the impact on the cellular pathways caused by the repertoire of DNA changes accrued in a cancer cell. Exploring the interactions between genomic aberrations and the expressed transcriptome can not only improve our understanding of the disease but also identify potential therapeutic approaches.

**Results** Using random forest models, we successfully identified transcriptional patterns associated with the loss of wild-type activity in cancer-related genes across various tumour types. While genes like *TP53* and *CDKN2A* exhibited unique pan-cancer transcriptional patterns, others like *ATRX*, *BRAF*, and *NRAS* showed tumour-type-specific expression patterns. We also observed that genes like *AR* and *ERBB4* did not lead to strong detectable patterns in the transcriptome when disrupted. Our investigation has also led to the identification of genes highly associated with transcriptional patterns. For instance, *DRG2* emerged as the top contributor in classification of *ATRX* alterations in lower-grade gliomas and was significantly downregulated in *ATRX* mutant tumours. Additionally, transcriptional features important in classification of *PTEN* aberrations, such as *CDC48*, *AURKA*, and *CDC20*, were found to be closely related to *PTEN* function.

**Conclusions** Our findings demonstrate the utility of machine learning in interpretation of cancer genomic data and provide new avenues for development of targeted therapies tailored to individual patients with cancer. Our analysis on the transcriptome revealed genes with expression levels strongly correlated with alterations in cancer-related genes. Additionally, we identified *AURKA* inhibitors as potential therapeutic option for tumours with alterations in tumour suppressors like *FBXW7* or *NSD1*.

**Keywords** Transcriptome, Machine learning, Cancer, Random Forest

## Background

Cancer remains a leading cause of mortality worldwide, with treatment efficacy and adverse effects varying among individual patients [1, 2]. Cancer drivers are frequently mutated genes, playing important roles in cancer initiation, progression, and maintenance [3–6]. Incorporating molecular data into clinical practice enables personalized therapies and improves patient outcomes [1, 2, 7]. Pivotal initiatives including The Cancer Genome Atlas (TCGA) and the British Columbia (BC) Cancer Agency's Personalized Oncogenomics (POG) program

\*Correspondence:

Steven J. M. Jones

sjones@bcgsc.ca

<sup>1</sup> Canada's Michael Smith Genome Sciences Centre at BC Cancer, Vancouver, BC, Canada

<sup>2</sup> Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada

<sup>3</sup> Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

<sup>4</sup> Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

have sequenced hundreds to thousands of tumour samples, generating genome sequences and gene expression profiles [7, 8].

While DNA sequencing is widely used in clinical settings, and most regulatory-approved targeted treatments by the United States Food and Drug Administration are based on genomic alterations [9–11], only a subset of tumours harbour actionable genomic alterations and therapies guided solely by DNA sequencing do not always lead to response. [2, 12]. Furthermore, a significant proportion of mutations identified through whole-genome sequencing occur in regions not transcribed as RNA for which it is challenging to infer role in tumour function [3]. Investigating the transcriptome aids in comprehending the effects of genetic aberrations on cellular pathways [1, 13, 14]. The dynamic nature of the transcriptome which is influenced by cell type, cellular state, regulatory mechanisms, and other factors provides insights into molecular changes in cancer [1]. Gene expression data has proven valuable for drug sensitivity prediction and detection of cancer subtypes with unique clinical outcomes [15–19].

Interpreting the impact of different types of variations on gene function is feasible through investigating the transcriptome [1]. Single-nucleotide variants (SNVs), small insertions/deletions (INDELs), copy number alterations (CNAs), and structural variants (SVs) are prevalent genomic aberrations which can be associated with transcriptional patterns and combined to predict tumour features [3, 9, 13, 14, 20]. Machine learning (ML) methods have demonstrated success in analysing pan-cancer genome-wide expression data for diverse applications, including predicting tumour types, identifying molecular signatures of metastasis, detecting pathway activation, and uncovering prognostic genes [21–24]. These methods have identified transcriptional signatures of disease with significance in prognosis and treatment [13, 14, 16]. However, prior studies were constrained to a limited number of genes and/or tumour types, while a broader gene list was explored only when the primary focus was on distinguishing cancer state from normal [25]. Consequently, comprehensive research regarding the impact of gene alterations on the transcriptome remains lacking.

In this study, we used ML to determine whether genetic aberrations in cancer driver genes generate specific and detectable transcriptional changes unique to those genes. Through an analysis of pan-cancer whole transcriptome data, we identified transcriptional changes associating with alterations in specific cancer-related genes, either spanning multiple tumour types or specific to a particular tumour type. Using random forest (RF) models, we pinpointed gene transcripts contributing to these patterns and examined their relationships with the driver genes

under investigation. These findings provide insights into oncogenic mechanisms and have the potential to help identify active and targetable cellular pathways in primary and metastatic tumours.

## Results

### Expression profiles from tumour cohorts

To explore transcriptional patterns related to tumour mutations, we collated 8726 tumour samples from TCGA and 608 from POG studies, a total of 9334 patients, which include complete DNA mutation and RNA expression data. 56,645 overlapping coding and non-coding genes were identified. PCA plots revealed clustering of samples from the two studies, despite being from various tumour types and a mix of primary and metastatic tumours (Additional file 1: Figures S1–S3). Given the absence of distinct systematic separation between the TCGA and POG datasets, the samples were merged for downstream classification analyses.

Frequently mutated and cancer driver genes were reviewed [3, 5, 6], to select 50 genes as the focus of this study (Additional file 1: Table S1). Using somatic and germline SNV/INDEL data, samples were categorized into three groups: with “impactful” mutations, including nonsense and missense events, with “non-impactful” mutations, including synonymous and UTR events, and with wild-type gene copies.

### Random Forest classification improved by combining SNV/INDEL and copy number data

To evaluate whether transcriptional patterns associated with SNVs/INDELs in the genes of interest exist, ML models were used to classify tumour transcriptomes based on mutational status of these genes. RF, Support Vector Machine (SVM), and Neural Network (NN) models were tested on mutational status of *TP53* (Additional file 1: Figure S4). Among these, RF outperformed the others and surpassed the performance of the XGBoost model previously used by Zhang et al. to classify p53 pathway activity [26]. Additionally, our prior in-depth analysis on *TP53* validated the suitability of RF models for this type of task [13]. Thus, RFs were used for all subsequent classification tasks.

Only samples harbouring “impactful” mutations and wild-type copies were included. Five-fold CV analyses were conducted, and F1 scores were calculated (Fig. 1A). For many genes, small variants were found to be relatively infrequent, which contributed to class imbalance and adversely affected the model’s overall performance as evidenced by low F1 scores. We therefore incorporated other types of somatic variation, CNA and SV data, into the analysis. Four different sets of analyses were conducted: the first focused solely on the impact of SNVs

and INDELS, while the other three examined combinations of SNVs/INDELS with CNAs, SNVs/INDELS with SVs, and finally SNVs/INDELS with both CNAs and SVs.

Average F1 scores indicated that adding copy number data significantly improved performance, increasing the F1 score by ~19.3% on average (Additional file 1: Figure S5). In contrast, inclusion of structural variant data only increased the F1 score by ~0.1%. The limited effect of SVs might be attributed to the small number of samples carrying such variations in our data, and given the difficulty in confirming the impact of SVs on the genes of interest which can lead to ambiguity during training, SV data was excluded from further analysis.

For most genes, the inclusion of CNAs enhanced the model's performance (Fig. 1B). Consequently, CNAs were incorporated into downstream analysis for all genes except *APC*, *ATRX*, *BRAF*, *KRAS*, and *TP53*, for which performance was not improved. Robustness of these findings was confirmed with consistent results on an independent 10% test set (Additional file 1: Figure S6). Incorporating CNAs not only aligns with biological understanding of different ways genes are impacted in cancer [27], but also helps mitigate the class imbalance problem by improving the mutant to wild-type ratio (Fig. 1C).

#### Tumour type-specific analysis and class imbalance at the tumour type level

Assessing the RF model's performance across the 33 TCGA cancer cohorts indicated that for certain genes, like *ARIDIA*, the model achieved high F1 scores across diverse tumour types (Fig. 2A). For other genes including *BRAF*, strong performance was confined to limited tumour types, particularly thyroid carcinoma and cutaneous melanoma (Fig. 2B). Therefore, it was crucial to ensure that the RF model was not simply learning to classify samples based on their cancer type. F1 scores for individual tumour types were calculated for each gene (Additional file 1: Figure S7), and a *z*-test was employed to find tumour types that could be effectively classified based on alterations in the 50 genes of interest (Additional file 1: Table S1). Evaluating performance on the selected tumour types compared to those from pan-cancer analyses revealed which genes benefitted from tumour-type-specific analysis (Fig. 3A). If the average F1

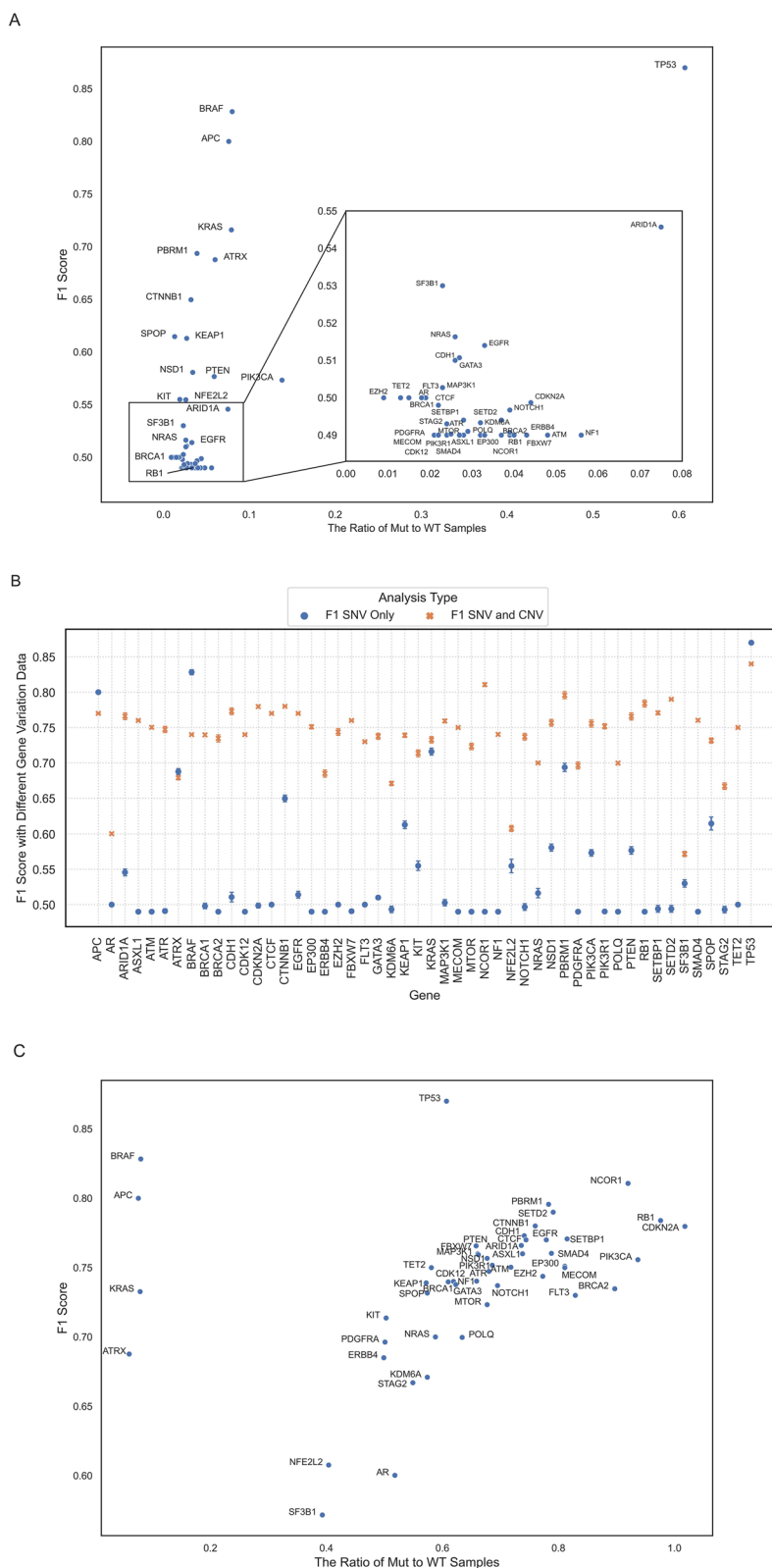
score improved by more than 5%—as observed for genes like *BRAF*, *ATRX*, and *NRAS*—selected tumour types were utilized for downstream analysis. Otherwise, the transcriptional pattern was interpreted to be more universal, and all tumour types were included to ensure a broader sample set and maximize the likelihood of identifying more generalizable transcriptional patterns.

We next explored the effect of class imbalance within individual tumour types. For *BRAF*, where the model performed well only in thyroid carcinoma and cutaneous melanoma, balancing the underperforming tumour types improved performance considerably in colorectal and stomach adenocarcinomas (Additional file 1: Figure S8). Similarly, balancing for all 50 genes (Additional file 1: Figure S9) identified balanced tumour types where the model performed strongly for each gene (Additional file 1: Table S2). Furthermore, analyses were conducted on balanced sets encompassing all tumour types. Results from pan-cancer analysis, specific tumour types, balanced sets of all tumour types, and balanced sets of specific tumour types were then compared (Fig. 3B). Down-sampling to balance sets of specific tumour types was pursued only when the average F1 score improved by over 5%, as applied to *ARIDIA*, *ASXL1*, *BRAF*, *BRCA1*, *CDK12*, *CTNNB1*, *KEAP1*, *MTOR*, *NOTCH1*, *PBRM1*, *PDGFRA*, *SETBP1*, *SETD2*, *SMAD4*, and *SPOP* genes, with a mean F1 improvement of 9.4% (range 6–17%). Balancing across all tumour types did not notably improve performance for any genes.

For genes where the focus of downstream analysis was on multiple tumour types (balanced or unbalanced), transcriptional modifications were investigated both within individual cancer types and across all selected tumour types (Additional file 1: Tables S3–S17). When necessary, analyses were repeated on a subset of cancer types (Additional file 1: Tables S18–S22). Subsequently top-ranked genes from the classification results were extracted, and genes in close chromosomal proximity were excluded (Additional file 1: Table S23). Final lists of highly contributing genes were reviewed (Additional file 1: Tables S24–S56) [28–163], which revealed adjustments necessary specifically for *APC* and *BRAF*. Nearly all the top genes with known associations with *APC* were linked to colorectal cancers, despite training across all tumour types (Additional file 1: Table S57) [28, 164–170],

(See figure on next page.)

**Fig. 1** F1 score of classification **A** when only samples with SNVs/INDELS were labeled as mutant versus the ratio of samples containing mutations over samples with wild-type gene copies, **B** with different combinations of gene alterations. Blue circles represent results when only samples with SNVs/INDELS were labeled as mutant and orange crosses show the results when samples with either SNVs/INDELS or CNAs were labeled as mutant, **C** when samples with either SNVs/INDELS or CNAs were labeled as mutant versus the ratio of samples containing alterations over samples with wild-type gene copies



**Fig. 1** (See legend on previous page.)

and the model's performance was lacking across all tumour types, except colorectal cancer (Additional file 1: Figure S10). As this suggests the model was capturing signals related to colorectal cancer rather than the effects of *APC* mutations, the second-best mode of analysis was selected which utilized a balanced set of colorectal cancers for analysis (Fig. 3B). Similarly, almost all genes with established literature associations to *BRAF* were reported in thyroid cancer, despite using balanced sets of thyroid (396 samples) and colorectal cancers (112 samples) in the analysis (Additional file 1: Table S28). To ensure that the larger number of thyroid tumours did not obscure the model's decision-making process, separate training was conducted on thyroid and colorectal samples. As anticipated, the majority of the top genes from thyroid classification overlapped with the initial list, while those important for classification of colorectal tumours were notably different (Additional file 1: Tables S58 and S59) [50–61, 171–174].

To further validate our classifier approach, we used the Cancer Dependency Map (DepMap), a collection of data from 1673 cancer cell lines from 96 primary diseases. Expression, point mutation, and copy number data for tumour cell lines were obtained from DepMap [175]. Due to notable differences in tumour type representation between TCGA and DepMap data, the three genes with strongest pan-cancer transcriptional patterns were selected, *CDKN2A*, *RBI*, and *TP53*. DepMap expression data only includes protein-coding genes, and therefore RF models were retrained using this subset of genes. The trained RFs were tested on the DepMap data and resulted in F1 scores of 0.57 for both *CDKN2A* and *RBI*, and 0.79 for *TP53*. These results are encouraging given the substantial difference between training and testing tumour types, including the presence of childhood cancers in DepMap not seen during training. Moreover, the variation in F1 scores among the three genes indicates that genes with higher F1 and top Gini scores have more generalizable patterns.

### Examining transcriptional patterns

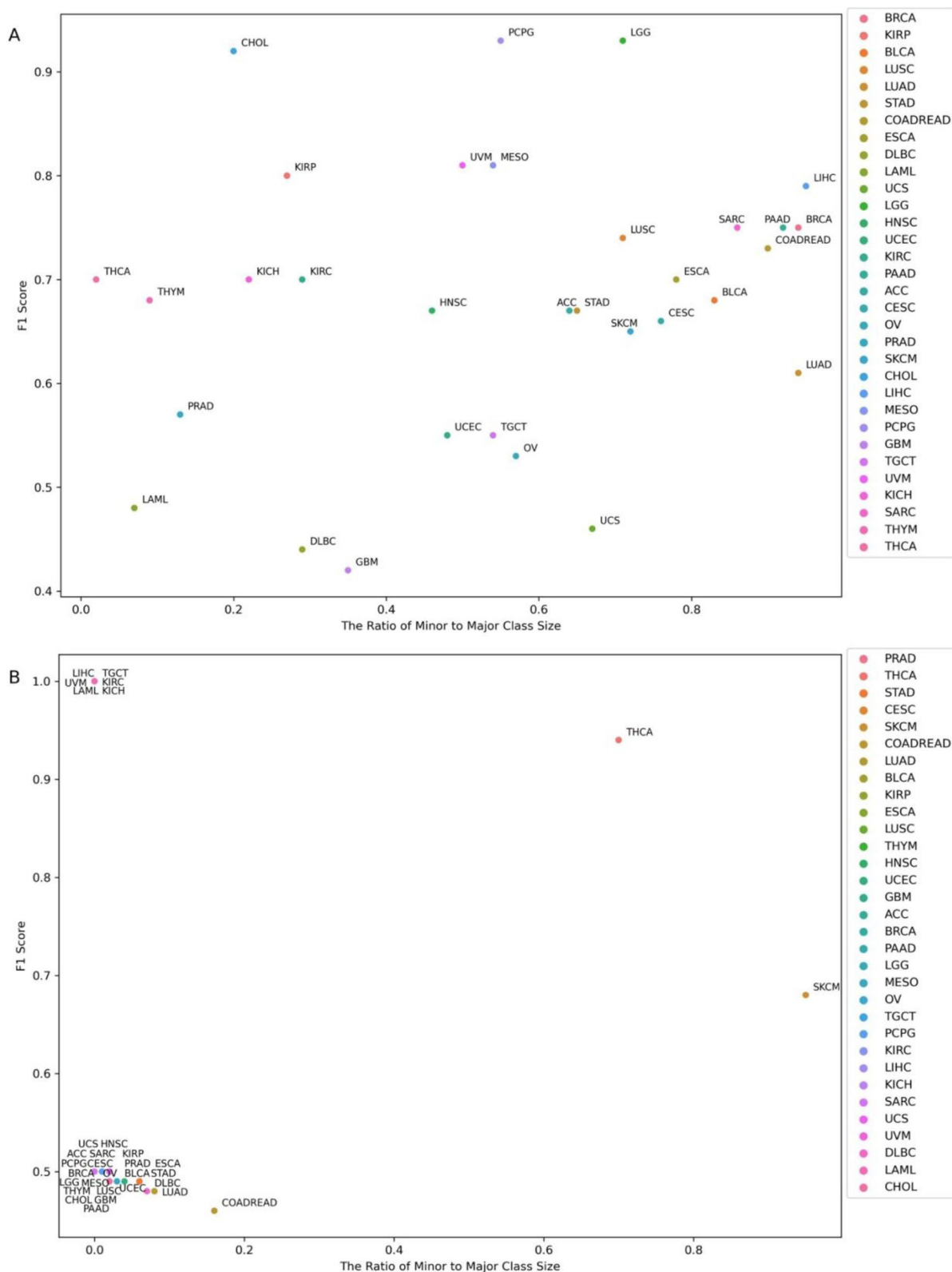
The optimal analysis settings for each of the 50 genes were used to train RF models, with chromosomally proximal genes removed. Genes contributing significantly to classification were identified based on Gini importance (examples in Additional file 1: Figures S11 and S12). Genes below significance thresholds (Additional file 1: Figures S13–S16; see Methods) were categorized as having no or weak transcriptional patterns associated with their genomic alterations, including *ERBB4*, *POLQ*, and *AR* (Fig. 4). Furthermore, a few genes showed high F1 but low Gini scores, suggesting potential overfitting (Fig. 4). This may be attributed to the limited sample sizes, as analyses for these genes were restricted to balanced sets of only one or two tumour types. These genes, along with those showing no or weak transcriptional patterns, were excluded from further analysis.

For the remaining genes, the top 15 features with the highest influence on classification were extracted (Additional file 1: Tables S24–S56). Number of highly contributing features to classification of all studied genes can be found in Fig. 5. A literature review identified known associations between these features and the target gene. While clear connections were found for well-studied genes, no known associations were identified for lesser-studied genes, like *KDM6A* and *NSD1*. Our findings suggest that there is value in further investigating these top-ranking genes through wet lab experiments, such as studies in cell lines.

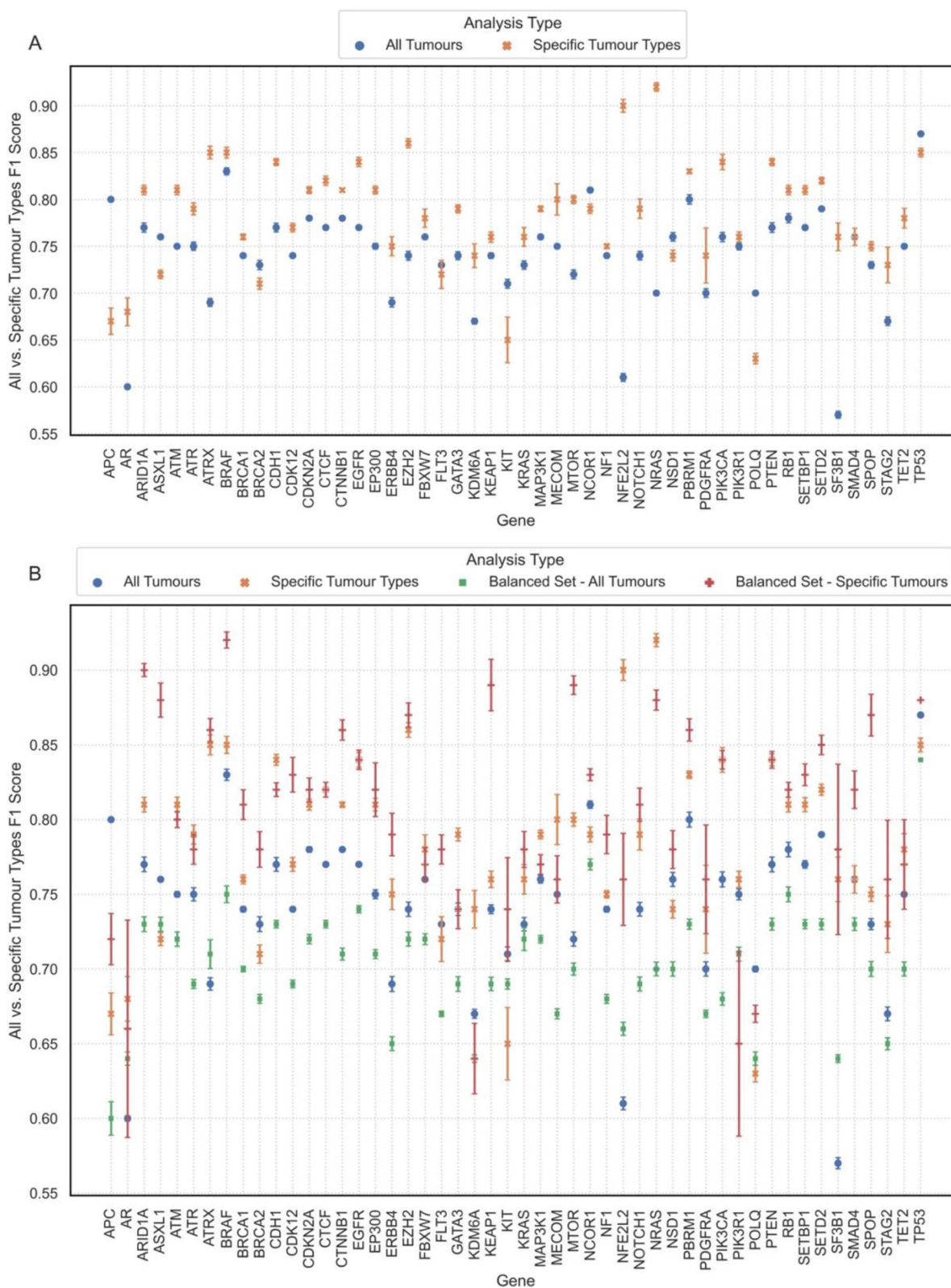
Comparison of overall F1 scores and top Gini scores illustrated stronger signals for genes such as *ATRX*, *BRAF*, and *TP53*. Consequently, it may be that targeting top-contributing genes to classification may offer therapeutic benefits in cancer treatment for patients carrying these driver mutations. It is speculated that the genes near the top right corner of the graph likely exhibit more generalizable transcriptional patterns. To further explore the nature of feature association types to mutational status of *ATRX*, *BRAF*, and *TP53* genes, SHAP analysis was conducted (Additional file 1: Figures S17–S19). This

(See figure on next page.)

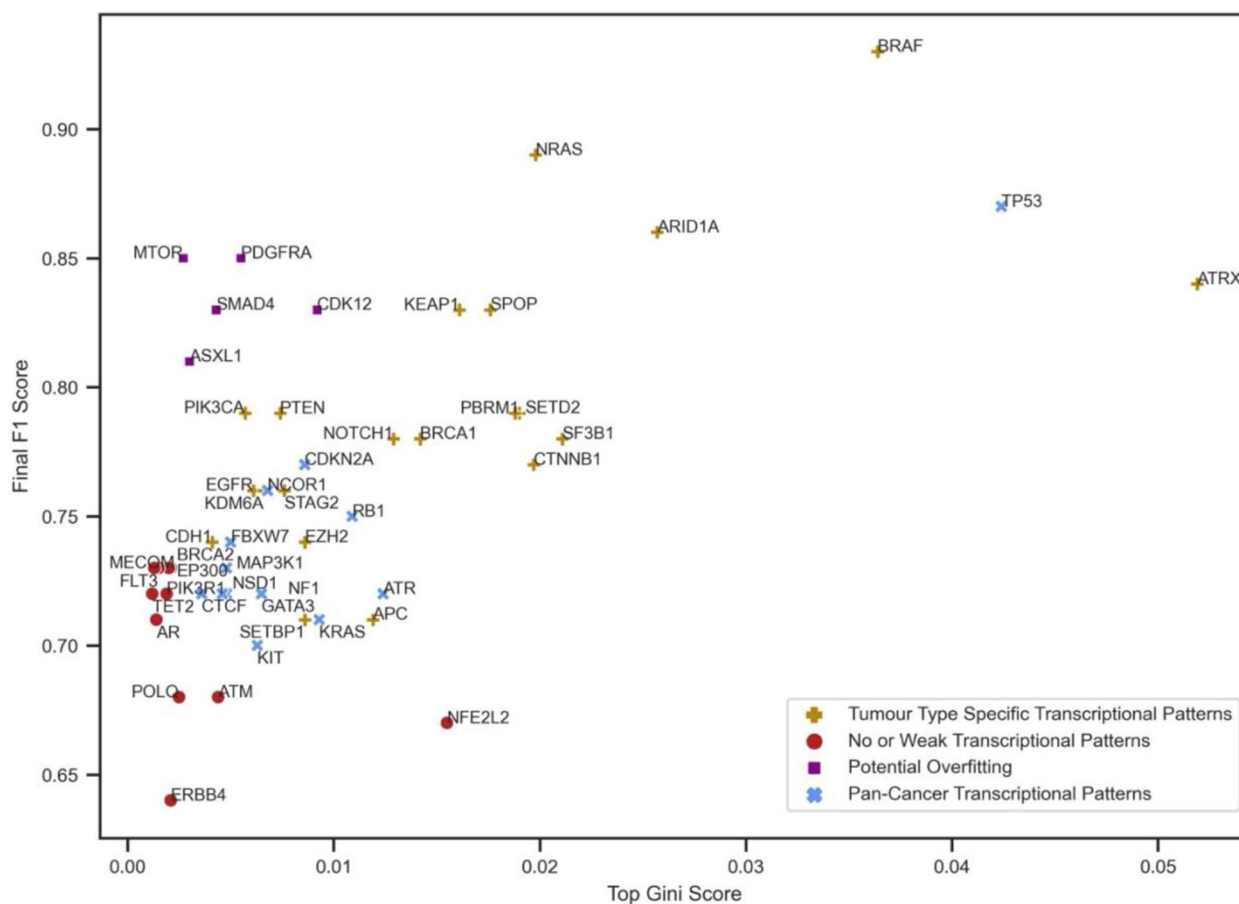
**Fig. 2** F1 scores for classification of samples based on **A** *ARID1A* and **B** *BRAF* alterations against the ratio of minor to major group sizes (with the minor group being the smaller of mutant or wild-type sample sets) across 33 TCGA tumour types (COAD and READ are combined). TCGA tumour types: ACC = Adrenocortical carcinoma, BLCA = Bladder Urothelial Carcinoma, BRCA = Breast invasive carcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL = Cholangiocarcinoma, COAD = Colon adenocarcinoma, DLBC = Lymphoid Neoplasm Diffuse Large B-cell Lymphoma, ESCA = Esophageal carcinoma, GBM = Glioblastoma multiforme, HNSC = Head and Neck squamous cell carcinoma, KICH = Kidney Chromophobe, KIRC = Kidney renal clear cell carcinoma, KIRP = Kidney renal papillary cell carcinoma, LAML = Acute Myeloid Leukaemia, LGG = Brain Lower Grade Glioma, LIHC = Liver hepatocellular carcinoma, LUAD = Lung adenocarcinoma, LUSC = Lung squamous cell carcinoma, MESO = Mesothelioma, OV = Ovarian serous cystadenocarcinoma, PAAD = Pancreatic adenocarcinoma, PCPG = Pheochromocytoma and Paraganglioma, PRAD = Prostate adenocarcinoma, READ = Rectum adenocarcinoma, SARC = Sarcoma, SKCM = Skin Cutaneous Melanoma, STAD = Stomach adenocarcinoma, TGCT = Testicular Germ Cell Tumours, THCA = Thyroid carcinoma, THYM = Thymoma, UCEC = Uterine Corpus Endometrial Carcinoma, UCS = Uterine Carcinosarcoma, UVM = Uveal Melanoma



**Fig. 2** (See legend on previous page.)



**Fig. 3** Comparison of F1 scores of classification based on alterations in genes of interest across **A** all versus specific tumour types, **B** four settings: 1. When all samples were used, 2. When samples from selected tumour types were used, 3. When balanced sets of all tumour types were used, and 4. When balanced sets of selected tumour types were used. The dots, crosses, and squares represent the average F1 scores, with whiskers indicating the standard deviation of F1 scores across 30 permutations per setting



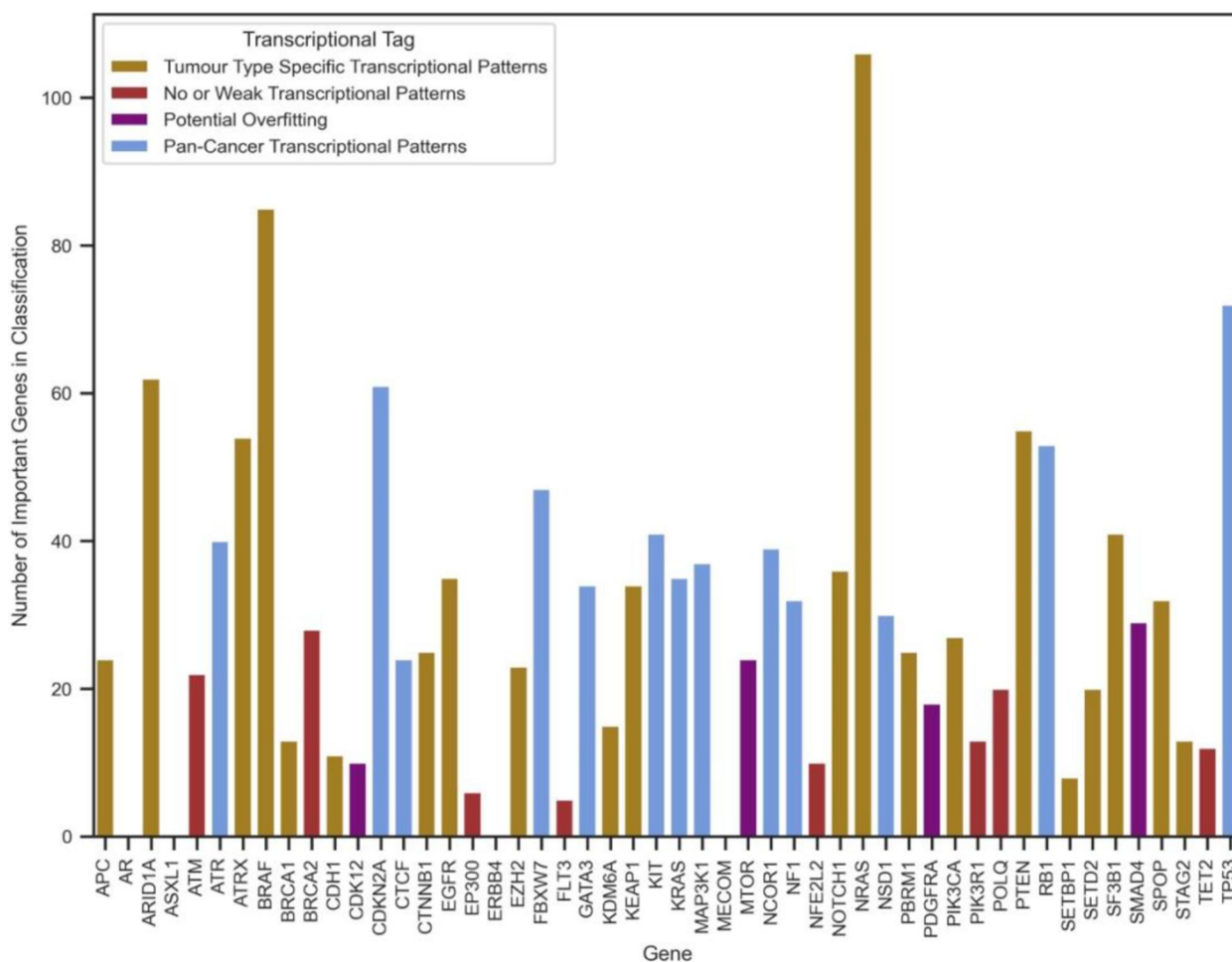
**Fig. 4** Final F1 score obtained from the optimal mode of analysis against the top Gini score belonging to the feature that contributed the most to classification. Colours and shapes correspond to identified transcriptional pattern categories associated with the genes studied in this work

approach further improves biological interpretation of the model by indicating whether a feature drives the prediction toward or away from the wild-type or mutated class. While for *BRAF*, higher expression of the top contributing genes was linked to a mutant prediction, for *ATRX* and *TP53*, elevated expression of top features was associated with wild-type function prediction. This pattern is evident in the SHAP value plots, where higher expression values (indicated in red) are shifted toward the negative class (mutant) for *BRAF* and the positive class (wild-type) for *ATRX* and *TP53*. This is an interesting observation since *BRAF* is an oncogene and its associated genes appear more active in the mutated state, while *ATRX* and *TP53* are tumour suppressors, whose functional networks are downregulated or disrupted in the presence of mutations.

We explored RF model's predictions for select tumour suppressor genes to assess what aligns with or adds to the existing knowledge about tumours with the associated driver mutations. *ATRX* showed one of the strongest transcriptional signatures in lower-grade gliomas

(LGG), with the *DRG2* gene emerging as the top feature in classification. *DRG2* was previously shown to be downregulated in *IDH1*-mutant gliomas, where *ATRX* mutations are common [47]. In our analysis, 179 out of 498 LGG samples (~36%) had an *ATRX* mutation, with *DRG2* downregulated in these samples (Fig. 6). While earlier research linked *DRG2* expression to *IDH1* mutations, our findings suggest that *DRG2* expression correlates more strongly with *ATRX* mutational status ( $p$ -value=6.79e-48) than with *IDH1* mutations ( $p$ -value=8.88e-7) (Fig. 6). SHAP analysis further supported this association between *DRG2* downregulation and *ATRX* mutation as lower expression values associated more strongly with the mutant class.

*RB1* exhibited a pan-cancer transcriptional signature. Among the 10 genes most associated with *RB1* classification, five were found to have known links to *RB1*. *AURKA*, second in the top-ranked set, was overexpressed in *RB1* mutant cells ( $P$ -value=3.40e-282; Fig. 7A). *AURKA* promotes cancer cell survival, and consistent with our model's finding, its inhibition has

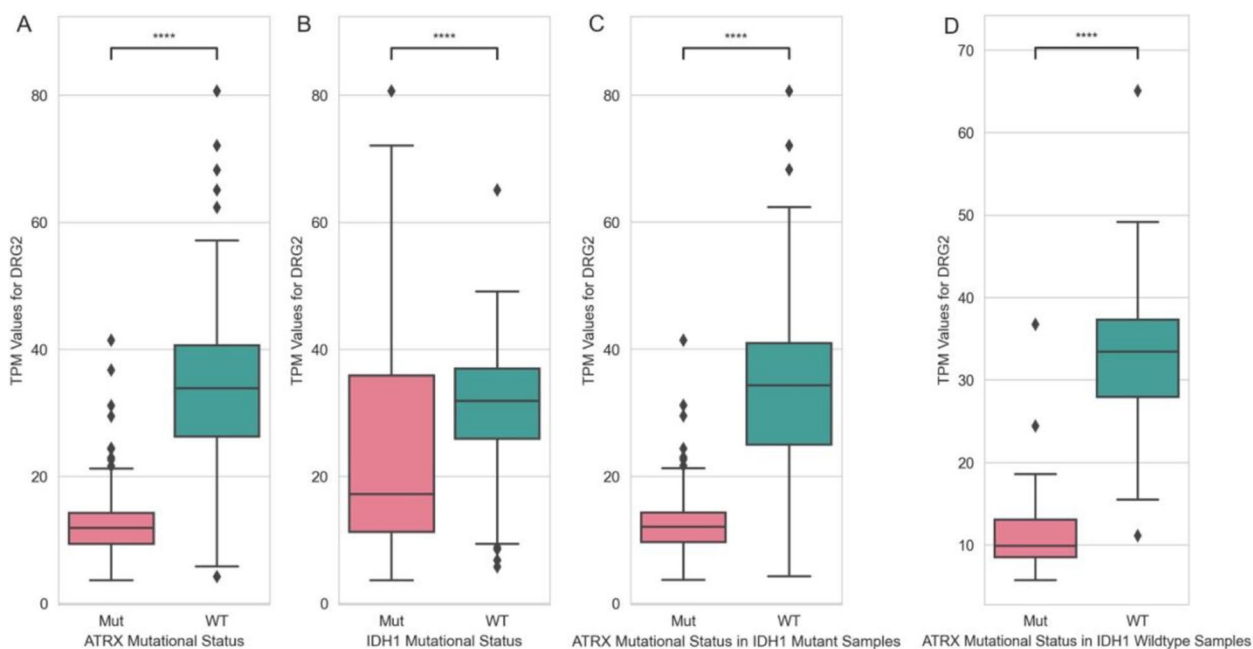


**Fig. 5** Number of genes found to be highly contributing to the classification task for each gene of interest. Colours represent the associated transcriptional pattern categories

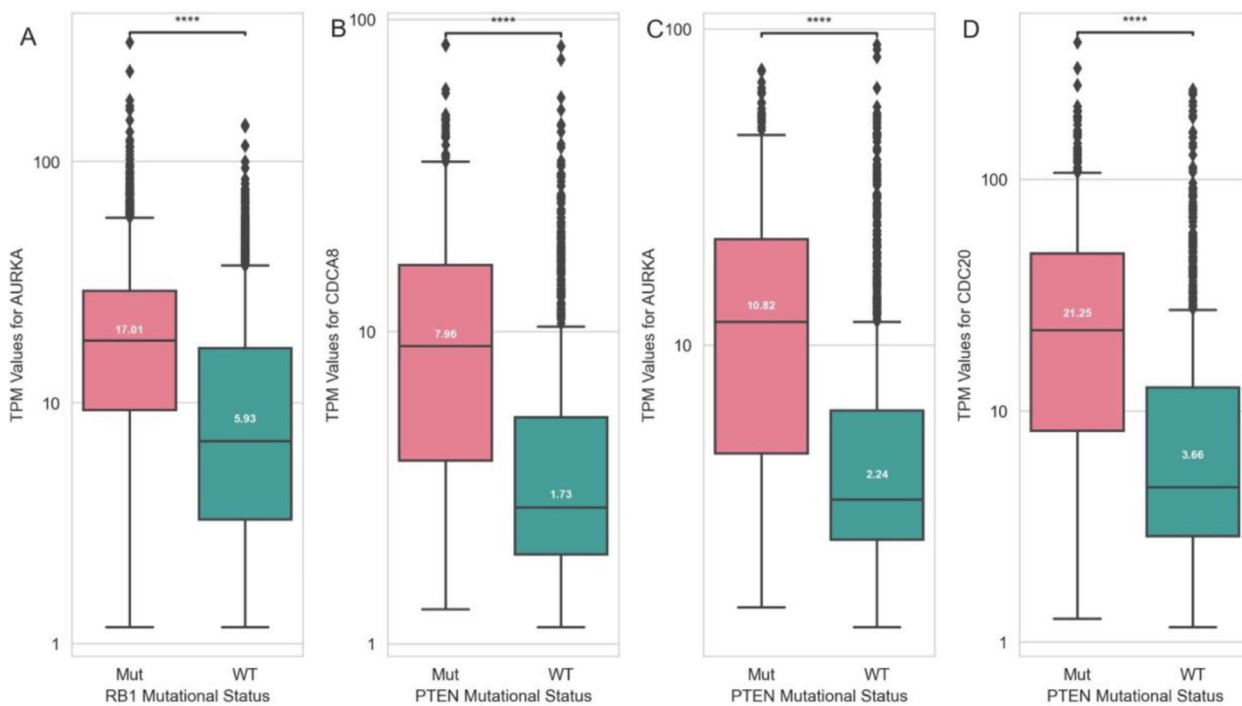
been shown to be effective in *RBI* mutant cells [149, 150]. Similarly, for *PTEN*, 8 out of top 10 genes in classification were found to have known associations with this gene. *AURKA*, ranked second, was upregulated in *PTEN* mutant LGG, PRAD, SARC, and SKCM tumours ( $P$ -value =  $1.04 \times 10^{-69}$ ; Fig. 7C). *CDCA8* and *CDC20*, ranked first and third, were also overexpressed in *PTEN* mutant samples ( $P$ -values of  $3.25 \times 10^{-72}$  and  $6.39 \times 10^{-74}$  respectively; Figs. 7B and 7D). Both genes are involved in cell cycle regulation and negatively affect survival [138, 140, 176], making them attractive targets for cancer therapies. Pathway analysis using top-ranked genes revealed changes in cancer-related pathways such as cell cycle, cell division, and mitosis (Additional file 1: Tables S60-S61) [177, 178].

A fully trained model on samples with impactful mutations and wild-type gene copies can reveal the effects of non-impactful mutations, as demonstrated in our previous work on *TP53* gene where RF classification correctly

identified silent mutations with splicing impacts [13]. The final trained RF models were therefore used to predict mutational status of samples harbouring non-impactful mutations (Additional file 1: Table S62). Notably, samples with non-impactful mutations predominantly predicted as mutant consistently involved intron variants. These samples were visualized in IGV (Additional file 1: Figures S20-S23), and it was observed that, in some cases, introns persisted in RNA molecules, indicating splicing errors. For example, *RBI* mutation c.1695+24,161 G>T or c.1695+29,182 insertion, between exons 17 and 18, led to un-spliced introns. There is also evidence that some variants might disrupt interactions between proteins and RNA strands, potentially affecting post-translational changes and protein production. For instance, *EZH2* mutation c.-8+9249 A>T, upstream of exon 1, might interfere with protein-RNA interactions, preventing splicing. Further investigation is required to



**Fig. 6** TPM expression values of *DRG2* gene in presence and absence of **A** *ATRX* gene mutations in all LGG samples, **B** *IDH1* gene mutations in all LGG samples, **C** *ATRX* gene mutations in *IDH1* mutant LGG samples, and **D** *ATRX* gene mutations in *IDH1* wild-type samples (*P*-values are obtained in a two-sided Mann–Whitney–Wilcoxon test with Bonferroni correction; *p*-value annotation legend: ns:  $5e-2 < p < 1$ ; \*:  $1e-2 < p < 5e-2$ ; \*\*:  $1e-3 < p < 1e-2$ ; \*\*\*:  $1e-4 < p < 1e-3$ ; \*\*\*\*:  $p < 1e-4$ )



**Fig. 7** Log 10 of TPM expression values for **A** *AURKA* gene in presence and absence of *RB1* gene mutations across all tumour types, **B** *CDC48*, **C** *AURKA*, and **D** *CDC20* genes in presence and absence of *PTEN* gene mutations across LGG, PRAD, SARC, and SKCM tumours (numbers inside the boxplots show median values)

understand how these variants affect the transcriptome and potentially play a role as oncogenic drivers.

Table 1 summarizes the genes studied in this work, with results from the tumour types where classification was most accurate. Details regarding top-ranked genes and affected pathways for *TP53* are omitted, as they were extensively discussed in prior work with very similar scores despite differences in genome versions [13].

## Discussion

The integration of transcriptomic data with genomic alterations gives a more comprehensive view of cancer biology and leads to discovery of novel biomarkers [13, 19]. This study examined transcriptomic changes arising from alterations in key cancer driver genes. Our findings revealed that cancer genes fall in distinct groups with respect to their impacts on the transcriptome. Many genes do not confer unique transcriptional patterns following alterations, which may reflect similar impacts on pathways and redundancy in transcriptional signalling. Some genes exhibit tumour-type-specific patterns, including *BRAF*, *ATRX*, and *PTEN*, where transcriptional signatures of mutation are strong but limited to specific tumour types. Conversely, specific cancer driver genes exhibit pan-cancer transcriptional patterns when altered. The most striking example is mutations in *TP53*, as previously described [13], but the findings presented here demonstrate that *CDKN2A*, *RBI*, and *NCOR1* also have notably unique pan-cancer transcriptional impacts. These relate to tumour biology and potentially to association with therapeutic sensitivity, particularly relevant for tumour suppressor mutations which are typically difficult to target [149, 150]. Classification of tumours based on *TP53* mutations revealed that *MYBL2* and *MDM2* play key roles in classification. *MYBL2* is upregulated in *TP53* mutant samples, while *MDM2* is upregulated in *TP53* wild-type samples [13]. This suggests targeting these genes in presence and absence of *TP53* mutations respectively could be an effective strategy for tumour suppression.

*ATRX* showed one of the strongest tumour-type-specific transcriptional patterns in LGG tumours, with *DRG2* gene having the highest Gini score. *DRG2* plays an important role in cell cycle arrest at the G2/M phase during mitosis and affects expression of cell cycle and checkpoint genes [179]. Meanwhile, *ATRX*, a chromatin remodeler, binds cell cycle transition genes, and its loss disrupts the maintenance of the G2/M checkpoint after irradiation [180]. Thus, we hypothesize that *ATRX* mutations affecting the cell cycle may influence *DRG2* expression level. Indeed, our analysis confirmed *DRG2* downregulation in *ATRX* mutant samples, with a stronger correlation to *ATRX* mutations than to *IDH1* mutations, while only correlation with *IDH1* mutations

have been discussed previously [47]. These findings suggest that restoring *DRG2* expression in *ATRX* mutant cases could hold therapeutic potential, especially since patients with *IDH1*-mutant tumours were shown to have better prognoses [181, 182].

*RBI* alterations had a pan-cancer transcriptional pattern. *AURKA*, a kinase with roles in mitosis and cancer cell survival [149, 150], contributed most significantly to *RBI* tumour classification. We observed *AURKA* overexpression in *RBI* mutant cells, making it a promising therapeutic target for patients with *RBI* mutations. Indeed, *AURKA* inhibitors have been described as effective against *RBI* mutant cells [149, 150]. This is particularly significant since *RBI*, a tumour suppressor, cannot be targeted directly.

*PTEN*, a frequently mutated tumour suppressor, had transcriptional patterns associated with several tumour types, with *AURKA* emerging as the second most important feature in classification. *AURKA* inhibitors have been shown to be effective in *PTEN* deficient mice [183], highlighting the effectiveness of our model in identifying actionable targets in *PTEN* mutant cases. We found that genes with the highest and third-highest Gini scores, *CDCA8* and *CDC20*, were overexpressed in *PTEN* mutant samples. *CDCA8* competes with *PTEN* for *AKT* binding [138], while *CDC20* physically interacts with *PTEN* in mitotic checkpoint complex [140], both negatively affecting survival [138, 176]. Thus, targeting them as well as *AURKA* could provide therapeutic benefits, particularly in *PTEN* mutant LGG, PRAD, SARC, and SKCM tumours.

Insights gained from the RF model can guide the discovery of new therapeutic strategies. For instance, *AURKA* inhibitors, effective in *RBI* and *PTEN* mutant tumours [149, 150, 183], could be trialed in tumours with mutations in less-studied tumour suppressors, such as *FBXW7* and *NSD1*, for which we found *AURKA* to be a key gene in classification. Additionally, *AURKA* is upregulated when *FBXW7* and *NSD1* are mutated. *FBXW7* has also been shown to negatively regulate *AURKA* [93, 94], further supporting the value in future experiments exploring targeting of *AURKA* in *FBXW7* mutant cell lines and tumours.

For genes with no or weak transcriptional patterns, several mechanisms could explain this outcome. One possibility is that the effect of these gene alterations may be dispersed across multiple genes and pathways, leading to weaker transcriptional signals. Alternatively, these alterations could primarily manifest at the protein level, influencing factors like post-translational modifications, stability, or availability, which may not be directly reflected in the transcriptome [184, 185]. Moreover, some alterations might be clonally diverse,

**Table 1** The 50 genes studied in this work, their mode of analysis, final F1 score, and corresponding top Gini score

Gene	Tumour types	Final F1 Score	Top Gini Score
APC	Balanced Set of COADREAD	0.71	0.0119
AR	KIRP	0.71	0.0014
ARID1A	Balanced Set of KICH, KIRP, LGG, PCPG and UVM	0.86	0.0257
ASXL1	Balanced Set of COADREAD	0.81	0.0030
ATM	BRCA, CESC, and PCPG	0.68	0.0044
ATR	All Tumour Types	0.72	0.0124
ATRX	LGG	0.84	0.0519
BRAF	Balanced Set of COADREAD and THCA	0.93	0.0364
BRAF	Balanced Set of COADREAD	0.86	0.0165
BRAF	Balanced Set of THCA	0.93	0.0487
BRCA1	Balanced Set of KICH, KIRP and UCEC	0.78	0.0142
BRCA2	All Tumour Types	0.73	0.0021
CDH1	KIRP, LIHC, PRAD, SARC, THYM, UCEC and UVM	0.74	0.0041
CDK12	Balanced Set of KIRP and UCEC	0.83	0.0092
CDKN2A	All Tumour Types	0.77	0.0086
CTCF	All Tumour Types	0.72	0.0036
CTNNB1	Balanced Set of HNSC, KIRC, PCPG and UVM	0.77	0.0197
EGFR	COADREAD, HNSC, KIRP, LGG and STAD	0.76	0.0061
EP300	BRCA, PCPG, THCA and UCEC	0.73	0.0020
ERBB4	CESC	0.64	0.0021
EZH2	COADREAD, KIRP, LGG and THYM	0.74	0.0086
FBXW7	All Tumour Types	0.74	0.0050
FLT3	All Tumour Types	0.73	0.0015
GATA3	All Tumour Types	0.72	0.0065
KDM6A	KIRP and UCEC	0.76	0.0066
KEAP1	Balanced Set of THCA	0.83	0.0161
KIT	All Tumour Types	0.70	0.0063
KRAS	All Tumour Types	0.71	0.0093
MAP3K1	All Tumour Types	0.73	0.0048
MECOM	All Tumour Types	0.73	0.0013
MTOR	Balanced Set of LGG and PCPG	0.85	0.0027
NCOR1	All Tumour Types	0.76	0.0068
NF1	All Tumour Types	0.72	0.0048
NFE2L2	KICH	0.67	0.0155
NOTCH1	Balanced Set of KIRC	0.78	0.0129
NRAS	LGG and PCPG	0.89	0.0198
NSD1	All Tumour Types	0.72	0.0046
PBRM1	Balanced Set of HNSC, KIRC, MESO, PCPG and UVM	0.79	0.0190
PDGFRA	Balanced Set of KICH	0.85	0.0055
PIK3CA	KIRP, PCPG and UVM	0.79	0.0057
PIK3R1	All Tumour Types	0.72	0.0019
POLQ	All Tumour Types	0.68	0.0025
PTEN	LGG, PRAD, SARC and SKCM	0.79	0.0074
RB1	All Tumour Types	0.75	0.0109
SETBP1	Balanced Set of COADREAD, HNSC and PRAD	0.71	0.0086
SETD2	Balanced Set of HNSC, KIRC, PCPG and UVM	0.79	0.0188
SF3B1	UVM	0.78	0.0211
SMAD4	Balanced Set of COADREAD	0.83	0.0043
SPOP	Balanced Set of KIRP, THCA and THYM	0.83	0.0176

**Table 1** (continued)

Gene	Tumour types	Final F1 Score	Top Gini Score
STAG2	KIRP	0.76	0.0076
TET2	All Tumour Types	0.72	0.0012
TP53	All Tumour Types	0.87	0.0424

affecting a subset of cells within the tumour [186], making effects harder to detect.

Using the RF classifier on samples with non-impactful mutations was especially valuable where alterations have ambiguous effects on RNA or protein function, helping to uncover insights that may otherwise remain unnoticed. In previous work, we demonstrated that samples with certain mutations labeled as silent actually caused splicing defects [13]. Here, we observed that intron variants might also have pathological effects, impairing the wild-type activity of genes like *NCOR1* and *RBI*. These results underscore the role of non-protein-altering mutations and the importance of detailed investigation of individual mutations.

This analysis is subject to several technical and biological confounding factors, including variations in sample acquisition, data processing, tumour purity, biopsy site, and stromal and immune composition. However, RF models are well-suited to overcome these challenges [187–189], and we observed that they outperformed SVMs, NNs, and XGBoost models. They are also less prone to overfitting compared to other models. However, certain limitations like small sample size and class imbalance can influence the performance, often compounding one another. For example, down-sampling to address class imbalance reduces available data, exacerbating the problem of small sample size. We observed signs of overfitting in some genes, particularly when the analysis was focused on balanced sets of a single or few tumour types. In such cases, although the F1 score was high, the top Gini score remained relatively low, suggesting that the model may have learned patterns specific to the limited dataset rather than generalizable trends. The best strategy to address these issues is to increase the sample size. Large ongoing projects, like the NIH Cancer Moonshot program and the UK Biobank, aim to recruit a large number of participants, to gather health data and sequence genomic material [190, 191]. The large datasets produced by these efforts will help uncover more generalizable patterns in genomics data. Moreover, follow-up studies investigating gene regulatory mechanisms, such as ChIP-seq experiments and protein interaction network analyses, can further enrich the findings from this work and provide validation. Therefore, it is essential to incorporate proteomic, epigenomic, and regulatory data

collection alongside genomic variation and expression profiling to support robust conclusions.

This study deepens our understanding of cancer biology and gene interactions, providing a framework for identifying molecular targets in tumours with previously untarargetable genes. Here, we showed the importance of combining transcriptomic and genomic data, with potential to facilitate development directions for new targeted therapies. These insights will be instrumental in navigating the ever-evolving landscape of cancer treatment and moving closer to more precise and effective solutions for cancer management.

## Conclusions

In this work, we demonstrated how ML can help unravel complexities of the cancer transcriptome, offering insights that accelerate the development of novel treatments. We identified potential drug targets for eliminating cancer cells by classifying samples based on gene alterations. This analysis led to discovery that *AURKA* inhibitors may be effective in cancers where tumour suppressors like *FBXW7* or *NSD1* are compromised. Additionally, we found a strong correlation between *DRG2* expression levels and *ATRX* gene mutations. These results underscore the potential of computational approaches in identifying new therapeutic strategies tailored to the unique molecular profiles of individual patients.

## Methods

### Data preprocessing

Expression and gene variation data for TCGA were downloaded from online resources [192–195], while data for the POG study was retrieved from Canada's Michael Smith Genome Sciences Centre servers [7]. Only samples with all data types were included, resulting in 8726 TCGA and 608 POG samples. A list of comparable genes overlapping between TCGA and POG datasets resulted in 56,645 transcribed genes, of which 18,606 (33%) are annotated as protein coding. Principal component analysis (PCA) plots were generated to visualize expression values (Additional file 1: Figures S1–S3). For validation analysis, tumour cell line data was also obtained from DepMap version 24Q4 (<https://depmap.org/portal>) [175].

The focus of the study was on frequently mutated genes and the ones essential in cancer biology. An overlap of

gene lists from related studies yielded 50 shared genes [3, 5, 6]. Tumour samples were grouped by mutational status (mutated vs wildtype) for each gene using somatic and germline mutation data. The samples in the mutated group were further divided into “impactful” and “non-impactful” categories based on expected consequences of mutations [13, 196]. Samples containing “impactful” mutations or wildtype gene copies were used to create feature matrices. For analyses including CNA and SV data, samples with copy number changes or structural variations were reclassified as “impactful”.

### Random Forests performance

Performance of RF, SVM, and NN models were compared on classification of *TP53* mutational status (Additional file 1: Figure S4). Main hyperparameters were fine-tuned using 90% of samples and validated on the remaining 10%. Samples were categorized as either wild-type or mutant based on SNV/INDEL data. The performance of the RF model was evaluated using 5-fold cross-validation (CV) across TCGA and POG datasets for all the genes of interest, and F1 scores were computed (Fig. 1A). To evaluate the impact of additional gene alterations, samples with CNAs or SVs were re-classified as mutant, and the RF model's performance was evaluated (Additional file 1: Figure S5). Since SVs had minimal impact on F1 scores, the focus was narrowed to SNVs/INDELS and CNAs (Fig. 1B). Samples with CNAs were only labeled as mutant if their inclusion resulted in a substantial improvement in F1 score. To validate this improvement, test F1 scores on 10% of samples were compared to the 5-fold CV F1 scores from the 90% training set (Additional file 1: Figure S6). The RF model was then trained on all TCGA and POG samples and the final F1 score were recorded (Fig. 1C).

### Tumour type-specific analysis and class imbalance at the tumour type level

F1 scores were computed across 33 TCGA tumour types (Fig. 2). To account for F1 score variability in some genes, a *z*-test with alpha of 0.1 was conducted, establishing a significance threshold of 0.728 (Additional file 1: Figure S7). For each gene of interest, tumour types with F1 scores above this threshold were selected for downstream analysis. If no tumour type met the threshold, the one with the highest F1 score was chosen (Additional file 1: Table S1). Model's performance was compared between pan-cancer and tumour type-specific analyses (Fig. 3A). Tumour-specific analyses were pursued only if the average F1 score improved by more than 5%.

The impact of class imbalance at the tumour type level was then examined. Balancing tumour types for *BRAF* mutations improved F1 scores (Additional file 1: Figure

S8), and thus, performance was tested on balanced sets of tumour types for all genes. A *z*-test on F1 scores with alpha of 0.1 resulted in a significance threshold of 0.785 (Additional file 1: Figure S9). Tumour types with F1 scores above this threshold or the one with the highest balanced F1 score (Additional file 1: Table S2) were used to perform 30 permutations of 5-fold CV. Additionally, balanced sets from all tumour types were analysed in the same way, and average F1 scores were compared across all settings (Fig. 3B). Subsequently, the analysis focused on balanced sets of specific tumour types only when the average F1 score showed an improvement of more than 5%. For multi-tumour-type analyses, model performance was investigated both within individual cancer types and across the whole set (Additional file 1: Tables S3-S17). If performance dropped for one of the tumour types, that tumour type was excluded, and performance was re-evaluated (Additional file 1: Tables S18-S22).

Subsequently, the lists of top-ranked genes in classification were examined. In some instances, the majority of top genes were located at nearby chromosomal regions to the gene under investigation, likely due to inclusion of CNAs in the analysis. To ensure that the observed transcriptional modifications were associated with alterations in the function of the genes of interest, nearby genes were iteratively removed until no cluster of physically adjacent genes appeared among the top-ranked genes (Additional file 1: Table S23) [197, 198]. The top genes in classification were then studied for their associations with the genes of interest, based on retrieved literature (Additional file 1: Tables S24-S56). In specific cases, like *APC* and *BRAF*, evaluating top genes in classification led to a revision of the analysis approach (Additional file 1: Figure S10 and Tables S57-S59).

### Examining transcriptional patterns

After training, thresholds were established for number of key genes in classification using a permutation-based method (Additional file 1: Figures S11 and S12) [13]. For certain genes of interest, no genes were found to have a significant impact, typically with low F1 or top Gini scores. These genes were considered to have no or weak transcriptional patterns (Additional file 1: Figures S13 and S14). To define a threshold, a *z*-test with alpha 0.1 was applied on 5-fold CV F1 scores and a lower percentile with alpha 0.25 on top Gini scores (Additional file 1: Figures S15 and S16). A lower threshold of 0.686 for F1 and a lower critical value of 0.0021 for Gini score were obtained. Genes below these thresholds were categorized as having no or weak transcriptional patterns (Fig. 4). Moreover, a few genes were identified with high F1 scores but low top Gini scores. These genes were flagged for potential overfitting, especially due to small training

sets. Consequently, these genes, along with those categorized as having no or weak transcriptional patterns, were excluded from further analysis.

For the remaining genes, associations between the top contributing genes and genes under investigation were gathered from the literature (Additional file 1: Tables S24–S56). The number of important genes in classification is shown in Fig. 5. These genes were further used in Gene Set Enrichment Analysis using the Database for Annotation, Visualization, and Integration Discovery [177, 178]. The top five enriched pathways with significant  $p$ -values are in Additional file 1: Table S60. SHAP importance was also found using the Python SHAP package for the top 15 features of the three genes with highest top Gini scores (Additional file 1: Figures S17–S19) [199]. The mutational status of samples with non-impactful mutations was also predicted. When the majority of samples were classified as mutant, these were further examined using Integrative Genomic Viewer (IGV) [200].

#### Abbreviations

ACC	Adrenocortical carcinoma
BC	British Columbia
BLCA	Bladder urothelial carcinoma
BRCA	Breast invasive carcinoma
CEC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	Cholangiocarcinoma
CNA	Copy number alteration
COAD	Colon adenocarcinoma
CV	Cross validation
DLBC	Lymphoid neoplasm diffuse large B-cell lymphoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and neck squamous cell carcinoma
IGV	Integrative Genomics Viewer
INDEL	Insertion/Deletion
KICH	Kidney chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute myeloid leukaemia
LGG	Brain lower-grade glioma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MESO	Mesothelioma
ML	Machine learning
NN	Neural Network
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCA	Principal component analysis
PCPG	Pheochromocytoma and Paraganglioma
POG	Personalized Oncogenomics
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
RF	Random Forest
SARC	Sarcoma
SKCM	Skin cutaneous melanoma
SNV	Single-nucleotide variation
STAD	Stomach adenocarcinoma
SV	Structural variant
SVM	Support Vector Machine
TCGA	The Cancer Genome Atlas
TGCT	Testicular germ cell tumours

THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine corpus endometrial carcinoma
UCS	Uterine carcinosarcoma
UVM	Uveal melanoma

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-025-02339-z>.

Additional file 1: This file contains a detailed description of the methodologies used for data analysis, along with supplementary tables S1–S62 and supplementary figures S1–S23. Table S1 – Tumour types selected for downstream analysis based on all samples. Table S2 – Tumour types selected for downstream analysis based on balanced sets of samples. Tables S3 to S17 – Number of mutant and wild-type samples and F1 scores based on alterations in *ARID1A*, *BRAF*, *BRCA1*, *CDH1*, *CTNNB1*, *EGFR*, *EZH2*, *KDM6A*, *NRAS*, *PBRM1*, *PIK3CA*, *PTEN*, *SETBP1*, *SETD2*, and *SPOP*. Tables S18 to S22 – Number of mutant and wild-type samples and F1 scores after excluding specific tumour types for *ARID1A*, *EGFR*, *EZH2*, *PBRM1*, and *SPOP*. Table S23 – Chromosomal regions excluded from analysis. Tables S24 to S56 – Top genes in classification of samples based on alterations in *APC*, *ARID1A*, *ATR*, *ATRX*, *BRAF*, *BRCA1*, *CDH1*, *CDKN2A*, *CTCF*, *CTNNB1*, *EGFR*, *EZH2*, *FBXW7*, *GATA3*, *KDM6A*, *KEAP1*, *KIT*, *KRAS*, *MAP3K1*, *NCOR1*, *NF1*, *NOTCH1*, *NRAS*, *NSD1*, *PBRM1*, *PIK3CA*, *PTEN*, *RB1*, *SETBP1*, *SETD2*, *SF3B1*, *SPOP*, and *STAG2*. Tables S57 to S59 – Top genes in classification of samples based on alterations in *APC* and *BRAF* under different settings. Table S60 – Top pathways affected by gene alterations. Table S61 – Top pathways affected by *BRAF* gene alterations in thyroid and colorectal cancers. Table S62 – Non-impactful mutations likely playing a role in pathogenesis. Figures S1 to S3 – PCA plots of POG, TCGA, and all samples. Figure S4 – Performance comparison across different models. Figure S5 – F1 scores based on different sets of gene alterations. Figure S6 – F1 score comparison between 5-fold CV and test set. Figure S7 – F1 scores distribution across all genes and tumour types. Figure S8 – Tumour-type-level F1 scores for *BRAF*. Figure S9 – F1 scores distribution based on balanced sets. Figure S10 – Tumour-type-level F1 scores for *APC*. Figures S11 to S14 – Gini scores based on true and randomly shuffled labels for *KRAS*, *PTEN*, *AR*, and *ERBB4*. Figure S15 – F1 score distribution across all genes. Figure S16 – Gini score distribution across all genes. Figures S17 to S19 – Top genes SHAP values for *ATRX*, *BRAF* and *TP53*. Figures S20 to S23 – Samples with intron variants predicted as mutant for *EGFR*, *EZH2*, *NCOR1*, and *RB1*.

#### Acknowledgements

This work would not be possible without the participation of our patients and families, the POG team, the GSC platform, and the generous support of the BC Cancer Foundation, Genome British Columbia (project B20POG), and the Terry Fox Research Institute's Marathon of Hope Cancer Centres Network (MOHCCN). We also acknowledge contributions towards equipment and infrastructure from Genome Canada and Genome BC (projects 202SEQ, 212SEQ, 262SEQ, 12002), Canada Foundation for Innovation (projects 20070, 30981, 30198, 33408 and 35444) and the BC Knowledge Development Fund. The results published here are in part based upon data generated by the following project: The Cancer Genome Atlas managed by the NCI and NHGRI (<http://cancergenome.nih.gov>).

#### Author contributions

SJM and FK designed and conceptualized the project. FK and EP acquired the relevant data. FK performed the analyses, generated the graphs, and drafted the manuscript. All authors interpreted the results and added relevant revisions to the manuscript. SJM supervised the project. All authors approved the final manuscript.

#### Funding

This research has been supported by the BC Cancer Foundation, Canada Research Chair Program funding to SJM and a Canadian Institutes of Health Research Doctoral Research Award and a University of British Columbia's 4-Year Doctoral Fellowship Award to FK.

Canadian Institutes of Health Research, University of British Columbia, BC Cancer Foundation

#### Data availability

- Expression matrices containing TPM (transcript per million) values and CNA data were obtained from the University of California Santa Cruz repository (192). SNV/INDEL data files were downloaded from GDC data portal (193), and germline mutation data file was downloaded from genomic data commons website (194). Structural variation files for TCGA study were downloaded from cBioPortal (195).
- Genomic and transcriptomic sequence datasets for the POG program are available at the European Genome-phenome Archive (EGA, <https://ega-archive.org/>) as part of the study EGAS00001001159.
- The code used in this project can be found in this GitHub Repository: [https://github.com/FaezeK/Cancer\\_Gene\\_Mutational\\_Status\\_Classifier](https://github.com/FaezeK/Cancer_Gene_Mutational_Status_Classifier).

#### Declarations

##### Ethics approval and consent to participate

TCGA data is available publicly. The POG program, registered under clinical trial number NCT02155621, was approved by the University of British Columbia – BC Cancer Research Ethics Board (H12-00137, H14-00681), and approved by the institutional review board.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 21 February 2025 Accepted: 14 July 2025

Published online: 07 August 2025

#### References

- Cieślak M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Vol. 19, *Nat Rev Genet*. Nature Publishing Group; 2018. p. 93–109.
- Massard C, Michiels S, Ferte C, Le Deley MC, Lacroix L, Hollebecque A, et al. High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: results of the MOSCATO 01 trial. *Cancer Discov*. 2017;7(6):586–95.
- Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502(7471):333–9.
- Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol*. 2016;34(2):155–63.
- Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;174(4):1034–5.
- Mendiratta G, Ke E, Aziz M, Liarakos D, Tong M, Stites EC. Cancer gene mutation frequencies for the U.S. population. *Nat Commun*. 2021;12(1):1–11. Available from: <https://doi.org/10.1038/s41467-021-26213-y>
- Pleasant E, Titmuss E, Williamson L, Kwan H, Culibrk L, Zhao EY, et al. Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes. *Nat Cancer*. 2020;1(4):452–68.
- Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45(10):1113–20.
- Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23(6):703–13.
- Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*. 2010;363(18):1693–703.
- Larkin J, Ascierto PA, Dréno B, Atkinson V, Liszkay G, Maio M, et al. Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N Engl J Med*. 2014;371(20):1867–76.
- Yang L, Lee MS, Lu H, Oh DY, Kim YJ, Park D, et al. Analyzing somatic genome rearrangements in human cancers by using whole-exome sequencing. *Am J Hum Genet*. 2016;98(5):843–56.
- Keshavarz-rahaghi F, Pleasant E, Kolisnik T, Jones SJM. A p53 transcriptional signature in primary and metastatic cancers derived using machine learning. *Front Genet*. 2022;13.
- Davis RJ, Gönen M, Margineantu DH, Handeli S, Swanger J, Hoellerbauer P, et al. Pan-cancer transcriptional signatures predictive of oncogenic mutations reveal that Fbw7 regulates cancer cell oxidative metabolism. *Proc Natl Acad Sci U S A*. 2018;115(21):5462–7.
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000. <https://doi.org/10.1038/35000501>.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530–6.
- Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98–110.
- Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol*. 2014;32(12):1202–12.
- Pleasant E, Bohm A, Williamson LM, Nelson JMT, Shen Y, Bonakdar M, et al. Whole genome and transcriptome analysis enhances precision cancer treatment options. *Ann Oncol*. 2022. <https://doi.org/10.1016/j.annonc.2022.05.522>.
- Kristensen VN, Vaske CJ, Ursini-Siegel J, Van Loo P, Nordgard SH, Sachidanandam R, et al. Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (DCIS) reveal differential vascular and interleukin signaling. *Proc Natl Acad Sci U S A*. 2012;109(8):2802–7.
- Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet*. 2003;33(1):49–54.
- Grewal JK, Tessier-Cloutier B, Jones M, Gakkhar S, Ma Y, Moore R, et al. Application of a neural network whole transcriptome-based pan-cancer method for diagnosis of primary and metastatic cancers. *JAMA Netw Open*. 2019;2(4):e192597.
- Anaya J, Reon B, Chen WM, Bekiranov S, Dutta A. A pan-cancer analysis of prognostic genes. *PeerJ*. 2016;2016(2).
- Way GP, Sanchez-Vega F, La K, Armenia J, Chatila WK, Luna A, et al. Machine Learning Detects Pan-cancer Ras Pathway Activation in The Cancer Genome Atlas. *Cell Rep*. 2018;23(1):72–80.
- Jha A, Quesnel-Vallières M, Wang D, Thomas-Tikhonenko A, Lynch KW, Barash Y. Identifying common transcriptome signatures of cancer by interpreting deep learning models. *Genome Biol*. 2022;23(1): 117.
- Zhang A, Liu C, Lin G. P53 Pathway Activate Detection based on Machine Learning. In: Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence. New York, NY, USA: ACM; 2021. p. 41–5.
- Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013;45(10):1134–40.
- Giordano G, Parceseppe P, D'Andrea MR, Coppola L, Di Raimo T, Remo A, et al. JAK/Stat5-mediated subtype-specific lymphocyte antigen 6 complex, locus G6D (LY6G6D) expression drives mismatch repair proficient colorectal cancer. *J Exp Clin Cancer Res*. 2019;38(1):28. <https://doi.org/10.1186/s13046-018-1019-5>.
- Grant A, Xicola RM, Nguyen V, Lim J, Thorne C, Salhia B, et al. Molecular drivers of tumor progression in microsatellite stable APC mutation-negative colorectal cancers. *Sci Rep*. 2021;11(1): 23507. <https://doi.org/10.1038/s41598-021-02806-x>.
- Lee SH, Koo BS, Kim JM, Huang S, Rho YS, Bae WJ, et al. Wnt/β-catenin signalling maintains self-renewal and tumorigenicity of head and neck squamous cell carcinoma stem-like cells by activating Oct4. *J Pathol*. 2014;234(1):99–107. <https://doi.org/10.1002/path.4383>.

31. Simó-Riudalbas L, Offner S, Planet E, Duc J, Abrami L, Dind S, et al. Transposon-activated *POU5F1B* promotes colorectal cancer growth and metastasis. *Nat Commun.* 2022;13(1): 4913. <https://doi.org/10.1038/s41467-022-32649-7>.
32. Schuijers J, Junker JP, Mokry M, Hatzis P, Koo BK, Sasselli V, et al. Ascl2 Acts as an R-spondin/Wnt-Responsive Switch to Control Stemness in Intestinal Crypts. *Cell Stem Cell.* 2015;16(2):158–70. <https://doi.org/10.1016/j.stem.2014.12.006>.
33. Ren Z, Wang Z, Gu D, Ma H, Zhu Y, Cai M, et al. Genome Instability and Long Noncoding RNA Reveal Biomarkers for Immunotherapy and Prognosis and Novel Competing Endogenous RNA Mechanism in Colon Adenocarcinoma. *Front Cell Dev Biol.* 2021;9. Available from: <https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/https://doi.org/10.3389/fcell.2021.740455>
34. Jin W, Wang X. PLAGL2 promotes the proliferation and migration of diffuse large B-cell lymphoma cells via Wnt/ $\beta$ -catenin pathway. *Ann Clin Lab Sci.* 2022;52(3):359–66.
35. Behrens J, Jerchow BA, Würtele M, Grimm J, Asbrand C, Wirtz R, et al. Functional interaction of an Axin homolog, Conductin, with  $\beta$ -catenin, APC, and GSK3 $\beta$ . *Science.* 1998;280(5363):596–9. <https://doi.org/10.1126/science.280.5363.596>.
36. Bond CE, McKeone DM, Kalimutho M, Bettington ML, Pearson SA, Dumenil TD, et al. RNF43 and ZNRF3 are commonly altered in serrated pathway colorectal tumorigenesis. *Oncotarget.* 2016;7(43):70589–600.
37. Ulicna L, Kimmey SC, Weber CM, Allard GM, Wang A, Bui NQ, et al. The interaction of SWI/SNF with the ribosome regulates translation and confers sensitivity to translation pathway inhibitors in cancers with complex perturbations. *Cancer Res.* 2022;82(16):2829–37. <https://doi.org/10.1158/0008-5472.CAN-21-1360>.
38. Du J, Liu Y, Sun J, Yao E, Xu J, Wu X, et al. *ARID1A* safeguards the canalization of the cell fate decision during osteoclastogenesis. *Nat Commun.* 2024;15(1): 5994. <https://doi.org/10.1038/s41467-024-50225-z>.
39. Zhang Z, Wang X, Hamdan FH, Likhobabina A, Patil S, Aperdanner L, et al. NFATc1 Is a Central Mediator of EGFR-Induced ARID1A Chromatin Dissociation During Acinar Cell Reprogramming. *Cell Mol Gastroenterol Hepatol.* 2023;15(5):1219–46. <https://www.sciencedirect.com/science/article/pii/S2352345X23000188>.
40. Dastsooz H, Cereda M, Donna D, Oliviero S. A comprehensive bioinformatics analysis of UBE2C in cancers. *Int J Mol Sci.* 2019;20(9): 2228.
41. Jiang H, Zheng S, Qian Y, Zhou Y, Dai H, Liang Y, et al. Restored UBE2C expression in islets promotes  $\beta$ -cell regeneration in mice by ubiquitinating PER1. *Cell Mol Life Sci.* 2023;80(8): 226. <https://doi.org/10.1007/s00018-023-04868-8>.
42. Saldívar JC, Hamperl S, Bocek MJ, Chung M, Bass TE, Cisneros-Soberanis F, et al. An intrinsic S/G2 checkpoint enforced by ATR. *Science.* 2018;361(6404):806–10. <https://doi.org/10.1126/science.aap9346>.
43. Feng H, Shen W. ACAA1 is a Predictive Factor of Survival and Is Correlated With T Cell Infiltration in Non-Small Cell Lung Cancer. *Front Oncol.* 2020;10.
44. Shan B, Zhao R, Zhou J, Zhang M, Qi X, Wang T, et al. AURKA Increase the Chemosensitivity of Colon Cancer Cells to Oxaliplatin by Inhibiting the TP53-Mediated DNA Damage Response Genes. *Biomed Res Int.* 2020;2020:1–11.
45. Trier I, Black EM, Joo YK, Kabeche L. ATR protects centromere identity by promoting DAXX association with PML nuclear bodies. *Cell Rep.* 2023;42(5): 112495.
46. Pressly JD, Hama T, Brien SO, Regner KR, Park F. *TRIP13*-deficient tubular epithelial cells are susceptible to apoptosis following acute kidney injury. *Sci Rep.* 2017;7(1): 43196. <https://doi.org/10.1038/srep43196>.
47. Pappula AL, Rasheed S, Mirzaei G, Petreaca RC, Bouley RA. A genome-wide profiling of glioma patients with an IDH1 mutation using the catalogue of somatic mutations in cancer database. *Cancers (Basel).* 2021;13(17): 4299.
48. Oppel F, Tao T, Shi H, Ross KN, Zimmerman MW, He S, et al. Loss of *atr* cooperates with p53-deficiency to promote the development of sarcomas and other malignancies. *PLoS Genet.* 2019;15(4): e1008039.
49. Xie Y, Tan Y, Yang C, Zhang X, Xu C, Qiao X, et al. Omics-based integrated analysis identified ATRX as a biomarker associated with glioma diagnosis and prognosis. *Cancer Biol Med.* 2019;16(4):784–96.
50. Lee JJ, Hsu YC, Huang WC, Cheng SP. Upregulation of dendrocyte-expressed seven transmembrane protein is associated with unfavorable outcomes in differentiated thyroid cancer. *Endocrine.* 2023;81(3):513–20.
51. Schulten HJ, Alotibi R, Al-Ahmadi A, Ata M, Karim S, Huwait E, et al. Effect of BRAF mutational status on expression profiles in conventional papillary thyroid carcinomas. *BMC Genomics.* 2015;16(S1): S6.
52. Kim HS, Kim DH, Kim JY, Jeoung NH, Lee IK, Bong JG, et al. Microarray analysis of papillary thyroid cancers in Korean. *Korean J Intern Med.* 2010;25(4):399–407.
53. Tanaka Y, Murata M, Shen CH, Furue M, Ito T. Nectin4: a novel therapeutic target for melanoma. *Int J Mol Sci.* 2021. <https://doi.org/10.3390/ijms22020976>.
54. Ruggiero CF, Malpicci D, Fattore L, Madonna G, Vanella V, Mallardo D, et al. ErbB3 phosphorylation as central event in adaptive resistance to targeted therapy in metastatic melanoma: early detection in CTCs during therapy and insights into regulation by autocrine neuregulin. *Cancers (Basel).* 2019. <https://doi.org/10.3390/cancers11101425>.
55. Mancikova V, Buj R, Castelblanco E, Inglada-Pérez L, Diez A, de Cubas AA, et al. DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. *Int J Cancer.* 2014;135(3):598–610.
56. Rusinek D, Swierniak M, Chmielik E, Kowal M, Kowalska M, Cyplinska R, et al. BRAFV600E-associated gene expression profile: early changes in the transcriptome, based on a transgenic mouse model of papillary thyroid carcinoma. *PLoS One.* 2015;10(12): e0143688.
57. Smalridge RC, Chindris AM, Asmann YW, Casler JD, Serie DJ, Reddi HV, et al. RNA Sequencing Identifies Multiple Fusion Transcripts, Differentially Expressed Genes, and Reduced Expression of Immune Function Genes in BRAF (V600E) Mutant vs BRAF Wild-Type Papillary Thyroid Carcinoma. *J Clin Endocrinol Metab.* 2014;99(2):E338–47.
58. Kalfert D, Ludvikova M, Pesta M, Hakala T, Dostalova L, Grundmanova H, et al. BRAF mutation, selected miRNAs and genes expression in primary papillary thyroid carcinomas and local lymph node metastases. *Pathol Res Pract.* 2024;258: 155319.
59. Borrelli N, Ugolini C, Giannini R, Antonelli A, Giordano M, Sensi E, et al. Role of gene expression profiling in defining indeterminate thyroid nodules in addition to <sc> BRAF </sc> analysis. *Cancer Cytopathol.* 2016;124(5):340–9.
60. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell.* 2014;159(3):676–90.
61. Hong S, Xie Y, Cheng Z, Li J, He W, Guo Z, et al. Distinct molecular subtypes of papillary thyroid carcinoma and gene signature with diagnostic capability. *Oncogene.* 2022;41(47):5121–32.
62. Huang F, Zhou P, Wang Z, Zhang XL, Liao FX, Hu Y, et al. Knockdown of TBRG4 suppresses proliferation, invasion and promotes apoptosis of osteosarcoma cells by downregulating TGF- $\beta$ 1 expression and PI3K/AKT signaling pathway. *Arch Biochem Biophys.* 2020;686: 108351.
63. Dubrovskaya A, Kanamoto T, Lomnytska M, Heldin CH, Volodko N, Souchelnytskyi S. TGF $\beta$ 1/Smad3 counteracts BRCA1-dependent repair of DNA damage. *Oncogene.* 2005;24(14):2289–97.
64. Yu X, Liu Y, Pan K, Sun P, Li J, Li L, et al. Breast cancer susceptibility gene 1 regulates oxidative damage via nuclear factor erythroid 2-related factor 2 in oral cancer cells. *Arch Oral Biol.* 2022;139: 105447.
65. Kang HJ, Hong Y Bin, Kim HJ, Wang A, Bae I. Bioactive food components prevent carcinogenic stress via Nrf2 activation in BRCA1 deficient breast epithelial cells. *Toxicol Lett.* 2012;209(2):154–60.
66. Uziel O, Yerushalmi R, Zuriano L, Naser S, Beery E, Nordenberg J, et al. *BRCA1/2* mutations perturb telomere biology: characterization of structural and functional abnormalities *in vitro* and *in vivo*. *Oncotarget.* 2016;7(3):2433–54.
67. Zhang Y, Xu S, Xu G, Gao Y, Li S, Zhang K, et al. Dynamic Expression of m6A Regulators During Multiple Human Tissue Development and Cancers. *Front Cell Dev Biol.* 2020;8: 629030.
68. Zirkel A, Lederer M, Stöhr N, Pazaitis N, Hüttelmaier S. IGF2BP1 promotes mesenchymal cell properties and migration of tumor-derived cells by enhancing the expression of LEF1 and SNAI2 (SLUG). *Nucleic Acids Res.* 2013;41(13):6618–36.
69. Yin N, Shi J, Wang D, Tong T, Wang M, Fan F, et al. IQGAP1 interacts with Aurora-A and enhances its stability and its role in cancer. *Biochem Biophys Res Commun.* 2012;421(1):64–9.

70. Song F, Kotoloshi R, Gajda M, Hölzer M, Grimm MO, Steinbach D. Reduced IQGAP2 promotes bladder cancer through regulation of MAPK/ERK pathway and cytokines. *Int J Mol Sci.* 2022;23(21): 13508.
71. Wang Q, Sun Z, Yang HS. Downregulation of tumor suppressor Pcd4 promotes invasion and activates both  $\beta$ -catenin/Tcf and AP-1-dependent transcription in colon carcinoma cells. *Oncogene.* 2008;27(11):1527–35.
72. Kariri Y, Toss MS, Alsaleem M, Elsharawy KA, Joseph C, Mongan NP, et al. Ubiquitin-conjugating enzyme 2C (UBE2C) is a poor prognostic biomarker in invasive breast cancer. *Breast Cancer Res Treat.* 2022;192(3):529–39.
73. Kandala S, Ramos M, Voith von Voithenberg L, Diaz-Jimenez A, Chocarro S, Keding J, et al. Chronic chromosome instability induced by Plk1 results in immune suppression in breast cancer. *Cell Rep.* 2023;42(12):113266.
74. Troiano G, Guida A, Aquino G, Botti G, Losito NS, Papagerakis S, et al. Integrative histologic and bioinformatics analysis of BIRC5/Survivin expression in oral squamous cell carcinoma. *Int J Mol Sci.* 2018;19(9): 2664.
75. Hélias-Rodzewicz Z, Lourenco N, Bakari M, Capron C, Emile JF. *CDKN2A* depletion causes aneuploidy and enhances cell proliferation in non-immortalized normal human cells. *Cancer Invest.* 2018;36(6):338–48.
76. Regneri J, Klotz B, Wilde B, Kottler VA, Hausmann M, Kneitz S, et al. Analysis of the putative tumor suppressor gene *cdkn2ab* in pigment cells and melanoma of *Xiphophorus* and medaka. *Pigment Cell Melanoma Res.* 2019;32(2):248–58.
77. Zhang D, Wang T, Zhou Y, Zhang X. Comprehensive analyses of cuproptosis-related gene *CDKN2A* on prognosis and immunologic therapy in human tumors. *Medicine.* 2023;102(14). [https://journals.lww.com/mdjournal/fulltext/2023/04070/comprehensive\\_analyses\\_of\\_cuproptosis\\_related\\_gene.19.aspx](https://journals.lww.com/mdjournal/fulltext/2023/04070/comprehensive_analyses_of_cuproptosis_related_gene.19.aspx)
78. Sur S, Steele R, Ko BCB, Zhang J, Ray RB. Long noncoding RNA ELDR promotes cell cycle progression in normal oral keratinocytes through induction of a CTCF-FOXM1-AURKA signaling axis. *J Biol Chem.* 2022. <https://doi.org/10.1016/j.jbc.2022.101895>.
79. Macúrek L, Lindqvist A, Lim D, Lampson MA, Klompmaker R, Freire R, et al. Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature.* 2008;455(7209):119–23.
80. Toyoshima-Morimoto F, Taniguchi E, Nishida E. Plk1 promotes nuclear translocation of human Cdc25C during prophase. *EMBO Rep.* 2002;3(4):341–8.
81. Hsiao JS, Germain ND, Wilderman A, Stoddard C, Wojenski LA, Villafano GJ, et al. A bipartite boundary element restricts *UBE3A* imprinting to mature neurons. *Proc Natl Acad Sci.* 2019;116(6):2181–6.
82. Liu M, Sun X, Shi S. MORC2 enhances tumor growth by promoting angiogenesis and tumor-associated macrophage recruitment via Wnt/ $\beta$ -catenin in lung cancer. *Cell Physiol Biochem.* 2018;51(4):1679–94.
83. She Q, Dong Y, Li D, An R, Zhou T, Nie X, et al. ABCB6 knockdown suppresses melanogenesis through the GSK3- $\beta$ / $\beta$ -catenin signaling axis in human melanoma and melanocyte cell lines. *J Dermatol Sci.* 2022;106(2):101–10.
84. Jalaledine N, El-Hajjar L, Dakik H, Shaito A, Saliba J, Safi R, et al. Pan-nexin1 is associated with enhanced epithelial-to-mesenchymal transition in human patient breast cancer tissues and in breast cancer cell lines. *Cancers (Basel).* 2019;11(12): 1967.
85. Xue W, Dong B, Zhao Y, Wang Y, Yang C, Xie Y, et al. Upregulation of TTYH3 promotes epithelial-to-mesenchymal transition through Wnt/ $\beta$ -catenin signaling and inhibits apoptosis in cholangiocarcinoma. *Cell Oncol.* 2021;44(6):1351–61.
86. Nakajima-Takagi Y, Oshima M, Takano J, Koide S, Itokawa N, Uemura S, et al. Polycomb repressive complex 1.1 coordinates homeostatic and emergency myelopoiesis. *Elife.* 2023. <https://doi.org/10.7554/eLife.83004>.
87. Chugh S, Meza J, Sheinin YM, Ponnusamy MP, Batra SK. Loss of N-acetyl-galactosaminyltransferase 3 in poorly differentiated pancreatic cancer: augmented aggressiveness and aberrant ErbB family glycosylation. *Br J Cancer.* 2016;114(12):1376–86.
88. Ji J, Li C, Wang J, Wang L, Huang H, Li Y, et al. Hsa\_circ\_0001756 promotes ovarian cancer progression through regulating IGF2BP2-mediated RAB5A expression and the EGFR/MAPK signaling pathway. *Cell Cycle.* 2022;21(7):685–96.
89. Sa R, Liang R, Qiu X, He Z, Liu Z, Chen L. IGF2BP2-dependent activation of ERBB2 signaling contributes to acquired resistance to tyrosine kinase inhibitor in differentiation therapy of radioiodine-refractory papillary thyroid cancer. *Cancer Lett.* 2022;527:10–23.
90. Xia L, Wang H, Xiao H, Lan B, Liu J, Yang Z. <sc>EEF1A2</sc> and <sc>ERN2</sc> could potentially discriminate metastatic status of mediastinal lymph node in lung adenocarcinomas harboring <sc>EGFR 19Del</sc> / <sc>L858R</sc> mutations. *Thorac Cancer.* 2020;11(10):2755–66.
91. Rinke J, Müller JP, Blaess MF, Chase A, Meggendorfer M, Schäfer V, et al. Molecular characterization of EZH2 mutant patients with myelodysplastic/myeloproliferative neoplasms. *Leukemia.* 2017;31(9):1936–43.
92. Liang W, Shi C, Hong W, Li P, Zhou X, Fu W, et al. Super-enhancer-driven lncRNA-DAW promotes liver cancer cell proliferation through activation of Wnt/ $\beta$ -catenin pathway. *Mol Ther Nucleic Acids.* 2021;26:1351–63. <https://doi.org/10.1016/j.omtn.2021.10.028>.
93. Mao JH, Perez-Iosada J, Wu D, DelRosario R, Tsunematsu R, Nakayama KI, et al. *Fbxw7/Cdc4* is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature.* 2004;432(7018):775–9.
94. Fujii Y, Yada M, Nishiyama M, Kamura T, Takahashi H, Tsunematsu R, et al. *Fbxw7* contributes to tumor suppression by targeting multiple proteins for ubiquitin-dependent degradation. *Cancer Sci.* 2006;97(8):729–36.
95. Zhang G, Zhu Q, Fu G, Hou J, Hu X, Cao J, et al. TRIP13 promotes the cell proliferation, migration and invasion of glioblastoma through the FBXW7/c-MYC axis. *Br J Cancer.* 2019;121(12):1069–78.
96. Li Z, Liu J, Chen T, Sun R, Liu Z, Qiu B, et al. HMGAI1-TRIP13 axis promotes stemness and epithelial mesenchymal transition of perihilar cholangiocarcinoma in a positive feedback loop dependent on c-Myc. *J Exp Clin Cancer Res.* 2021;40(1): 86.
97. Wang Z, Liu Y, Zhang P, Zhang W, Wang W, Curr K, et al. FAM83D promotes cell proliferation and motility by downregulating tumor suppressor gene FBXW7. *Oncotarget.* 2013;4(12):2476–86.
98. Jiang X, Wang Y, Guo L, Wang Y, Miao T, Ma L, et al. The FBXW7-binding sites on FAM83D are potential targets for cancer therapy. *Breast Cancer Res.* 2024;26(1): 37.
99. Bailey ML, Singh T, Mero P, Moffat J, Hieter P. Dependence of human colorectal cells lacking the FBW7 tumor suppressor on the spindle assembly checkpoint. *Genetics.* 2015;201(3):885–95.
100. Liu R, Lu Y, Li J, Yao W, Wu J, Chen X, et al. Annexin A2 combined with TTK accelerates esophageal cancer progression via the Akt/mTOR signaling pathway. *Cell Death Dis.* 2024;15(4):291.
101. Buratin A, Borin C, Tretti Parenzan C, Dal Molin A, Orsi S, Binatti A, et al. CircFBXW7 in patients with T-cell ALL: depletion sustains MYC and NOTCH activation and leukemia cell viability. *Exp Hematol Oncol.* 2023;12(1):12.
102. Lu H, Yao B, Wen X, Jia B. FBXW7 circular RNA regulates proliferation, migration and invasion of colorectal carcinoma through NEK2, mTOR, and PTEN signaling pathways in vitro and in vivo. *BMC Cancer.* 2019;19(1):918.
103. Takada M, Zhang W, Suzuki A, Kuroda TS, Yu Z, Inuzuka H, et al. FBW7 loss promotes chromosomal instability and tumorigenesis via cyclin E1/CDK2-mediated phosphorylation of CENP-A. *Cancer Res.* 2017;77(18):4881–93.
104. Jiang S, Katayama H, Wang J, Li SA, Hong Y, Radvanyi L, et al. Estrogen-induced Aurora kinase-A (AURKA) gene expression is activated by GATA-3 in estrogen receptor-positive breast cancer cells. *Horm Cancer.* 2010;1(1):11–20.
105. Jiang X, Yuan Y, Tang L, Wang J, Liu Q, Zou X, et al. Comprehensive pan-cancer analysis of the prognostic and immunological roles of the METTL3/lncRNA-SNHG1/miRNA-140-3p/UBE2C axis. *Front Cell Dev Biol.* 2021. <https://doi.org/10.3389/fcell.2021.765772>.
106. Messeha SS, Zarmouh NO, Maku H, Gendy S, Yedjou CG, Elhag R, et al. Prognostic and therapeutic implications of cell division cycle 20 homolog in breast cancer. *Cancers (Basel).* 2024. <https://doi.org/10.3390/cancers16142546>.
107. Xie X, Jiang S, Li X. Nuf2 Is a Prognostic-Related Biomarker and Correlated With Immune Infiltrates in Hepatocellular Carcinoma. *Front Oncol.* 2021;11.
108. Zheng B, Wang S, Yuan X, Zhang J, Shen Z, Ge C. NUF2 is correlated with a poor prognosis and immune infiltration in clear cell renal cell carcinoma. *BMC Urol.* 2023;23(1):82.

109. Liu LYD, Chang LY, Kuo WH, Hwa HL, Shyu MK, Chang KJ, et al. *In silico* prediction for regulation of transcription factors on their shared target genes indicates relevant clinical implications in a breast cancer population. *Cancer Inform.* 2012;11: CIN.S8470.
110. Bildik G, Gray JP, Mao W, Yang H, Ozyurt R, Orellana VR, et al. DIRAS3 induces autophagy and enhances sensitivity to anti-autophagic therapy in KRAS-driven pancreatic and ovarian carcinomas. *Autophagy.* 2024;20(3):675–91.
111. Ma H, Xie C, Chen Z, He G, Dai Z, Cai H, et al. MFG-E8 alleviates intervertebral disc degeneration by suppressing pyroptosis and extracellular matrix degradation in nucleus pulposus cells via Nrf2/TXNIP/NLRP3 axis. *Cell Death Discov.* 2022;8(1): 209.
112. Yi YW, Oh S. Comparative analysis of NRF2-responsive gene expression in AcPC-1 pancreatic cancer cell line. *Genes Genomics.* 2015;37(1):97–109.
113. Mbiandjeu SCT, Siciliano A, Mattè A, Federti E, Perduca M, Melisi D, et al. Nrf2 plays a key role in erythropoiesis during aging. *Antioxidants.* 2024;13(4): 454.
114. Sanghvi VR, Mohan P, Singh K, Cao L, Berishaj M, Wolfe AL, et al. NRF2 activation confers resistance to *elf4A* inhibitors in cancer therapy. *Cancers (Basel).* 2021. <https://doi.org/10.3390/cancers13040639>.
115. Oblinger JL, Wang J, Wetherell GD, Agarwal G, Wilson TA, Benson NR, et al. Anti-tumor effects of the *elf4A* inhibitor didesmethylrocaamide and its derivatives in human and canine osteosarcomas. *Sci Rep.* 2024;14(1): 19349.
116. Burroughs AF, Eluhu S, Whalen D, Goodwin JS, Sakwe AM, Arinze IJ. PML-nuclear bodies regulate the stability of the fusion protein *Dendra2-Nrf2* in the nucleus. *Cell Physiol Biochem.* 2018;47(2):800–16.
117. Greiner G, Witzneder N, Klein K, Tangermann S, Kodajova P, Jaeger E, et al. Tumor necrosis factor  $\alpha$  promotes clonal dominance of *KIT* D816V+ cells in mastocytosis: role of survival and impact on prognosis. *Blood.* 2024;143(11):1006–17.
118. Vivarelli S, Falzone L, Candido S, Bonavida B, Libra M. YY1 silencing induces 5-fluorouracil-resistance and *BCL2L15* downregulation in colorectal cancer cells: diagnostic and prognostic relevance. *Int J Mol Sci.* 2021;22(16): 8481.
119. Rayon-Estrada V, Harjanto D, Hamilton CE, Berchiche YA, Gantman EC, Sakmar TP, et al. Epitranscriptomic profiling across cell types reveals associations between APOBEC1-mediated RNA editing, gene expression outcomes, and cellular function. *Proc Natl Acad Sci.* 2017;114(50):13296–301.
120. Pandey R, Zhou M, Islam S, Chen B, Barker NK, Langlais P, et al. Carcinoembryonic antigen cell adhesion molecule 6 (CEACAM6) in pancreatic ductal adenocarcinoma (PDA): an integrative analysis of a novel therapeutic target. *Sci Rep.* 2019;9(1): 18347.
121. Singh A, Greninger P, Rhodes D, Koopman L, Violette S, Bardeesy N, et al. A Gene Expression Signature Associated with K-Ras Addiction Reveals Regulators of EMT and Tumor Cell Survival. *Cancer Cell.* 2009;15(6):489–500. <https://doi.org/10.1016/j.ccr.2009.03.022>.
122. Babicky ML, Harper MM, Chakedis J, Cazes A, Mose ES, Jaquish DV, et al. MST1R kinase accelerates pancreatic cancer progression via effects on both epithelial cells and macrophages. *Oncogene.* 2019;38(28):5599–611.
123. Lefebvre AM, Adam J, Nicolazzi C, Larois C, Attenot F, Falda-Buscaiot F, et al. The search for therapeutic targets in lung cancer: preclinical and human studies of carcinoembryonic antigen-related cell adhesion molecule 5 expression and its associated molecular landscape. *Lung Cancer.* 2023;184: 107356.
124. Wang X, Ye J, Gao M, Zhang D, Jiang H, Zhang H, et al. Nifuroxazide inhibits the growth of glioblastoma and promotes the infiltration of CD8 T cells to enhance antitumor immunity. *Int Immunopharmacol.* 2023;118: 109987.
125. Hu P, Huang Q, Li Z, Wu X, Ouyang Q, Chen J, et al. Silencing MAP3K1 expression through RNA interference enhances paclitaxel-induced cell cycle arrest in human breast cancer cells. *Mol Biol Rep.* 2014;41(1):19–24.
126. Adinew GM, Messeha S, Taka E, Soliman KFA. The prognostic and therapeutic implications of the chemoresistance gene *BIRC5* in triple-negative breast cancer. *Cancers (Basel).* 2022;14(21): 5180.
127. Li X, McDonnell DP. The transcription factor B-Myb is maintained in an inhibited state in target cells through its interaction with the nuclear corepressors N-CoR and SMRT. *Mol Cell Biol.* 2002;22(11):3663–73.
128. Doherty CJ, Kay SA. Circadian Surprise—It's Not All About Transcription. *Science.* 2012;338(6105):338–40. <https://doi.org/10.1126/science.1230008>.
129. Morf J, Rey G, Schneider K, Stratmann M, Fujita J, Naef F, et al. Cold-inducible RNA-binding protein modulates circadian gene expression posttranscriptionally. *Science.* 2012;338(6105):379–83. <https://doi.org/10.1126/science.1217726>.
130. Patel AV, Eaves D, Jessen WJ, Rizvi TA, Ecsedy JA, Qian MG, et al. Ras-driven transcriptome analysis identifies Aurora kinase A as a potential malignant peripheral nerve sheath tumor therapeutic target. *Clin Cancer Res.* 2012;18(18):5020–30.
131. Pucciarelli D, Angus SP, Huang B, Zhang C, Nakaoka HJ, Krishnamurthi G, et al. *Nf1*-mutant tumors undergo transcriptome and kinase remodeling after inhibition of either mTOR or MEK. *Mol Cancer Ther.* 2020;19(11):2382–95.
132. Jour G, Illa-Bochaca I, Ibrahim M, Donnelly D, Zhu K, Miera EVS de, et al. Genomic and Transcriptomic Analyses of NF1-Mutant Melanoma Identify Potential Targeted Approach for Treatment. *Journal of Investigative Dermatology.* 2023;143(3):444–455.e8. Available from: <https://www.sciencedirect.com/science/article/pii/S0022202X22017808>
133. Morris BB, Smith JP, Zhang Q, Jiang Z, Hampton OA, Churchman ML, et al. Replicative instability drives cancer progression. *Biomolecules.* 2022;12(11): 1570.
134. Liu H, Liu L, Rosen CJ. PTH and the regulation of mesenchymal cells within the bone marrow niche. *Cells.* 2024;13(5):406.
135. Cao H, Yang P, Liu J, Shao Y, Li H, Lai P, et al. MYL3 protects chondrocytes from senescence by inhibiting clathrin-mediated endocytosis and activating of Notch signaling. *Nat Commun.* 2023;14(1): 6190.
136. Tosello V, Di Martino L, Papathanassiou AE, Santa SD, Pizzi M, Mussolin L, et al. *BCAT1* is a NOTCH1 target and sustains the oncogenic function of NOTCH1. *Haematologica.* 2024.
137. Tang J, Zhong G, Zhang H, Yu B, Wei F, Luo L, et al. LncRNA DANCER upregulates PI3K/AKT signaling through activating serine phosphorylation of RXRA. *Cell Death Dis.* 2018;9(12): 1167.
138. Zhou Q, Huang W, Xiong J, Guo B, Wang X, Guo J. CDCA8 promotes bladder cancer survival by stabilizing HIF1 $\alpha$  expression under hypoxia. *Cell Death Dis.* 2023;14(10): 658.
139. Li L, Song Y, Liu Q, Liu X, Wang R, Kang C, et al. Low expression of PTEN is essential for maintenance of a malignant state in human gastric adenocarcinoma via upregulation of p-AURKA mediated by activation of AURKA. *Int J Mol Med.* 2018.
140. Choi BH, Xie S, Dai W. PTEN is a negative regulator of mitotic checkpoint complex during the cell cycle. *Exp Hematol Oncol.* 2017;6(1):19.
141. Li D, Yue Y, Feng X, Lv W, Fan Y, Sha P, et al. MicroRNA-542-3p targets Pten to inhibit the myoblasts proliferation but suppresses myogenic differentiation independent of targeted Pten. *BMC Genomics.* 2024;25(1): 325.
142. Bochkareva E, Korolev S, Lees-Miller SP, Bochkarev A. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* 2002;21(7):1855–63.
143. Wang G, Li Y, Wang P, Liang H, Cui M, Zhu M, et al. PTEN regulates RPA1 and protects DNA replication forks. *Cell Res.* 2015;25(11):1189–204.
144. Wu X, Chen H, You C, Peng Z. A potential immunotherapeutic and prognostic biomarker for multiple tumors including glioma: *SHOX2*. *Hereditas.* 2023;160(1):21.
145. Singharajkomron N, Yodsurang V, Seephan S, Kungsukool S, Petchjorm S, Maneeganjanasing N, et al. Evaluating the expression and prognostic value of genes encoding microtubule-associated proteins in lung cancer. *Int J Mol Sci.* 2022;23(23): 14724.
146. van Ree JH, Nam HJ, Jeganathan KB, Kanakkanthara A, van Deursen JM. Pten regulates spindle pole movement through Dlg1-mediated recruitment of Eg5 to centrosomes. *Nat Cell Biol.* 2016;18(7):814–21.
147. Imada EL, Sanchez DF, Dinalankara W, Vidotto T, Ebot EM, Tyekucheva S, et al. Transcriptional landscape of PTEN loss in primary prostate cancer. *BMC Cancer.* 2021;21(1): 856.
148. Wang X, Cao X, Sun R, Tang C, Tzankov A, Zhang J, et al. Clinical significance of PTEN deletion, mutation, and loss of PTEN expression in de novo diffuse large B-cell lymphoma. *Neoplasia.* 2018;20(6):574–93.

149. Gong X, Du J, Parsons SH, Merzoug FF, Webster Y, Iversen PW, et al. Aurora A kinase inhibition is synthetic lethal with loss of the *RB1* tumor suppressor gene. *Cancer Discov.* 2019;9(2):248–63.
150. Lyu J, Yang EJ, Zhang B, Wu C, Pardeshi L, Shi C, et al. Synthetic lethality of *RB1* and *aurora A* is driven by stathmin-mediated disruption of microtubule dynamics. *Nat Commun.* 2020;11(1): 5105.
151. Yang J, Li Y, Han Y, Feng Y, Zhou M, Zong C, et al. Single-cell transcriptome profiling reveals intratumoral heterogeneity and malignant progression in retinoblastoma. *Cell Death Dis.* 2021;12(12): 1100.
152. Sumi T, Hirai S, Yamaguchi M, Tanaka Y, Tada M, Niki T, et al. Trametinib downregulates survivin expression in *RB1*-positive *KRAS*-mutant lung adenocarcinoma cells. *Biochem Biophys Res Commun.* 2018;501(1):253–8.
153. Raj D, Liu T, Samadashwily G, Li F, Grossman D. Survivin repression by p53, Rb and E2F2 in normal human melanocytes. *Carcinogenesis.* 2008;29(1):194–201.
154. Zhu Y, Yang Y, Jethava Y, Frech I, Tricot GJ, Zhan F. Targeting *NEK2* induces cellular senescence in B-cell malignancies through p53-independent signaling pathways. *Blood.* 2019;134(Supplement\_1):3102–3102.
155. Suzuki K, Kokuryo T, Senga T, Yokoyama Y, Nagino M, Hamaguchi M. Novel combination treatment for colorectal cancer using *Nek2* siRNA and cisplatin. *Cancer Sci.* 2010;101(5):1163–9.
156. Gao X, Wen X, He H, Zheng L, Yang Y, Yang J, et al. Knockdown of *CDCA8* inhibits the proliferation and enhances the apoptosis of bladder cancer cells. *PeerJ.* 2020;8: e9078.
157. Date DA, Jacob CJ, Bekier ME, Stiff AC, Jackson MW, Taylor WR. *Borealin* is repressed in response to p53/Rb signaling. *Cell Biol Int.* 2007;31(12):1470–81.
158. Amato A, Schillaci T, Lentini L, Di Leonardo A. CENPA overexpression promotes genome instability in pRb-depleted human cells. *Mol Cancer.* 2009;8(1): 119.
159. Lütge A, Lu J, Hüllein J, Walther T, Sellner L, Wu B, et al. Subgroup-specific gene expression profiles and mixed epistasis in chronic lymphocytic leukemia. *Haematologica.* 2023;108(10):2664–76.
160. Dolatshad H, Pellagatti A, Liberante FG, Llorian M, Repapi E, Steeples V, et al. Cryptic splicing events in the iron transporter *ABC7* and other key target genes in *SF3B1*-mutant myelodysplastic syndromes. *Leukemia.* 2016;30(12):2322–31.
161. Clough CA, Pangallo J, Sarchi M, Ilagan JO, North K, Bergantinos R, et al. Coordinated missplicing of *TMEM14C* and *ABC7* causes ring sideroblast formation in *SF3B1*-mutant myelodysplastic syndrome. *Blood.* 2022;139(13):2038–49.
162. Visconte V, Rogers HJ, Singh J, Barnard J, Bupathi M, Traina F, et al. *SF3B1* haploinsufficiency leads to formation of ring sideroblasts in myelodysplastic syndromes. *Blood.* 2012;120(16):3173–86.
163. Obeng EA, Chappell RJ, Seiler M, Chen MC, Campagna DR, Schmidt PJ, et al. Physiologic Expression of *Sf3b1*<sup>K700E</sup> Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell.* 2016;30(3):404–17. <https://doi.org/10.1016/j.ccell.2016.08.006>.
164. Olsen AK, Coskun M, Bzorek M, Kristensen MH, Danielsen ET, Jørgensen S, et al. Regulation of APC and AXIN2 expression by intestinal tumor suppressor *CDX2* in colon cancer cells. *Carcinogenesis.* 2013;34(6):1361–9.
165. Laurent E, McCoy JW, Macina RA, Liu W, Cheng G, Robine S, et al. *Nox1* is over-expressed in human colon cancers and correlates with activating mutations in *K-Ras*. *Int J Cancer.* 2008;123(1):100–7.
166. Zhang X, Abreu JG, Yokota C, MacDonald BT, Singh S, Coburn KLA, et al. *Tiki1* is required for head formation via Wnt cleavage-oxidation and inactivation. *Cell.* 2012;149(7):1565–77. <https://doi.org/10.1016/j.cell.2012.04.039>.
167. Matsumoto A, Shimada Y, Nakano M, Oyanagi H, Tajima Y, Nakano M, et al. *RNF43* mutation is associated with aggressive tumor biology along with *BRAF* V600E mutation in right-sided colorectal cancer. *Oncol Rep.* 2020. <https://doi.org/10.3892/or.2020.7561>.
168. Garinchesa P, Sakamoto J, Welt S, Real F, Rettig W, Old L. Organ-specific expression of the colon cancer antigen A33, a cell surface target for antibody-based therapy. *Int J Oncol.* 1996;9(3):465–71.
169. Hryniuk A, Grainger S, Savory JGA, Lohnes D. *Cdx1* and *Cdx2* function as tumor suppressors. *J Biol Chem.* 2014;289(48):33343–54.
170. Zong H, Zou JQ, Huang JP, Huang ST. Potential role of long noncoding RNA RP5–881L22.5 as a novel biomarker and therapeutic target of colorectal cancer. *World J Gastrointest Oncol.* 2022;14(11):2108–21.
171. Lind KT, Chatwin HV, DeSisto J, Coleman P, Sanford B, Donson AM, et al. Novel *RAF* fusions in pediatric low-grade gliomas demonstrate *MAPK* pathway activation. *J Neuropathol Exp Neurol.* 2021;80(12):1099–107.
172. Park HJ, Baek I, Cheang G, Solomon JP, Song W. Comparison of RNA-based next-generation sequencing assays for the detection of *NTRK* gene fusions. *J Mol Diagn.* 2021;23(11):1443–51.
173. Chakraborty A, Diefenbacher ME, Mylona A, Kassel O, Behrens A. The E3 ubiquitin ligase *Trim7* mediates *c-Jun/AP-1* activation by *Ras* signalling. *Nat Commun.* 2015;6(1):6782.
174. Morafraila EC, Saiz-Ladera C, Nieto-Jiménez C, Györfy B, Nagy A, Velasco G, et al. Mapping immune correlates and surfaceome genes in *BRAF* mutated colorectal cancers. *Curr Oncol.* 2023;30(3):2569–81.
175. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a Cancer Dependency Map. *Cell.* 2017;170(3):564–576.e16.
176. Maes A, Maes K, De Raeve H, De Smedt E, Vlummens P, Szablewski V, et al. The anaphase-promoting complex/cyclosome: a new promising target in diffuse large B-cell lymphoma and mantle cell lymphoma. *Br J Cancer.* 2019;120(12):1137–46.
177. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57.
178. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37(1):1–13.
179. Jang SH, Kim AR, Park NH, Park JW, Han IS. *DRG2* regulates G2/M progression via the *Cyclin B1-Cdk1* complex. *Mol Cells.* 2016;39(9):699–704.
180. Qin T, Mullan B, Ravindran R, Messinger D, Siada R, Cummings JR, et al. *Atrx* loss in glioma results in dysregulation of cell-cycle phase transition and *ATM* inhibitor radio-sensitization. *Cell Rep.* 2022. <https://doi.org/10.1016/j.celrep.2021.110216>.
181. Richardson TE, Kumar A, Xing C, Hatanpaa KJ, Walker JM. Overcoming the odds: toward a molecular profile of long-term survival in glioblastoma. *J Neuropathol Exp Neurol.* 2020;79(10):1031–7.
182. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and molecular epidemiology of adult diffuse glioma. *Nat Rev Neurol.* 2019;15(7):405–17.
183. Yang Y, Wang J, Wan J, Cheng Q, Cheng Z, Zhou X, et al. *PTEN* deficiency induces an extrahepatic cholangitis-cholangiocarcinoma continuum via *aurora kinase A* in mice. *J Hepatol.* 2024;81(1):120–34.
184. Harper JW, Bennett EJ. Proteome complexity and the forces that drive proteome imbalance. *Nature.* 2016;537(7620):328–38.
185. Nusinow DP, Szpyt J, Ghandi M, Rose CM, McDonald ER, Kalocsay M, et al. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell.* 2020;180(2):387–402.e16.
186. Gerstung M, Jolly C, Leshchiner I, Drento SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. *Nature.* 2020;578(7793):122–8.
187. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
188. Khoshgoftaar TM, Dittman DJ, Wald R, Awada W. A Review of Ensemble Classification for DNA Microarrays Data. In: 2013 IEEE 25th International Conference on Tools with Artificial Intelligence. IEEE; 2013. p. 381–9.
189. Maurya NS, Kushwaha S, Chawade A, Mani A. Transcriptome profiling by combined machine learning and statistical R analysis identifies *TMEM236* as a potential novel diagnostic biomarker for colorectal cancer. *Sci Rep.* 2021;11(1): 14304.
190. UK Biobank. 2024. Available from: <https://www.ukbiobank.ac.uk/>
191. National Institute of Health. Cancer Moonshot. Available from: <https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative>
192. TCGA Pan-Cancer (PANCAN). 2021. Available from: <https://xenabrowser.net/datapages/>
193. GDC Data Portal. 2022. Available from: <https://portal.gdc.cancer.gov/>
194. Genomic Data Commons. 2023. Available from: <https://gdc.cancer.gov/about-data/publications/PanCanAtlas-Germline-AWG>
195. cBioPortal. 2023. Available from: <https://www.cbioportal.org/>
196. EMBL-EBI. Ensembl Variation - Calculated variant consequences. 2023. Available from: [https://useast.ensembl.org/info/genome/variation/prediction/predicted\\_data.html](https://useast.ensembl.org/info/genome/variation/prediction/predicted_data.html)

197. Index of /pub/gdp. 2023. Available from: <https://ftp.ncbi.nlm.nih.gov/pub/gdp/>
198. Ensembl. Available from: <https://www.ensembl.org/biomart/martview/8907a97ca0c1d33ecef4e3b8096ea8c>
199. Lundberg S. SHAP. 2018. Available from: <https://shap.readthedocs.io/en/latest/index.html>
200. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24–6.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.