Revealing cancer driver genes through integrative transcriptomic and epigenomic analyses with Moonlight

Mona Nourbakhsh[1], Yuanning Zheng[2], Humaira Noor[2], Matteo Tiberti[3], Olivier Gevaert[2,4], Elena Papaleo*[1,3]

[1]Cancer Systems Biology, Section for Bioinformatics, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark
[2]Stanford Center for Biomedical Informatics Research, Department of Medicine, Stanford University, Stanford, CA, USA
[3]Cancer Structural Biology, Danish Cancer Institute, Copenhagen, Denmark
[4]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA

*Corresponding author: Elena Papaleo, elpap@dtu.dk, elenap@cancer.dk

Running title
Methylation-driven discovery of cancer drivers

Keywords
Cancer, driver, gene expression, methylation, oncogene, tumor suppressor

Abstract
Cancer involves dynamic changes caused by (epi)genetic alterations such as mutations or abnormal DNA methylation patterns which occur in cancer driver genes. These driver genes are divided into oncogenes and tumor suppressors depending on their function and mechanism of action. Discovering driver genes in different cancer (sub)types is important not only for increasing current understanding of carcinogenesis but also from prognostic and therapeutic perspectives. We have previously developed a framework called Moonlight which uses a systems biology multi-omics approach for prediction of driver genes. Here, we present further updates to Moonlight by incorporating a DNA methylation layer which provides epigenetic evidence for deregulated expression profiles of driver genes. To this end, we present a novel functionality called Gene Methylation Analysis (GMA) which investigates abnormal DNA methylation patterns to predict driver genes. This is achieved by integrating the tool EpiMix which is designed to detect such aberrant DNA methylation patterns in a cohort of patients and further couples these patterns with changes in gene expression. To showcase GMA, we applied it to three cancer (sub)types (basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma) where we discovered 33, 190, and 263 epigenetically driven genes, respectively. A subset of these driver genes had prognostic effects with expression levels significantly affecting survival of the patients. Moreover, a subset of the driver genes demonstrated therapeutic potential as drug targets. This study provides a framework for exploring the driving forces behind cancer and provides novel insights into the landscape of three cancer sub(types) by integrating gene expression

and methylation data. Moonlight2R is available on GitHub (https://github.com/ELELAB/Moonlight2R) and BioCondcutor (https://bioconductor.org/packages/release/bioc/html/Moonlight2R.html) and the associated case studies presented in this study are available on GitHub (https://github.com/ELELAB/Moonlight2_GMA_case_studies) and OSF (https://osf.io/j4n8q/).

1 Introduction

Cancer is a complex and heterogeneous disease and a leading cause of death globally [1]. This widespread disease is categorized into multiple (sub)types and is characterized by stepwise accumulation of (epi)genetic alterations in cancer driver genes [2]. These driver genes are classified according to their function where oncogenes (OCGs) are activated by gain-of-function mechanisms and tumor suppressor genes (TSGs) are inactivated by loss-of-function mechanisms [3]. Recently, dual role genes have also emerged which show context-dependent behavior and can thus act as both OCGs and TSGs in different biological contexts [4,5]. Driver genes participate in several cellular pathways conceptualized in the Hallmarks of Cancer, a collection of functional capabilities that cells gain during their transition from a normal state to a tumor cell [6–8]. Distinct driver genes can initiate cancer development in different cancer types and even within subtypes of cancers originating from the same tissue. Thus, context-specific discovery of driver genes in light of the cancer hallmarks is essential. Numerous tools have been developed for prediction of driver genes based on varying computational methods which we recently reviewed [9]. Prediction of driver genes is essential for increasing current knowledge of cancer development and for analyzing and interpreting today's vast amount of data in relation to the cancer phenotypes. This knowledge can be a step towards reversing these phenotypes, discovering novel drug targets, facilitating new treatment strategies, and designing precision medicine strategies [10–13]. We have also contributed to this field with Moonlight which uses a multi-omics systems biology approach for prediction of driver genes [14,15].

The accumulated (epi)genetic alterations in driver genes include mutations, copy number variations, aberrant methylation levels, and histone modifications [3,16]. While abnormal methylation patterns are recognized as cancer-causing mechanisms, they have been described to a lesser extent compared to mutations [9]. Hypomethylation and hypermethylation, respectively representing loss and gain of methylation compared to normal conditions, have been described as activating and inactivating mechanisms of OCGs and TSGs, respectively [17–19]. For instance, Søes et al. found promoter hypomethylation and increased expression of putative OCG *ELMO3* to be associated with development of non-small cell lung cancer [20]. Other studies have also identified driver genes with aberrant methylation profiles [21–28]. For example, one study identified 118 differentially expressed methylation-driven genes associated with lung adenocarcinoma of which five genes (*CCDC181*, *PLAU*, *S1PR1*, *ELF3*, *KLHDC9*) were used to construct a prognostic risk model which could divide patients with lung

adenocarcinoma into high- and low-risk groups [28]. Additionally, promoter hypermethylation of *BRCA1*, a well-known TSG, has been reported in breast cancers [29,30].

In this study, we present novel functionalities to Moonlight, expanding upon features presented in our previous work [15]. Specifically, we incorporate methylation evidence to Moonlight's predicted driver genes as a source of epigenetic explanation of the deregulated expression of these genes. Information about methylation state is provided by EpiMix, an integrative tool for detecting aberrant DNA methylation patterns connected with expression changes in patient cohorts [31]. To validate and showcase this new feature, we apply it to three cancer (sub)types (basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma) and discover driver genes in the context of cell proliferation and apoptosis, two well-established cancer hallmarks, and explore the prognostic and therapeutic potentials of the predicted driver genes. We apply our new method on data from The Cancer Genome Atlas (TCGA), available at the Genomic Data Commons [32,33]. The wealth of available data is fundamental in this context towards identifying crucial molecular drivers in carcinogenesis.

2 Methods

2.1 Design and implementation of new functionalities in Moonlight

Here, we present new functionalities to Moonlight, a framework that we have previously developed for driver gene prediction [14,15]. In brief, Moonlight requires a set of differentially expressed genes (DEGs) as input and is built up on two layers: a primary layer discovering putative driver genes, termed oncogenic mediators, that uses gene expression changes and information about cancer-related biological processes; and a secondary layer that couples mechanistic evidence to the oncogenic mediators by investigating (epi)genetic alterations in the oncogenic mediators. From this secondary layer, the critical driver genes are predicted among the oncogenic mediators. We recently presented an update to Moonlight implementing a secondary mutational layer that explores driver mutations in the oncogenic mediators for prediction of driver genes [15]. In this current contribution, we add a new functionality to Moonlight, implementing a secondary methylation layer to predict driver genes, according to changes in methylation levels between normal and tumor samples. Depending on the user's data availability and aim, these two secondary layers do not necessarily both need to be applied following the primary layer.

This new functionality is termed Gene Methylation Analysis (GMA) and should be applied following the Pattern Recognition Analysis (PRA) function which predicts the oncogenic mediators in the primary layer (**Figure 1A**). The biological foundation for GMA lies within the observed roles of DNA methylation in both physiological and cancer states. Under normal healthy conditions, DNA methylation serves an essential regulatory role in cells by regulating expression of genes [34]. However, in cancer, DNA methylation processes are altered, where hypo- and hypermethylation can activate

and inactivate OCGs and TSGs, respectively, leading to overexpression of OCGs and silencing of TSGs [17–19] (**Figure 1B**).

GMA predicts methylation-driven driver genes by using EpiMix [31]. EpiMix models DNA methylation in patient cohorts and predicts differential methylation associated with gene expression and further allows for DNA methylation analysis of non-coding regulatory regions [31], therefore being perfectly suitable to integrate with Moonlight's primary layer. Moreover, EpiMix is easily available as a R BioConductor package, which allows for easy integration with Moonlight. A key result of EpiMix is a table which includes functional CpG-gene pairs containing differentially methylated CpG sites whose DNA methylation state is associated with the expression of the corresponding genes they map to. Moreover, the methylation state (e.g. hypo- or hypermethylated) of each CpG site is reported. This table is integrated with the main output table from Moonlight's primary layer, specifically the output from PRA, which provides a list of oncogenic mediators and their putative driver role (e.g. putative TSG or OCG) (**Figure 1C**). This integration step involves the following: for each oncogenic mediator, the number of associated CpG sites is summarized. EpiMix's predictions of methylation state and Moonlight's predictions of driver gene role are then compared and used to assess whether the gene's methylation status supports the putative role (OCG or TSG) of the oncogenic mediator. These comparisons are subsequently used to define the driver genes (see section 2.2).

As input, GMA requires i) a gene expression matrix with genes in rows and tumor and normal samples in columns, ii) a methylation matrix with CpG sites in rows and tumor and normal samples in columns which should be the same samples as in the expression data, iii) output of PRA from Moonlight's primary layer, i.e. the predicted oncogenic mediators and their putative driver role, and finally, iv) output of a differential expression analysis (DEA) which includes information about the DEGs. In return, GMA outputs the following: i) list of predicted driver genes categorized into TSGs and OCGs, ii) a summary of the oncogenic mediators which includes the number of associated CpG sites and evidence label, iii) a summary of various annotations found to all DEGs input to Moonlight on the gene and methylation level, and iv) raw EpiMix results corresponding to applying EpiMix on the input data independent of the GMA function.

2.2 Evidence integration for defining driver genes

As mentioned in section 2.1, the driver genes are defined by comparing EpiMix's predictions of methylation state and Moonlight's predictions of driver role. This is based on the assumption with which methylation changes can activate or inactivate driver genes. Suppose EpiMix predicts a gene to contain one or more hypermethylated CpG sites whereas PRA predicts this gene to be an OCG. This indicates conflicting evidence between the mechanism of (in)activation and putative driver role between the two tools. Consequently, this gene is labeled with "conflicting" evidence. Contrary, if EpiMix for instance predicts a gene to be associated with hypermethylated CpG site(s) and Moonlight predicts this gene as a TSG, this gene is labeled with an "agreement" evidence to signify correspondence between EpiMix and Moonlight. Those oncogenic

mediators labeled with an “agreement” evidence are retained as the final set of predicted driver genes (**Figure 1D**).

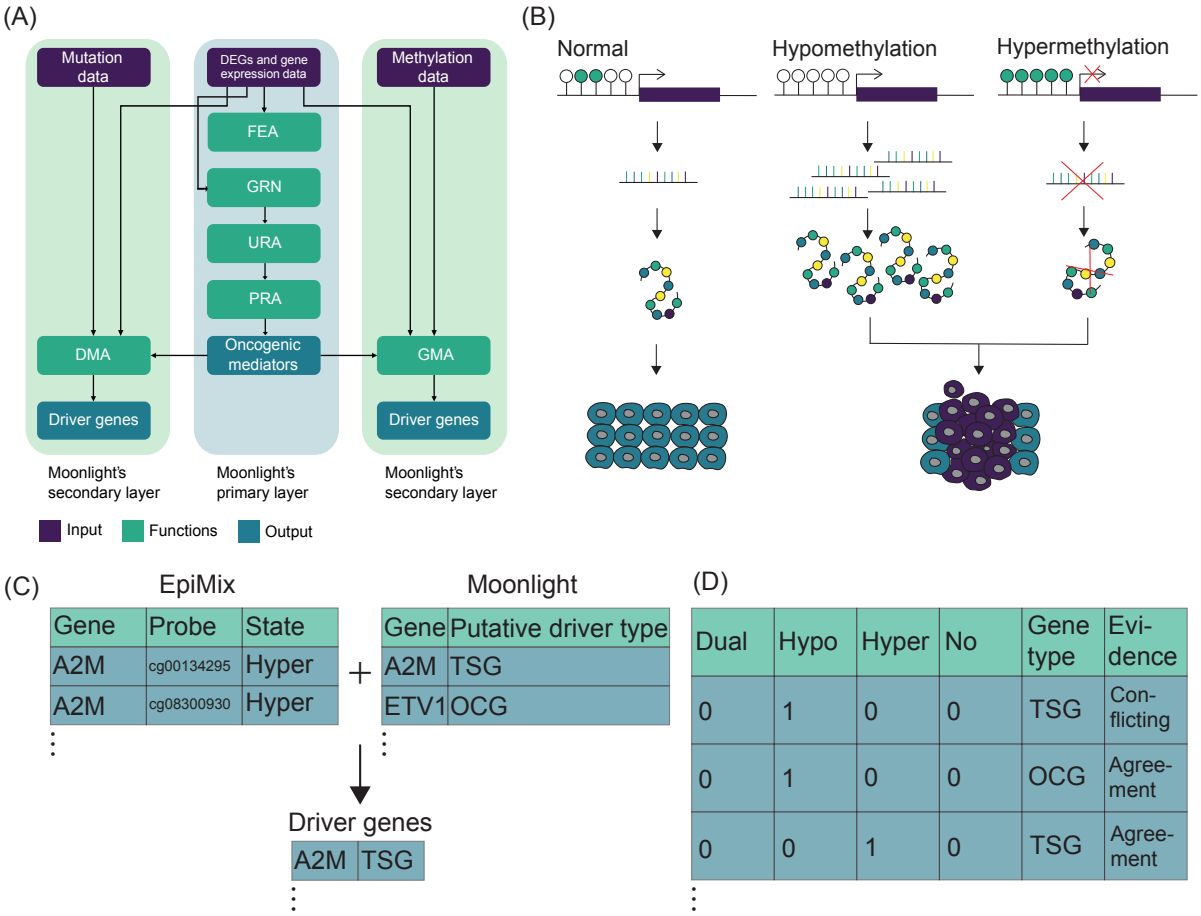

Figure 1. Overview of the Moonlight framework with new methylation functionality. **(A)** Moonlight consists of a primary layer requiring differentially expressed genes and gene expression data as input. The primary layer predicts oncogenic mediators through a series of functions called functional enrichment analysis (FEA), gene regulatory network analysis (GRN), upstream regulator analysis (URA), and pattern recognition analysis (PRA). Moonlight’s secondary mutation layer requires mutation data as input and is carried out via the driver mutation analysis (DMA) function and similarly, Moonlight’s secondary methylation layer implemented in the gene methylation analysis (GMA) function requires methylation data as input. The secondary layer results in the final prediction of driver genes. **(B)** DNA methylation is a mechanism occurring under physiological conditions in cells which functions to regulate gene expression. However, in cancer, the DNA methylation process is altered. A loss of methylation called hypomethylation can occur which can lead to increased expression of a gene and thus an increased amount of the resulting protein. In contrast, gain of methylation called hypermethylation can also occur which can silence gene expression and lead to decreased protein expression. These two mechanisms can finally lead to cancer. Hypo- and hypermethylation can activate and inactivate oncogenes and tumor suppressors, respectively, the biological principle that GMA is built on. **(C)** The outputs of EpiMix and Moonlight are integrated to predict driver genes. EpiMix outputs a table of CpG-gene pairs containing differentially methylated CpG sites whose DNA methylation state is associated with gene expression. Moonlight outputs a list of oncogenic mediators and their putative driver role as tumor suppressors or oncogenes. **(D)** Driver genes are defined in GMA by comparing EpiMix’s predictions of methylation state and Moonlight’s predictions of driver role in “evidence” categories.

2.3 Visualization functions of genes and methylation states

Accompanying the new GMA function, we have also created three functions for visualizing genes and methylation states consistent with our previous effort [15]. The first visualization function is called plotGMA which visualizes the number of differentially methylated hypo-, hyper- or dual-methylated CpG sites. The second visualization function, plotMoonlightMet, visualizes the effect of genes on biological processes estimated in Moonlight's primary layer. The genes to be visualized can be e.g. oncogenic mediators, driver genes or a user-supplied gene list containing specific genes of interest. Finally, the third function, plotMetExp, calls a visualization function from EpiMix, EpiMix_PlotModel, to display gene expression and methylation levels of a specific gene and CpG site [31].

2.4 Application of new functionality to three cancer (sub)types

Following implementation of the new functionality to Moonlight, we applied it to basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma data from The Cancer Genome Atlas (TCGA) to discover methylation-driven driver genes. We obtained gene expression, methylation, and clinical data from TCGA using the R Bioconductor package TCGAbiolinks [35–37]. For basal-like breast cancer, we retrieved this particular subtype using the PanCancerAltas_subtype function in TCGAbiolinks. We processed the gene expression data as previously described using TCGAbiolinks [14,15,35–37] and the methylation data using the Preprocess_DNAMethylation function from EpiMix as previously done [31,38]. We kept only patients who had both expression and methylation data resulting in 144, 474, and 551 patients for basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma, respectively. Subsequently, we performed DEA between each of the cancer (sub)types and corresponding normal samples using limma-voom including tissue source site (TSS) as a covariate in the linear modelling to account for TSS as a batch factor as previously outlined [14,15]. Finally, we applied the Moonlight pipeline including the new GMA functionality and predicted driver genes using apoptosis and cell proliferation as the underlying biological foundation given the recognized role of these two processes in carcinogenesis [14,15].

2.5 Enrichment analysis of predicted driver genes

We performed enrichment analyses of predicted driver genes using the enrichR R package [39–41] and the "MSigDB Hallmark 2020" database. We used the enrichr function to perform enrichment analyses on the predicted driver gene sets of each cancer (sub)type and selected significantly enriched terms as those with an adjusted p-value < 0.05.

2.6 Survival analysis of predicted driver genes

We conducted survival analysis of predicted driver genes in the three cancer (sub)types using Cox proportional hazards regression and Kaplan-Meier survival analysis implemented in the survival [42,43], survminer [44], and survMisc [45] R packages. We used the last follow-up or death time and vital status (dead or alive) of the patients as the survival data. First, we tested the proportional hazards assumption using the cox.zph function and proceeded with genes satisfying this assumption. For these genes, we next fit a univariate Cox regression model using the coxph function with survival data as the response variable and expression values of the respective gene as a continuous explanatory variable. Following FDR correction of resulting *p*-values, genes whose expression had a significant effect on survival were retained (FDR < 0.05) and subject to subsequent multivariate Cox regression analysis. Here, tumor stage, age of patients, and sex of patients were included as additional explanatory variables to ensure that the effect of the gene on survival was in fact due to its expression and not because of confounding factors such as stage, age, and sex. Sex of patients were not included for basal-like breast cancer as all patients were female. Genes whose expression showed a significant effect on survival at the multivariate level (*p*-value < 0.05) were considered prognostic. Moreover, we performed Kaplan-Meier survival analysis on these prognostic genes to also evaluate differences in survival between two discrete expression groups. Patients with expression values above and below the median expression level of the respective gene were divided into a high and low expression group, respectively. We fit survival curves using the discrete expression group as the explanatory variable and evaluated the significance of difference in survival between the two groups by means of a log-rank test as *p*-value < 0.05.

2.7 Exploration of predicted driver genes as drug targets

To investigate driver gene-drug interactions, we applied the R package rDGIdb [46,47] to query the Drug-Gene Interaction Database (DGIdb) using only cancer-specific data sources: CGI, CIViC, COSMIC, CancerCommons, ClearityFoundationBiomarkers, ClearityFoundationClinicalTrial, DoCM, JAX-CKB, MyCancerGenome, MyCancerGenomeClinicalTrial, NCI, OncoKB, and TALC [48].

3 Results

3.1 Case study: Prediction of driver genes with differential methylation in basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma using Moonlight

To showcase the new functionality in Moonlight and predict driver genes driven by methylation changes, we applied Moonlight on three cancer (sub)types: basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma. First, we performed DEA

between each of these cancer tissues and corresponding normal samples as this is the input to Moonlight's primary layer. We identified 4292, 4468, and 2972 DEGs between each of the cancer (sub)types (basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma) and corresponding normal samples, respectively (**Table 1**). Moonlight's primary layer predicted 159, 1228, and 1598 oncogenic mediators in these three cancer (sub)types, respectively (**Table 1**).

Table 1. Number of predicted DEGs, oncogenic mediators, and driver genes in three cancer (sub)types: basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma. The oncogenic mediators and driver genes predicted by Moonlight's primary and secondary methylation layer, respectively, are divided into (putative) TSGs and OCGs.

Cancer (sub)type	DEGs	Oncogenic mediators [putative TSGs/putative OCGs]	Driver genes [TSGs/OCGs]
Basal-like breast cancer	4292	159 [125/34]	33 [32/1]
Lung adenocarcinoma	4468	1228 [521/707]	190 [110/80]
Thyroid carcinoma	2972	1598 [118/1480]	263 [5/258]

Abbreviations: DEGs, differentially expressed genes; OCGs, oncogenes; TSGs, tumor suppressor genes.

Additionally, EpiMix alone identified 9483, 10018, and 6142 functional gene-CpG pairs in these three cancer (sub)types, respectively. These functional gene-CpG pairs represent differentially methylated CpG sites whose DNA methylation state is associated with the expression of the corresponding genes they map to. The number of hits discovered individually from EpiMix and Moonlight's primary layer indicate a substantial volume of significant associations. Consequently, integrating the results from EpiMix with the oncogenic mediators identified in Moonlight's primary layer, as implemented in GMA presented here, can narrow down the most critical findings and yield the benefits of both approaches. From GMA, we found that those oncogenic mediators in basal-like breast cancer that are associated with differentially methylated CpGs include 38 hypomethylated CpGs, 165 hypermethylated CpGs, and 22 methylated CpGs with a dual status, meaning the CpG site was found hypomethylated in cancer tissues from some patients, while hypermethylated in other patients. Similarly, oncogenic mediators in lung adenocarcinoma that are associated with differentially methylated CpGs include in total 218 hypomethylated CpGs, 625 hypermethylated CpGs, and 48 dual-methylated CpGs. Finally, oncogenic mediators in thyroid carcinoma associated with differentially methylated CpGs contain in total 945 hypomethylated CpGs, 305 hypermethylated CpGs, and 230 dual-methylated CpGs (**Figure 2A**).

Across all three cancer (sub)types, the number of differentially methylated CpG sites mapped to the oncogenic mediators ranges between 0 and 28. The classifications of methylation status in the oncogenic mediators in basal-like breast cancer are shown in **Figure 2B**, generated with the plotGMA function. Next, we compared Moonlight's oncogenic mediators with EpiMix' functional genes. For this, we included only those functional genes that contained the same methylation state in all of its associated CpGs and moreover, the dual states were excluded. In basal-like breast cancer, this comparison revealed 109 oncogenic mediators not associated with differentially methylated CpGs, 2754 functional genes not predicted as oncogenic mediators, 17 oncogenic mediators with a "conflicting" evidence label, and 33 oncogenic mediators with an "agreement" evidence label (**Figure 2C**). Consequently, these 33 oncogenic mediators are retained as the final set of driver genes divided into 32 TSGs and 1 OCG (**Table 1**). Next, we visualized the effect of these predicted driver genes in basal-like breast cancer on two well-known cancer hallmarks, apoptosis and proliferation of cells, using the function plotMoonlightMet. These effects define the basis upon which the oncogenic mediators are predicted from the PRA step in Moonlight's primary layer, demonstrating that the predicted OCGs have a positive effect on proliferation of cells and a negative effect on apoptosis and vice versa for the predicted TSGs (**Figure 2D**). Similar overviews for lung adenocarcinoma and thyroid carcinoma are shown in **Supplementary Figure S1**, which resulted in a final prediction of 190 driver genes divided into 110 TSGs and 80 OCGs in lung adenocarcinoma and 263 driver genes categorized into 5 TSGs and 258 OCGs in thyroid carcinoma (**Table 1**). We did not discover any dual role genes across the three cancer (sub)types, i.e. genes predicted as OCGs in one of the three cancer (sub)types and as TSGs in another cancer (sub)type and vice versa.

Next, we examined what difference the new functionality, GMA, in Moonlight's secondary layer makes with respect to Moonlight's primary layer. For this, we compared the predicted driver genes with the predicted oncogenic mediators in each cancer (sub)type. We quantified these comparisons in terms of overlaps with genes reported in the COSMIC Cancer Gene Census (CGC) [49]. Specifically, we computed the precision as (TP/(TP+FP))*100 and sensitivity as (TP/(TP+FN))*100. We defined the true positives (TP) as the overlap between the gene set (either the driver genes or the oncogenic mediators) and the CGC, whereas the false positives (FP) are those genes found in the gene set but are not included in CGC. In contrast, the false negatives (FN) comprise those genes reported in CGC but are not predicted in our gene set. For all three cancer (sub)types, we found that GMA had a greater precision and lower sensitivity compared to using only Moonlight's primary layer (**Figure 2E**). A higher precision of GMA is desirable as it indicates that the predicted driver gene sets have a higher fraction of genes from the CGC compared to the oncogenic mediator sets. On the other hand, the higher sensitivity of using only Moonlight's primary layer compared to also using GMA might be attributed to the larger numbers of oncogenic mediators. A larger number of oncogenic mediators results in a larger overlap between the CGC and the oncogenic mediators, thereby lowering the number of FNs and increasing the sensitivity. In this case, prioritizing higher precision over sensitivity is preferable since

our aim is to find the most crucial driver genes among the oncogenic mediators. Thus, a higher precision indicates a greater proportion of TPs, corresponding with our objective.

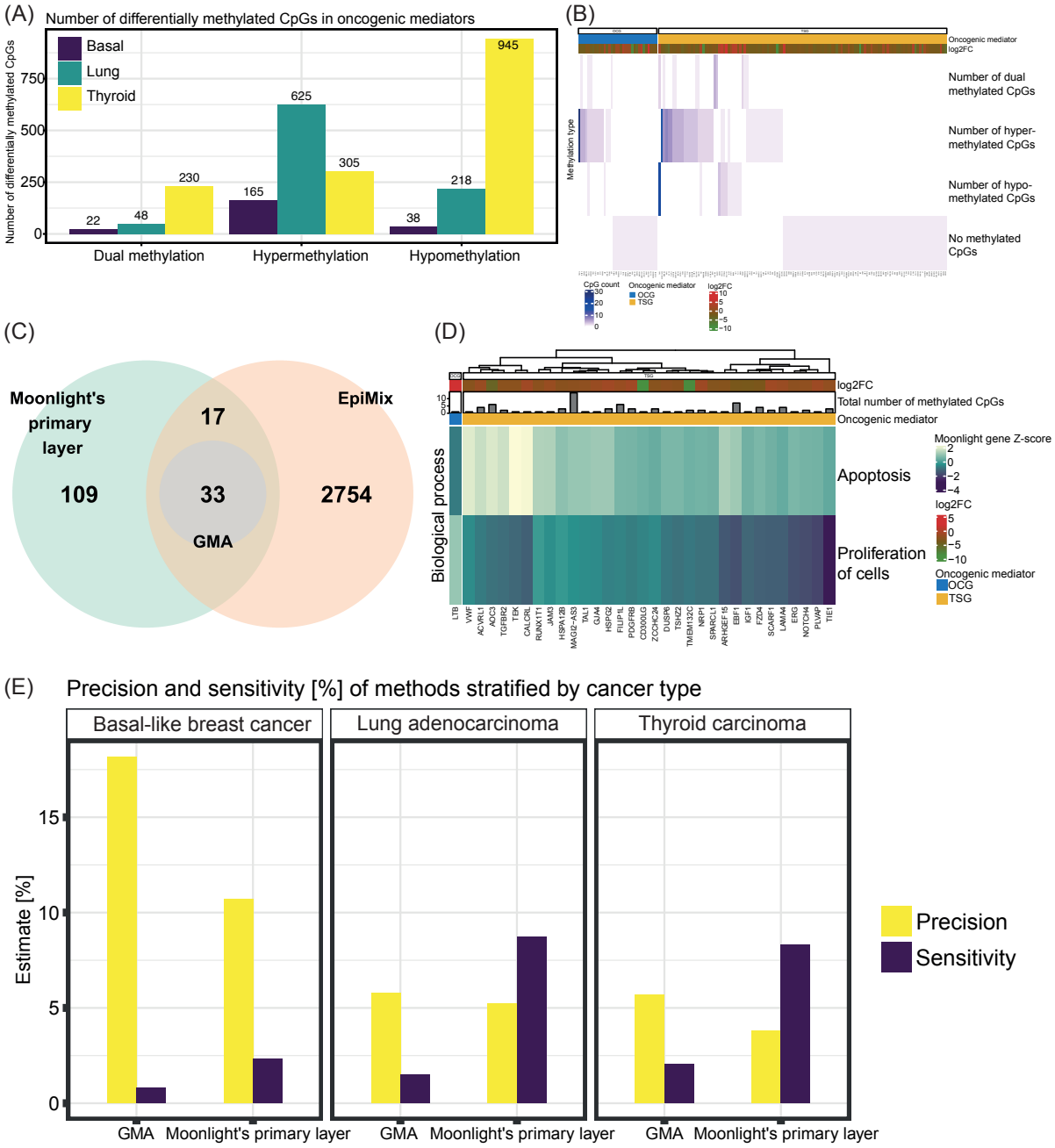

Figure 2. Integration of Moonlight and EpiMix for prediction of cancer driver genes. **(A)** Number of differentially methylated CpGs as found from EpiMix in oncogenic mediators predicted from Moonlight's primary layer. The differentially methylated CpGs are categorized into methylation status and stratified by cancer (sub)type. **(B)** Heatmap showing number of differentially methylated CpGs and classifications of methylation status in the oncogenic mediators in basal-like breast cancer. The heatmap was generated using the plotGMA function. **(C)** Venn diagram comparing oncogenic mediators predicted from Moonlight's primary layer with functional genes predicted from EpiMix in basal-like breast cancer. The functional genes are genes containing differentially methylated CpG pairs whose DNA methylation state is associated with expression of the gene. Only those functional genes that contained the same methylation state in all of its associated CpGs were included in this comparison, and moreover, the dual

methylation states were excluded. **(D)** Heatmap showing the effect of the predicted driver genes in basal-like breast cancer on apoptosis and proliferation of cells. This heatmap was generated using the function plotMoonlightMet. These effects define the basis upon which the oncogenic mediators are predicted from the PRA step in Moonlight's primary layer. **(E)** Comparison between the predicted driver genes with the predicted oncogenic mediators in all three cancer (sub)types where the driver genes were predicted with the new functionality GMA in Moonlight's secondary layer, and the oncogenic mediators were predicted with Moonlight's primary layer. The comparisons were quantified in terms of overlaps with genes reported in the COSMIC Cancer Gene Census (CGC) by computing the precision and sensitivity. The precision was calculated as (TP/(TP+FP))*100 and sensitivity as (TP/(TP+FN))*100. The true positives (TP) are the overlap between the gene set (either the driver genes or the oncogenic mediators) and the CGC. The false positives (FP) are those genes found in the gene set but are not included in CGC. The false negatives (FN) comprise those genes reported in CGC but are not predicted in our gene set.

Next, we also evaluated the significance of association between the gene sets and the CGC using a Fisher's exact test (**Table 2**). We only found the oncogenic mediator and driver gene sets from basal-like breast cancer to have a significant association with genes in the CGC (*p*-value = 0.000392 for the oncogenic mediators predicted using Moonlight's primary layer and *p*-value = 0.00228 for the driver genes predicted using GMA). However, in all three cancer (sub)types, we found the driver genes to have a higher odds ratio than the oncogenic mediators, demonstrating a greater association between the driver gene sets and the CGC compared to the oncogenic mediators (**Table 2**).

While these results together demonstrate the added value of GMA, it is worth highlighting certain limitations associated with these quantitative statistical measures. Notably, the driver genes reported in CGC are mainly based on mutation evidence, however, in this study, we have used abnormal DNA methylation levels as evidence for deregulated expression of the driver genes. Hence, these methylation patterns may not be fully captured in the CGC, challenging our comparison with the CGC. However, to date, no golden standard of cancer drivers exists, and the CGC stands as the most robust and comprehensive resource available. Thus, it serves as the main reference point that the majority of studies use to evaluate their predicted driver genes and method [50–61]. To our knowledge, a similar well-curated resource of cancer driver genes driven by methylation changes does not exist. Moreover, performing cancer type-specific comparisons would be more desirable. While the CGC reports which cancer types the driver genes are associated with, these annotations are limited in scope. Therefore, while ideal, performing such cancer type-specific comparisons do not contain enough power. Finally, the quantitative statistical measures are not taking into account that some of our predicted driver genes may be novel. Consequently, some FPs may in fact be TPs but are not included in CGC, and some FNs may not necessarily be FNs; rather, they may not represent drivers in the specific cancer (sub)type.

Table 2. Significance of association between Moonlight's gene sets and genes from the Cancer Gene Census (CGC) evaluated using Fisher's exact test in three cancer (sub)types: basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma. The gene sets from Moonlight were found using Moonlight's primary layer and Moonlight's secondary layer through the Gene Methylation Analysis (GMA) functionality. *p*-values and odds ratios from Fisher's exact test are included. **p*-value < 0.05.

Cancer (sub)type	Method	*p*-value	Odds ratio
Basal-like breast cancer	Moonlight's primary layer	0.000392*	2.77
	GMA	0.00228*	5.09
Lung adenocarcinoma	Moonlight's primary layer	0.0764	1.28
	GMA	0.272	1.40
Thyroid carcinoma	Moonlight's primary layer	0.472	0.895
	GMA	0.215	1.39

Abbreviations: CGC, Cancer Gene Census; GMA, Gene Methylation Analysis.

To investigate biological roles of the predicted driver genes, we performed enrichment analyses (**Figure 3**). The predicted driver genes are involved in various signaling pathways such as KRAS signaling in basal-like breast cancer and thyroid carcinoma, mTORC1 signaling in lung adenocarcinoma, and TNF−alpha signaling via NF−kB and p53 pathway in thyroid carcinoma. Previously, TP53 and TNF signaling have been associated with the onset of cancer among epigenetically modified pathways [62]. Furthermore, IL−6/JAK/STAT3 signaling was significantly enriched among the predicted driver genes in basal-like breast cancer (**Figure 3**). Basal-like breast cancers overexpress Interleukin 6 (IL-6), a pro-inflammatory cytokine, and it has been reported that p53 absence triggers an IL-6 dependent epigenetic reprogramming driving breast cancer cells towards a basal-like/stem cell-like gene expression profile [63]. Additionally, epithelial-mesenchymal transition (EMT) is another recurring enriched term, observed in both lung adenocarcinoma and thyroid carcinoma. Epigenetic regulation of EMT has previously been described, and DNA methylation and demethylation plays a key role in this regulation [64–67].

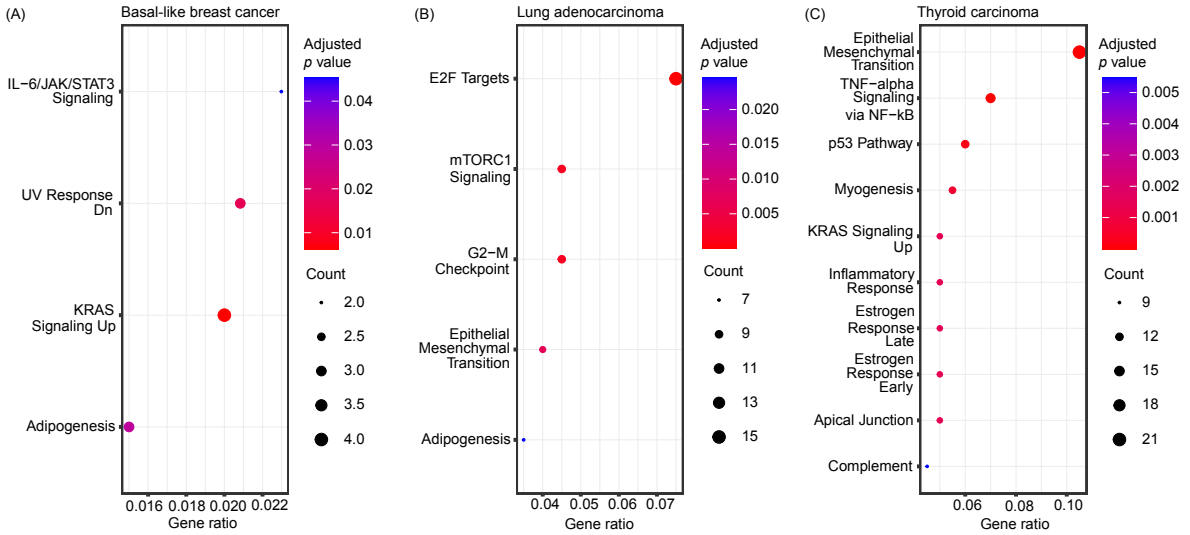

Figure 3. Enrichment analyses of predicted driver genes in **(A)** basal-like breast cancer, **(B)** lung adenocarcinoma, and **(C)** thyroid carcinoma using the "MSigDB Hallmark 2020" database. The top 10 most significantly enriched terms (adjusted *p*-value < 0.05) are included. The gene ratio on the *x* axis is the ratio between the number of predicted driver genes that intersect with genes annotated in the given hallmark gene set and the total number of genes annotated in the respective hallmark gene set. The point sizes reflect the number of driver genes playing a role in the respective hallmark gene set.

3.2 Association between expression of predicted driver genes and survival of cancer patients

Following discovery of driver genes in the three cancer (sub)types, we performed survival analysis to evaluate their prognostic potential. For this purpose, we first used Cox proportional hazards regression and found that the expression level of 20 of the predicted OCGs in lung adenocarcinoma had a significant effect on survival at the multivariate level when accounting for tumor stage, age of patients, and sex of patients. Similarly, expression of two of the predicted OCGs in thyroid carcinoma had a significant effect on survival. Thus, we deemed these 22 OCGs as prognostic (**Figure 4A**). Next, we examined whether high or low expression of these prognostic genes were associated with survival of the patients. For this, we divided the patients into high and low expression groups and assessed differences in survival through Kaplan-Meier survival analyses and log-rank tests. These analyses revealed a significant difference in survival between patients with high and low expression of 18 of the 20 prognostic OCGs in lung adenocarcinoma. The two OCGs that did not show a significant difference were *RPL39L* and *GINS2*. On the other hand, we did not observe a significant difference in survival between patients with high and low expression of the two predicted OCGs in thyroid carcinoma. These results together indicate a greater prognostic potential of OCGs compared to TSGs and additionally, a greater presence of prognostic OCGs in lung adenocarcinoma compared to basal-like breast cancer and thyroid carcinoma. It is, however, also worth mentioning that a smaller subset of driver genes were predicted in basal-like breast cancer with only one predicted OCG, indicating a more limited search pool for prognostic OCGs.

To highlight a few examples, multivariate Cox regression analysis identified *GNPNAT1*, *RRM2*, and *SLC2A1* as prognostic OCGs in lung adenocarcinoma with hazard ratios of 1.4, 1.3, and 1.3, respectively. In all three cases, patients with high expression of the OCG had a significantly lower survival probability compared to patients with low expression of these OCGs (**Figure 4B-D)** (*p*-values: < 0.0001, 0.0015 and 0.00055 for *GNPNAT1*, *RRM2* and *SLC2A1*, respectively). This aligns with the anticipated role of OCGs which are typically upregulated in cancer, indicating a worse prognosis.

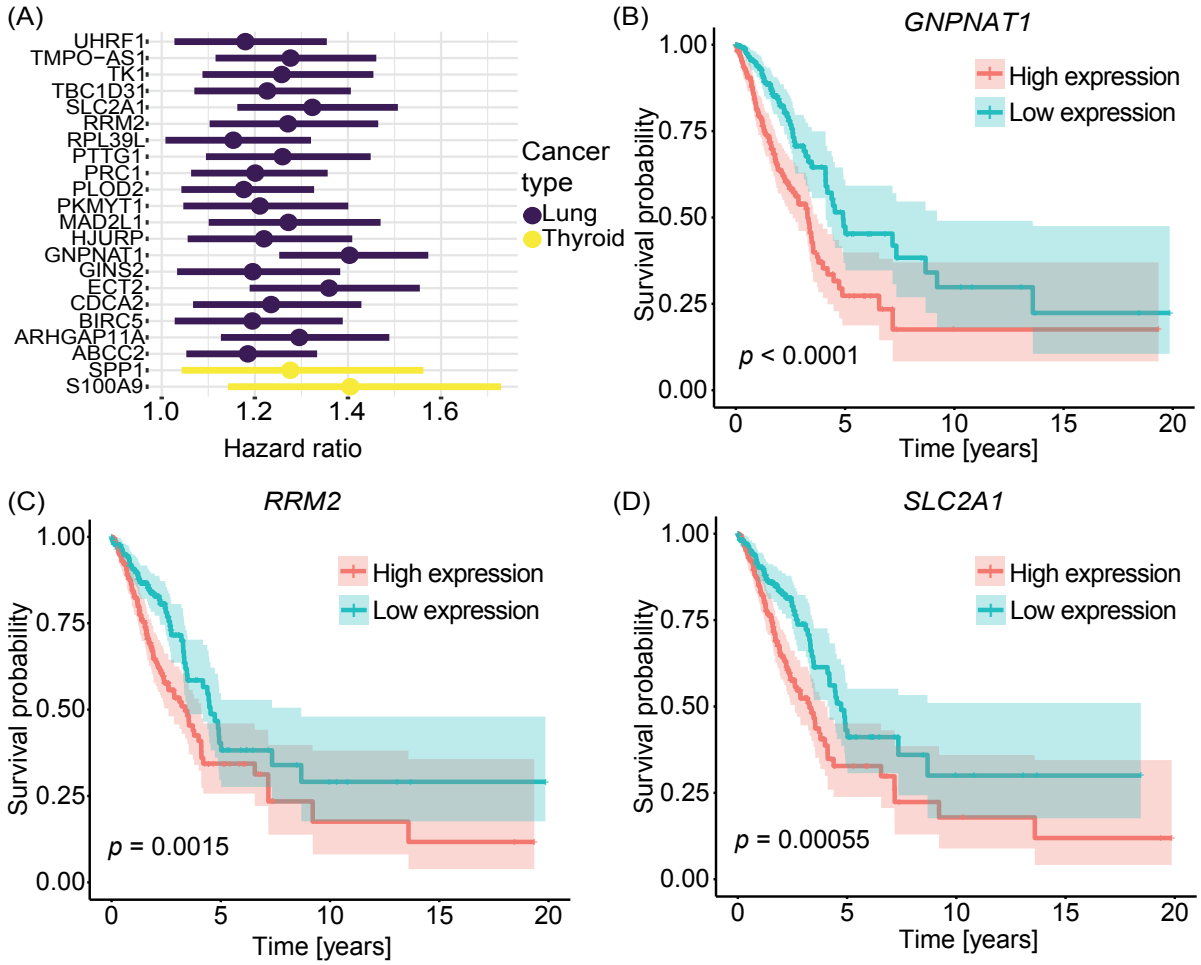

Figure 4. Survival analysis of predicted driver genes. **(A)** Hazard ratios from multivariate Cox proportional hazards regression of 20 of the predicted OCGs in lung adenocarcinoma and of two of the predicted OCGs in thyroid carcinoma. **(B-D)** Kaplan-Meier survival plots of three of the predicted OCGs in lung adenocarcinoma which were deemed prognostic from multivariate Cox regression analysis: **(B)** *GNPNAT1*, **(C)** *RRM2*, and **(D)** *SLC2A1*. Patients with expression values above and below the median expression level of the respective gene were divided into a high and low expression group, respectively. The *p*-values represent the significance of difference in survival between the two groups for each gene.

3.3 Predicted driver genes have therapeutic potential as drug targets

The potential of cancer driver genes as drug targets have previously been highlighted [68–70] and targeted therapies have been developed towards these genes. For example, a well-known example in breast cancer is the discovery of *HER2* as a cancer drug target which led to the development of the monoclonal antibody trastuzumab for targeting HER2 [10]. Thus, we next investigated the therapeutic potential of the predicted driver genes as drug targets by querying DGIdb for driver gene-drug interactions using only cancer-specific data sources. In basal-like breast cancer, we identified seven TSGs documented to interact with drugs in DGIdb. In lung adenocarcinoma, both OCGs and TSGs, numbering 12 each, were reported as drug targets. Finally, in thyroid carcinoma, 23 OCGs were reported as interacting with drugs (**Supplementary Figure S2**). Across all three cancer (sub)types, the number of driver gene-drug interactions varied between one and 55. Roughly half of all predicted driver genes interacted with one drug while the other half interacted with two or more drugs (**Figure 5A**).

Next, we examined those driver gene-drug interactions for which the interaction type was known. In basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma, we found three (all TSGs), six (three OCGs and three TSGs), and five (all OCGs) driver genes, respectively, for which the interaction type was known (**Figure 5B-D**). The majority of the drugs were inhibitors. The two driver genes with the most interactions were *PDGFRB* in basal-like breast cancer and *MET* in thyroid cancer. We predicted *PDGFRB* as a TSG in basal-like breast cancer which is annotated to interact with 16 inhibitors and three drugs with antagonist or inhibitor interactions. These drugs exert inhibitory mechanisms for targeting an OCG role of *PDGFRB*. As the gene-drug target interactions are not specific for a certain cancer type, these results might suggest a potential dual role of *PDGFRB*. On the other hand, *MET* predicted as an OCG in thyroid cancer interacted with 19 inhibitors, in accordance with the OCG role of *MET*. Moreover, in lung adenocarcinoma, the predicted OCG *RRM2*, which we also identified as a prognostic gene above, interacted with one inhibitor, gemcitabine. Previously, one study investigated the mRNA expression of RRM1 and RRM2 in tumors from patients with lung adenocarcinoma treated with docetaxel/gemcitabine. They found low RRM2 mRNA expression to be associated with a higher response rate to treatment compared to patients with high expression [71]. Similarly, in thyroid carcinoma, we observed an interaction between *ERBB3*, a member of the epidermal growth factor receptor (EGFR) family of receptor tyrosine kinases, and four inhibitors (sapitinib, poziotinib, gefitinib, and dacomitinib). These inhibitors, all classified as tyrosine kinase inhibitors [72–79], align with *ERBB3*'s predicted role as an OCG. Another example is the interaction between *EpCAM* and solitomab in lung adenocarcinoma. *EpCAM* is an epithelial cell adhesion molecule which plays a role in cell proliferation, migration, and signaling and is frequently overexpressed on the cell surface of several human carcinomas [80–82]. For instance, *EpCAM* was recently found to be upregulated in primary lung cancer compared to normal lung tissues caused by gene amplification and promoter hypomethylation [83]. Solitomab is a bispecific antibody binding to *EpCAM* and *CD3* [80] which previously has shown preliminary signs of antitumor activity [84].

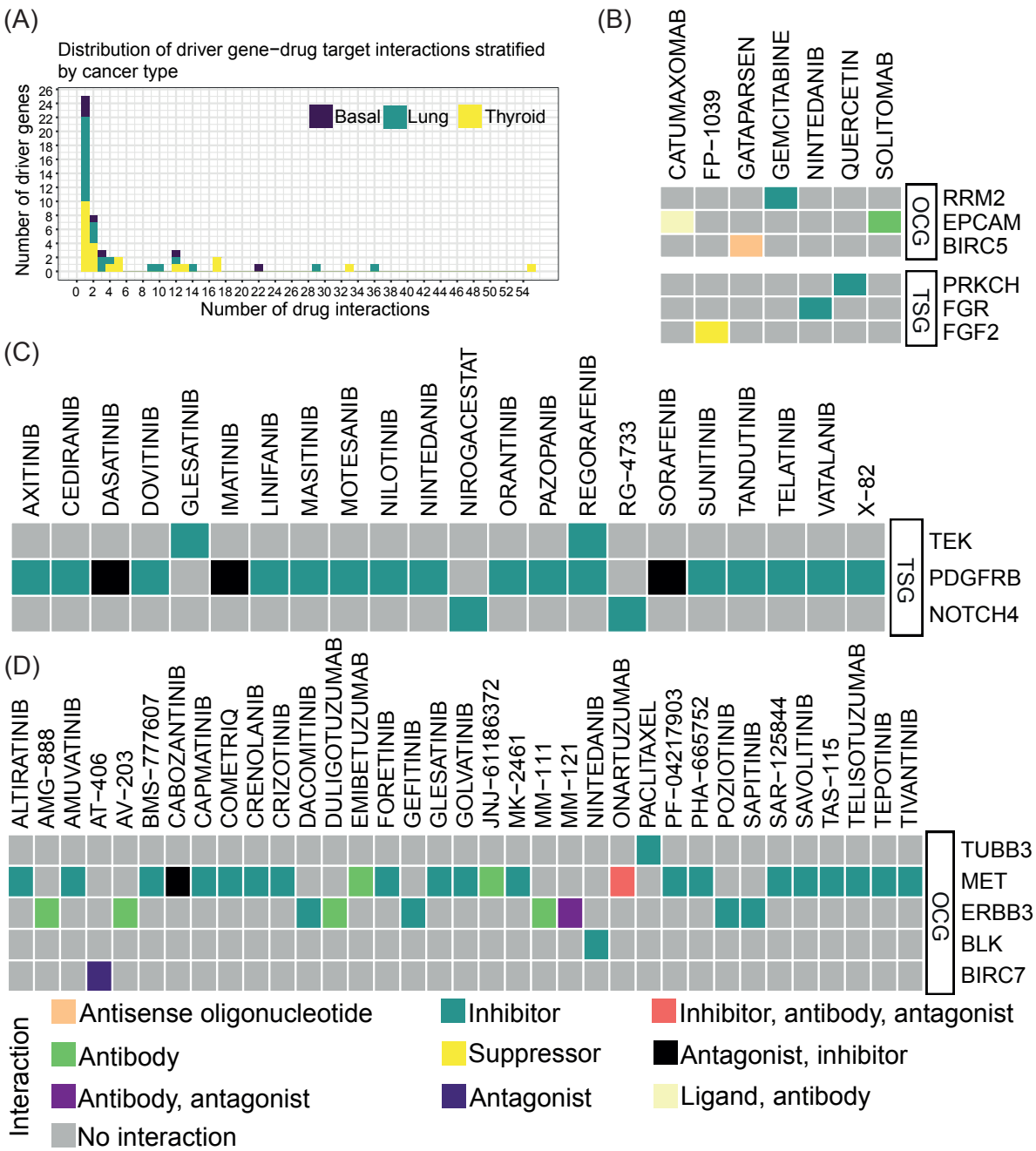

Figure 5. Exploration of predicted driver genes as drug targets. **(A)** Distribution of driver gene-drug interactions stratified by cancer type with the number of drug interactions on the *x* axis and number of driver genes on the *y* axis. **(B-D)** Heatmaps visualizing driver gene-drug interactions in **(B)** lung adenocarcinoma, **(C)** basal-like breast cancer, and **(D)** thyroid carcinoma. Only those driver gene-drug interactions where the interaction type was known are included in the heatmaps. The type of interaction is shown in different colors. The driver genes are divided into OCGs and TSGs.

Finally, we also discovered the family of *BIRC* genes as potential drug targets where we predicted *BIRC5* and *BIRC7* as OCGs in lung adenocarcinoma and thyroid carcinoma, respectively (**Figure 5B, D**). These two genes belong to the family of inhibitors of apoptosis proteins (IAPs) which are antiapoptotic proteins modulating signal transduction, cell cycle, and apoptosis [85]. We found an interaction between *BIRC5* and

gataparsen, an antisense oligonucleotide, which has previously been tested in a phase I study [86]. *BIRC5* inhibits apoptosis and promotes cell proliferation and angiogenesis which together with its upregulation in various human malignancies makes it a promising target for anti-cancer therapy [86–89]. Additionally, one study found *BIRC5* to promote lung adenocarcinoma progression by modulating PD-L1 expression via mTOR signaling [90]. In terms of *BIRC7*, we found an interaction of this gene with AT−406 which has been reported as a potent antagonist of multiple IAPs [91]. AT-406 has further been found to inhibit progression of human ovarian cancer [92].

4 Discussion

In this study, we present updates to our previously developed framework for driver gene discovery by incorporating abnormal methylation patterns as mechanistic explanations of deregulated expression profiles of driver genes. To this end, we integrated EpiMix into Moonlight in a new functionality called GMA. Afterwards, we applied GMA to three cancer (sub)types and predicted driver genes with prognostic and therapeutic potential.

Previous efforts have also integrated expression and methylation data to predict cancer biomarkers and driver genes with prognostic potential. Shen et al. used methylation and expression data to predict driver genes in breast cancer and found 89 basal-like specific driver genes [23]. Xu et al. also integrated DNA methylation and expression data and found four cervical cancer-specific methylation biomarkers which could distinguish tumors from normal tissues [93]. Similarly, Li et al. combined methylation and expression data with survival analysis and found eight mRNAs as potential biomarkers for predicting prognosis of lung adenocarcinoma [94]. Few driver gene prediction tools have been developed that integrate epigenetic modifications including DNA methylation [9] and even fewer tools classify driver genes into TSGs and OCGs [9]. Following rigorous exploration of driver gene prediction tools to compare our predicted driver genes with, we encountered a scarcity of suitable tools for this purpose [9]. For example, MethSig [21] uses a novel statistical inference framework to estimate tumor promoter hypermethylation to infer DNA methylation cancer drivers. However, here, we are also interested in examining hypomethylated mechanisms. DORGE [95] is another tool which incorporates epigenetic features in its machine learning model for prediction of both TSGs and OCGs. However, its availability appears to be ad-hoc rather than an installable tool. As we also reviewed in [9], no universal method for driver gene prediction exists, complicating benchmarking and comparisons of these tools.

In this study, we found various OCGs and TSGs annotated as drug targets in DGIdb that also previously have been described and evaluated as drug targets in literature (*RRM2*, *ERBB3*, *EpCAM*, *BIRC5*, and *BIRC7*), demonstrating Moonlight's ability to find critical driver genes with high relevance. For example, we found an interaction between *BIRC5* predicted as an OCG in lung adenocarcinoma and gataparsen, an antisense oligonucleotide, which has previously been tested in a phase I study [86].

Identifying driver genes is crucial for discovering novel drug targets and designing new treatment strategies. In particular, basal-like breast cancer faces a clinical need as this subtype is negative for HER2 and hormone receptors which can be targeted in other breast cancer subtypes. This also contributes to the aggressive nature of the basal-like subtype, resulting in a poorer prognosis and limited array of treatment options compared to the other subtypes [96]. Similarly, novel therapeutic interventions are also needed for lung adenocarcinoma. This subtype demonstrates high frequency of a few drivers [97–102] towards which targeted therapies have been developed showing great progress [103]. Nevertheless, lung adenocarcinoma has a poor prognosis and low survival rate [104] due to drug resistance [105] and metastasis [106]. Thus, discovery of novel driver genes with therapeutic potential is crucial in lung adenocarcinoma. Similarly to lung adenocarcinoma, targeted therapies have also been developed for thyroid carcinoma for instance directed towards mutated *BRAF*, however, resistance has been observed [107–110].

The majority of targeted therapies involve inhibitors of OCGs as developing drugs towards inactivated TSGs is more challenging [111,112] which can explain why most of our predicted driver genes interact with inhibitors (**Figure 5**). In fact, kinase inhibitors are one of the most successful categories of cancer drugs developed [111]. Nevertheless, strategies to target TSGs have emerged. This includes for example indirect targeting exploiting the fact that inactivation of a TSG results in the activation of downstream signaling pathways such as the activation of kinases which can be targeted [111,112]. Moreover, oligonucleotide therapies have been proposed to restore the function of TSGs [112]. Thus, prediction of driver genes in the context of signaling pathways is vital and greater knowledge of these biological processes are needed for these purposes.

What sets Moonlight apart from similar tools is its ability to classify driver genes into OCGs and TSGs which few driver gene prediction tools do [9]. Distinguishing between these two driver gene classes is important given their opposing biological roles. OCGs drive cell proliferation and survival whereas TSGs to a larger extent protects genome integrity and inhibits cell growth [113,114]. Understanding the balance and interplay between these gene roles is key for untangling the complex landscape of cancer. Additionally, Moonlight offers a flexible multi-omics approach where the framework can be used with just gene expression and at least one additional -omics layer such as mutation or methylation data. This means that the tool is applicable even if the user has only mutation or methylation data and does not necessarily require both. Furthermore, Moonlight allows for prediction of driver genes in connection with the Hallmarks of Cancer [6–8]. The previous functionality that we presented [15] focused on illuminating the “genome instability & mutation” enabling characteristic. The new functionality that we present here focus on the “nonmutational epigenetic programming” enabling characteristic, added in the most recent version of the cancer hallmarks [8]. The conducted case studies take their starting points in the “sustaining proliferative signaling” and “resisting cell death” core hallmarks. Driver gene discovery in the context of these cancer hallmarks is important for increasing our understanding of these phenotypes to eventually revert them.

In the future, we envision incorporation of additional secondary -omics layers such as chromatin accessibility and copy number variation coupled with expression changes. Moreover, we would like to implement proteomics and single-cell RNA sequencing data as additional input data types. In the future, cancer driver genes will likely continue to be an active area of investigation given their key role in driving cancer development, potential as drug targets, and importance in facilitating more effective treatment strategies and ultimately, precision medicine strategies.

5 Conclusion

In this study, we presented updates to the Moonlight framework, expanding upon features that we previously developed. The novel functionality presented here called GMA investigates abnormal DNA methylation patterns manifested in deregulated expression profiles to predict driver genes. We achieved this by integrating the EpiMix tool. Driver gene discovery is important for increasing our knowledge of cancer development and for finding cancer (sub)type-specific biomarkers with prognostic and therapeutic potential. Here, we discovered 33, 190, and 263 driver genes in basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma, respectively, using GMA where a subset of these demonstrated prognostic and therapeutic potential. In the future, we envision inclusion of additional -omics layers and compliance with proteomics and single-cell RNA-sequencing as input data in Moonlight. Additionally, with Moonlight, we categorize predicted driver genes into OCGs and TSGs, a crucial task given their opposing roles in cancer biology, ultimately important for developing novel and effective treatment strategies.

Acknowledgements

EP's group is supported by Elegant North (EN; Exploring Leukemia: Education Genetics And Technology; New Option for Rare diseases Towards Health), a collaboration between Oslo University Hospital, Capital Region of Denmark, Technical University Denmark (DTU), Abzu, Plesner, Region Skåne. EP's group is also supported by Danmarks Grundforskningsfond (Grant/Award Number: DNRF125), Hartmanns Fond (R241-A33877), LEO Foundation (LF17006), Novo-Nordisk Fonden Bioscience and Basic Biomedicine (NNF20OC0065262), and travel grants from Knud Højgaards Fond, William Demant Fonden, and Otto Mønsteds Fond. The results published here are in whole or in part based upon data generated by The Cancer Genome Atlas (TCGA) Research Network: https://www.cancer.gov/tcga. Research reported here was further supported by the National Cancer Institute (NCI) under award: R01 CA260271. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data accessibility

The data that support the findings of this study are openly available in The Cancer Genome Atlas. The data used for this analysis are available at the Genomic Data Commons (https://portal.gdc.cancer.gov). GitHub and OSF repositories associated with this study are available at https://github.com/ELELAB/Moonlight2R, https://github.com/ELELAB/Moonlight2_GMA_case_studies, and https://osf.io/j4n8q/.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. A Cancer Journal for Clinicians 2021; 71:209–249
2. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009; 458:719–724
3. Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. Science (1979) 2013; 340:1546–1558
4. Shen L, Shi Q, Wang W. Double agents: Genes with both oncogenic and tumor-suppressor functions. Oncogenesis 2018; 7:1–14
5. Datta N, Chakraborty S, Basu M, et al. Tumor suppressors having oncogenic functions: The double agents. Cells 2021; 10:1–26
6. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell 2000; 100:57–70
7. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell 2011; 144:646–674
8. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov 2022; 12:31–46
9. Nourbakhsh M, Degn K, Saksager AB, et al. Prediction of cancer driver genes and mutations: the potential of integrative computational frameworks. Brief Bioinform 2024; 25:1–16
10. Konda P, Garinet S, Van Allen EM, et al. Genome-guided discovery of cancer therapeutic targets. Cell Rep 2023; 42:1–18
11. Liu Y, Hu X, Han C, et al. Targeting tumor suppressor genes for cancer therapy. BioEssays 2015; 37:1277–1286
12. Yu X, Zhao H, Wang R, et al. Cancer epigenetics: from laboratory studies and clinical trials to precision medicine. Cell Death Discov 2024; 10:1–12
13. Iannuccelli M, Micarelli E, Surdo P Lo, et al. CancerGeneNet: Linking driver genes to cancer hallmarks. Nucleic Acids Res 2020; 48:D416–D421
14. Colaprico A, Olsen C, Bailey MH, et al. Interpreting pathways to discover cancer driver genes with Moonlight. Nat Commun 2020; 11:1–17
15. Nourbakhsh M, Saksager A, Tom N, et al. A workflow to study mechanistic indicators for driver gene prediction with Moonlight. Brief Bioinform 2023; 24:1–13
16. Ciriello G, Miller ML, Aksoy BA, et al. Emerging landscape of oncogenic signatures across human cancers. Nat Genet 2013; 45:1127–1133
17. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics 2009; 1:239–259
18. Esteller M. Epigenetics in Cancer. New England Journal of Medicine 2008; 358:1148–1159
19. Lakshminarasimhan R, Liang G. The role of DNA methylation in cancer. Adv Exp Med Biol 2016; 945:151–172

20. Søes S, Daugaard IL, Singers Sørensen B, et al. Hypomethylation and increased expression of the putative oncogene ELMO3 are associated with lung cancer development and metastases formation. Oncoscience 2014; 1:367–374
21. Pan H, Renaud L, Chaligne R, et al. Discovery of candidate DNA methylation cancer driver genes. Cancer Discov 2021; 11:2266–2281
22. Sun X, Yi J, Yang J, et al. An integrated epigenomic-transcriptomic landscape of lung cancer reveals novel methylation driver genes of diagnostic and therapeutic relevance. Theranostics 2021; 11:5346–5364
23. Shen X, Li S, Zhang L, et al. An Integrated Approach to Uncover Driver Genes in Breast Cancer Methylation Genomes. PLoS One 2013; 8:1–7
24. Zheng Y, Huang Q, Ding Z, et al. Genome-wide DNA methylation analysis identifies candidate epigenetic markers and drivers of hepatocellular carcinoma. Brief Bioinform 2018; 19:101–108
25. Li G xun, Ding Z yang, Wang Y wei, et al. Integrative analysis of DNA methylation and gene expression identify a six epigenetic driver signature for predicting prognosis in hepatocellular carcinoma. J Cell Physiol 2019; 234:11942–11950
26. Zhao SG, Chen WS, Li H, et al. The DNA methylation landscape of advanced prostate cancer. Nat Genet 2020; 52:778–789
27. Heery R, Schaefer MH. DNA methylation variation along the cancer epigenome and the identification of novel epigenetic driver events. Nucleic Acids Res 2021; 49:12692–12705
28. Gao C, Zhuang J, Li H, et al. Exploration of methylation-driven genes for monitoring and prognosis of patients with lung adenocarcinoma. Cancer Cell Int 2018; 18:1–11
29. Oubaddou Y, Oukabli M, Fenniche S, et al. BRCA1 Promoter Hypermethylation in Malignant Breast Tumors and in the Histologically Normal Adjacent Tissues to the Tumors: Exploring Its Potential as a Biomarker and Its Clinical Significance in a Translational Approach. Genes (Basel) 2023; 14:1–15
30. Hsu NC, Huang YF, Yokoyama KK, et al. Methylation of BRCA1 Promoter Region Is Associated with Unfavorable Prognosis in Women with Early-Stage Breast Cancer. PLoS One 2013; 8:1–6
31. Zheng Y, Jun J, Brennan K, et al. EpiMix is an integrative tool for epigenomic subtyping using DNA methylation. Cell Reports Methods 2023; 3:1–23
32. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. Contemp Oncol 2015; 19:A68–A77
33. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. Cell 2018; 173:283–285
34. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology 2013; 38:23–38
35. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016; 44:1–11
36. Mounir M, Lucchetta M, Silva TC, et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEX. PLoS Comput Biol 2019; 15:1–18
37. Silva TC, Colaprico A, Olsen C, et al. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. F1000Res 2016; 5:1–61
38. Gevaert O, Tibshirani R, Plevritis SK. Pancancer analysis of DNA methylation-driven genes using MethylMix. Genome Biol 2015; 16:1–13
39. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013; 14:1–14

40. Xie Z, Bailey A, Kuleshov M V., et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc 2021; 1:1–84
41. Kuleshov M V., Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016; 44:W90–W97
42. Therneau T. A package for survival analysis in R. R package, https://CRAN.R-project.org/package=survival 2022
43. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. 2000
44. Kassambara A, Kosinski M, Biecek P, et al. survminer: Drawing Survival Curves using 'ggplot2'. R package, https://CRAN.R-project.org/package=survminer 2021
45. Dardis C. survMisc: Miscellaneous Functions for Survival Data. R package, https://CRAN.R-project.org/package=survMisc 2022
46. Thurnherr T, Singer F, Stekhoven DJ, et al. Genomic variant annotation workflow for clinical applications. F1000Res 2016; 5:1–13
47. Wagner AH, Coffman AC, Ainscough BJ, et al. DGIdb 2.0: Mining clinically relevant drug-gene interactions. Nucleic Acids Res 2016; 44:D1036–D1044
48. Cannon M, Stevenson J, Stahl K, et al. DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. Nucleic Acids Res 2024; 52:D1227–D1235
49. Sondka Z, Bamford S, Cole CG, et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer 2018; 18:696–705
50. Rahimi M, Teimourpour B, Marashi SA. Cancer driver gene discovery in transcriptional regulatory networks using influence maximization approach. Comput Biol Med 2019; 114:1–11
51. Akhavan-Safar M, Teimourpour B. KatzDriver: A network based method to cancer causal genes discovery in gene regulatory network. BioSystems 2021; 201:1–13
52. Wei PJ, Zhang D, Li HT, et al. DriverFinder: A gene length-based network method to identify cancer driver genes. Complexity 2017; 2017:1–11
53. Pham VVH, Liu L, Bracken CP, et al. CBNA: A control theory based method for identifying coding and non-coding cancer drivers. PLoS Comput Biol 2019; 15:23
54. Dinstag G, Shamir R. PRODIGY: Personalized prioritization of driver genes. Bioinformatics 2020; 36:1831–1839
55. Wei PJ, Zhang D, Xia J, et al. LNDriver: Identifying driver genes by integrating mutation and expression data based on gene-gene interaction network. BMC Bioinformatics 2016; 17:1–10
56. Zhang D, Bin Y. DriverSubNet: A Novel Algorithm for Identifying Cancer Driver Genes by Subnetwork Enrichment Analysis. Front Genet 2021; 11:1–10
57. Li A, Chapuy B, Varelas X, et al. Identification of candidate cancer drivers by integrative Epi-DNA and Gene Expression (iEDGE) data analysis. Sci Rep 2019; 9:1–12
58. Han Y, Yang J, Qian X, et al. DriverML: A machine learning algorithm for identifying driver genes in cancer sequencing studies. Nucleic Acids Res 2019; 47:1–12
59. Gu H, Xu X, Qin P, et al. FI-Net: Identification of Cancer Driver Genes by Using Functional Impact Prediction Neural Network. Front Genet 2020; 11:1–15
60. Hou Y, Gao B, Li G, et al. MaxMIF: A New Method for Identifying Cancer Driver Genes through Effective Data Integration. Advanced Science 2018; 5:1–9
61. Zapata L, Susak H, Drechsel O, et al. Signatures of positive selection reveal a universal role of chromatin modifiers as cancer driver genes. Sci Rep 2017; 7:1–15
62. Terekhanova N V., Karpova A, Liang WW, et al. Epigenetic regulation during cancer transitions across 11 tumour types. Nature 2023; 623:432–441

63. D'Anello L, Sansone P, Storci G, et al. Epigenetic control of the basal-like gene expression profile via Interleukin-6 in breast cancer cells. Mol Cancer 2010; 9:1–13
64. Lin Y Te, Wu KJ. Epigenetic regulation of epithelial-mesenchymal transition: Focusing on hypoxia and TGF-β signaling. J Biomed Sci 2020; 27:1–10
65. Skrypek N, Goossens S, De Smedt E, et al. Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity. Trends in Genetics 2017; 33:943–959
66. Liu QL, Luo M, Huang C, et al. Epigenetic Regulation of Epithelial to Mesenchymal Transition in the Cancer Metastatic Cascade: Implications for Cancer Therapy. Front Oncol 2021; 11:1–13
67. Lu W, Kang Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. Dev Cell 2019; 49:361–374
68. Yang H, Gan L, Chen R, et al. From multi-omics data to the cancer druggable gene discovery: A novel machine learning-based approach. Brief Bioinform 2023; 24:1–11
69. Yoshimaru T, Nakamura Y, Katagiri T. Functional genomics for breast cancer drug target discovery. J Hum Genet 2021; 66:927–935
70. Zsákai L, Sipos A, Dobos J, et al. Targeted drug combination therapy design based on driver genes. Oncotarget 2019; 10:5255–5266
71. Souglakos J, Boukovinas I, Taron M, et al. Ribonucleotide reductase subunits M1 and M2 mRNA expression levels and clinical outcome of lung adenocarcinoma patients treated with docetaxel/gemcitabine. Br J Cancer 2008; 98:1710–1715
72. Atwell B, Chen CY, Christofferson M, et al. Sorting nexin-dependent therapeutic targeting of oncogenic epidermal growth factor receptor. Cancer Gene Ther 2023; 30:267–276
73. Robichaux JP, Le X, Vijayan RSK, et al. Structure-based classification predicts drug response in EGFR-mutant NSCLC. Nature 2021; 597:732–737
74. Robichaux JP, Elamin YY, Tan Z, et al. Mechanisms and clinical activity of an EGFR and HER2 exon 20-selective kinase inhibitor in non-small cell lung cancer. Nat Med 2018; 24:638–646
75. Cohen P, Cross D, Jänne PA. Kinase drug discovery 20 years after imatinib: progress and future directions. Nat Rev Drug Discov 2021; 20:551–569
76. Poels KE, Schoenfeld AJ, Makhnin A, et al. Identification of optimal dosing schedules of dacomitinib and osimertinib for a phase I/II trial in advanced EGFR-mutant non-small cell lung cancer. Nat Commun 2021; 12:1–12
77. van Alderwerelt van Rosenburgh IK, Lu DM, Grant MJ, et al. Biochemical and structural basis for differential inhibitor sensitivity of EGFR with distinct exon 19 mutations. Nat Commun 2022; 13:1–16
78. Pu X, Zhou Y, Kong Y, et al. Efficacy and safety of dacomitinib in treatment-naïve patients with advanced NSCLC harboring uncommon EGFR mutation: an ambispective cohort study. BMC Cancer 2023; 23:1–10
79. Momeny M, Zarrinrad G, Moghaddaskho F, et al. Dacomitinib, a pan-inhibitor of ErbB receptors, suppresses growth and invasive capacity of chemoresistant ovarian carcinoma cells. Sci Rep 2017; 7:1–12
80. Brischwein K, Schlereth B, Guller B, et al. MT110: A novel bispecific single-chain antibody construct with high efficacy in eradicating established tumors. Mol Immunol 2006; 43:1129–1143
81. Ferrari F, Bellone S, Black J, et al. Solitomab, an EpCAM/CD3 bispecific antibody construct (BiTE®), is highly active against primary uterine and ovarian carcinosarcoma cell lines in vitro. Journal of Experimental and Clinical Cancer Research 2015; 34:1–8

82. Keller L, Werner S, Pantel K. Biology and clinical relevance of EpCAM. Cell Stress 2019; 3:165–180
83. Cui Y, Li J, Liu X, et al. Dynamic Expression of EpCAM in Primary and Metastatic Lung Cancer Is Controlled by Both Genetic and Epigenetic Mechanisms. Cancers (Basel) 2022; 14:1–12
84. Kebenko M, Goebeler ME, Wolf M, et al. A multicenter phase 1 study of solitomab (MT110, AMG 110), a bispecific EpCAM/CD3 T-cell engager (BiTE®) antibody construct, in patients with refractory solid tumors. Oncoimmunology 2018; 7:1–12
85. Liang J, Zhao W, Tong P, et al. Comprehensive molecular characterization of inhibitors of apoptosis proteins (IAPs) for therapeutic targeting in cancer. BMC Med Genomics 2020; 13:1–13
86. Tanioka M, Nokihara H, Yamamoto N, et al. Phase I study of LY2181308, an antisense oligonucleotide against survivin, in patients with advanced solid tumors. Cancer Chemother Pharmacol 2011; 68:505–511
87. Xu L, Yu W, Xiao H, et al. BIRC5 is a prognostic biomarker associated with tumor immune cell infiltration. Sci Rep 2021; 11:1–13
88. Ryan BM, O'Donovan N, Duffy MJ. Survivin: A new target for anti-cancer therapy. Cancer Treat Rev 2009; 35:553–562
89. Garg H, Suri P, Gupta JC, et al. Survivin: A unique target for tumor therapy. Cancer Cell Int 2016; 16:1–14
90. Ma T, Gu J, Wen H, et al. BIRC5 Modulates PD-L1 Expression and Immune Infiltration in Lung Adenocarcinoma. J Cancer 2022; 13:3140–3150
91. Cai Q, Sun H, Peng Y, et al. A potent and orally active antagonist (SM-406/AT-406) of multiple inhibitor of apoptosis proteins (IAPs) in clinical development for cancer treatment. J Med Chem 2011; 54:2714–2726
92. Brunckhorst MK, Lerner D, Wang S, et al. AT-406, an orally active antagonist of multiple inhibitor of apoptosis proteins, inhibits progression of human ovarian cancer. Cancer Biol Ther 2012; 13:804–811
93. Xu W, Xu M, Wang L, et al. Integrative analysis of DNA methylation and gene expression identified cervical cancer-specific diagnostic biomarkers. Signal Transduct Target Ther 2019; 4:1–11
94. Li R, Yang YE, Yin YH, et al. Methylation and transcriptome analysis reveal lung adenocarcinoma-specific diagnostic biomarkers. J Transl Med 2019; 17:1–14
95. Lyu J, Jessica Li J, Su J, et al. DORGE: Discovery of Oncogenes and tumoR suppressor genes using Genetic and Epigenetic features. Sci Adv 2020; 6:6784–6795
96. Mahmoud R, Ordóñez-Morán P, Allegrucci C. Challenges for Triple Negative Breast Cancer Treatment: Defeating Heterogeneity and Cancer Stemness. Cancers (Basel) 2022; 14:1–15
97. Mack PC, Klein MI, Ayers KL, et al. Targeted Next-Generation Sequencing Reveals Exceptionally High Rates of Molecular Driver Mutations in Never-Smokers With Lung Adenocarcinoma. Oncologist 2022; 27:476–486
98. Sun Y, Ren Y, Fang Z, et al. Lung adenocarcinoma from East Asian never-smokers is a disease largely defined by targetable oncogenic mutant kinases. Journal of Clinical Oncology 2010; 28:4616–4620
99. Sholl LM, Aisner DL, Varella-Garcia M, et al. Multi-institutional oncogenic driver mutation analysis in lung adenocarcinoma: The lung cancer mutation consortium experience. Journal of Thoracic Oncology 2015; 10:768–777

100. Zhang Y, Sun Y, Pan Y, et al. Frequency of driver mutations in lung adenocarcinoma from female never-smokers varies with histologic subtypes and age at diagnosis. Clinical Cancer Research 2012; 18:1947–1953
101. Hsu KH, Ho CC, Hsia TC, et al. Identification of Five Driver Gene Mutations in Patients with Treatment-Naïve Lung Adenocarcinoma in Taiwan. PLoS One 2015; 10:1–14
102. Kris MG, Johnson BE, Berry LD, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA 2014; 311:1998–2006
103. Saito M, Suzuki H, Kono K, et al. Treatment of lung adenocarcinoma by molecular-targeted therapy and immunotherapy. Surg Today 2018; 48:1–8
104. Lin JJ, Cardarella S, Lydon CA, et al. Five-year survival in EGFR-mutant metastatic lung adenocarcinoma treated with EGFR-TKIs. Journal of Thoracic Oncology 2016; 11:556–565
105. Kara A, Özgür A, Nalbantoğlu S, et al. DNA repair pathways and their roles in drug resistance for lung adenocarcinoma. Mol Biol Rep 2021; 48:3813–3825
106. Lin YY, Wang YC, Yeh DW, et al. Gene expression profile in primary tumor is associated with brain-tropism of metastasis from lung adenocarcinoma. Int J Mol Sci 2021; 22:1–15
107. Chen D, Su X, Zhu L, et al. Papillary thyroid cancer organoids harboring BRAFV600E mutation reveal potentially beneficial effects of BRAF inhibitor-based combination therapies. J Transl Med 2023; 21:1–18
108. Lee M, Untch BR, Xu B, et al. Genomic and Transcriptomic Correlates of Thyroid Carcinoma Evolution after BRAF Inhibitor Therapy. Molecular Cancer Research 2022; 20:45–55
109. Danysh BP, Rieger EY, Sinha DK, et al. Long-term vemurafenib treatment drives inhibitor resistance through a spontaneous KRAS G12D mutation in a BRAF V600E papillary thyroid carcinoma model. Oncotarget 2016; 7:30907–30923
110. Owen DH, Konda B, Sipos J, et al. KRAS G12V mutation in acquired resistance to combined BRAF and MEK inhibition in papillary thyroid cancer. Journal of the National Comprehensive Cancer Network 2019; 17:409–413
111. Morris LGT, Chan TA. Therapeutic targeting of tumor suppressor genes. Cancer 2015; 121:1357–1368
112. Gregory GL, Copple IM. Modulating the expression of tumor suppressor genes using activating oligonucleotide technologies as a therapeutic approach in cancer. Mol Ther Nucleic Acids 2023; 31:211–223
113. Croce CM. Oncogenes and Cancer. New England Journal of Medicine 2008; 358:502–511
114. Wang LH, Wu CF, Rajasekaran N, et al. Loss of tumor suppressor gene function in human cancer: An overview. Cellular Physiology and Biochemistry 2019; 51:2647–2693

Supporting Information

Supplementary_Figure_S1.pdf
Visualization of integration of Moonlight and EpiMix for prediction of cancer driver genes in lung adenocarcinoma and thyroid carcinoma.

Supplementary_Figure_S2.pdf
Visualization of number of driver gene-drug interactions in basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma.