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

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

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

*Corresponding author
E-mail: elpap@dtu.dk, elenap@cancer.dk (EP)

Short 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 an important development in Moonlight2 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 gene expression changes. 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). The associated case studies

presented here are available on GitHub (https://github.com/ELELAB/Moonlight2_GMA_case_studies) and OSF (https://osf.io/j4n8q/).

Author summary

Cancer is a complex disease and a main cause of mortality worldwide. This heterogeneous disease arises due to accumulation of changes which occur in driver genes that drive cancer progression when they are altered. These driver genes are commonly divided into oncogenes, which promote cancer, and tumor suppressors, which prevent it. A major goal of cancer research is identifying these driver genes, crucial for increasing our current understanding of cancer biology and for developing novel treatment approaches. A large number of cancer driver genes have already been identified. However, the underlying mechanisms for the alterations in these genes is challenging to predict given their context-dependent behavior and the complexity of cancer. Such explanations are the focus of this study with the aim of providing evidence of why certain genes do not function normally in cancer. Within this context, we present new functionalities to our previously developed cancer driver predictive framework, Moonlight. These new functionalities integrate multiple data types to predict oncogenes and tumor suppressors in a systems-biology-oriented manner that is freely available as a R package for the community.

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].

Driver genes are classified according to their function, i.e. oncogenes (OCGs) activated by gain-of-function mechanisms and tumor suppressor genes (TSGs) inactivated by loss-of-function mechanisms [3]. Recently, dual role genes also emerged which show context-dependent behavior and can 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 normal to tumor cells [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 the 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 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 mech-

anisms 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].

Here, we present novel functionalities to Moonlight2, expanding upon features presented in our previous work [15]. Specifically, we incorporate methylation evidence to Moonlight2 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 [21]. To 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) [22,23].

Design and implementation

Design and implementation of new functionalities in Moonlight

Here, we present new functionalities to Moonlight, our framework 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 (namely, mechanistic indicators). From this secondary layer, the critical driver genes are predicted among the oncogenic mediators. We recently presented Moonlight2 with the overall goal of implementing new functionalities to provide standardized and automatized solutions to the analysis of the mechanistic indicators. At first, we developed a secondary layer for mechanistic indicators based on mutational data [15].

In this contribution, we tackled the challenge of adding functionalities to Moonlight2 to cover mechanistic indicators related to methylation changes. Each layer can be applied together or separately depending on the source of -omics data available for the samples under analysis. 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 (Fig 1A). The biological foundation for GMA lies within the observed roles of DNA methylation in both physiological and cancer states. Under healthy conditions, DNA methylation serves an essential regulatory role in cells by regulating expression of genes [24]. 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] (Fig 1B).

GMA predicts methylation-driven driver genes by using EpiMix [21]. 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 [21], therefore being perfectly suitable to integrate with Moonlight's primary layer. Moreover, EpiMix is 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

149 DNA methylation state is associated with the expression of the corresponding genes
150 they map to. Moreover, the methylation state (e.g. hypo- or hypermethylated) of each
151 CpG site is reported. This table is integrated with the main output table from Moon-
152 light's primary layer, specifically the output from PRA, which provides a list of onco-
153 genic mediators and their putative driver role (e.g. putative TSG or OCG) (Fig 1C).
154 This integration step involves the following: for each oncogenic mediator, the number
155 of associated CpG sites is summarized. EpiMix's predictions of methylation state and
156 Moonlight's predictions of driver gene role are then compared and used to assess
157 whether the gene's methylation status supports the putative role (OCG or TSG) of the
158 oncogenic mediator. These comparisons are subsequently used to define the driver
159 genes (Fig 1D). Those oncogenic mediators with correspondence between methyla-
160 tion state and putative driver role from EpiMix and Moonlight's primary layer are re-
161 tained as the final set of driver genes. See S1 Text for a detailed description of this
162 comparison.
163 As input, GMA requires i) a gene expression matrix with genes in rows and tumor and
164 normal samples in columns, ii) a methylation matrix with CpG sites in rows and tumor
165 and normal samples in columns which should be the same samples as in the expres-
166 sion data, iii) output of PRA from Moonlight's primary layer, i.e. the predicted onco-
167 genic mediators and their putative driver role, and finally, iv) output of a differential
168 expression analysis (DEA) which includes information about the DEGs. In return, GMA
169 outputs the following: i) a list of predicted driver genes categorized into TSGs and
170 OCGs, ii) a summary of the oncogenic mediators which includes the number of asso-
171 ciated CpG sites and evidence label, iii) a summary of various annotations found to all
172 DEGs input to Moonlight on the gene and methylation level, and iv) raw EpiMix results
173 corresponding to applying EpiMix on the input data independent of the GMA function.

We have also created three functions for visualizing genes and methylation states: plotGMA which visualizes the number of differentially methylated hypo-, hyper- or dual-methylated CpG sites, plotMoonlightMet which visualizes the effect of genes on biological processes estimated in Moonlight's primary layer, and plotMetExp which calls a visualization function from EpiMix, EpiMix_PlotModel, to display gene expression and methylation levels of a specific gene and CpG site [21].

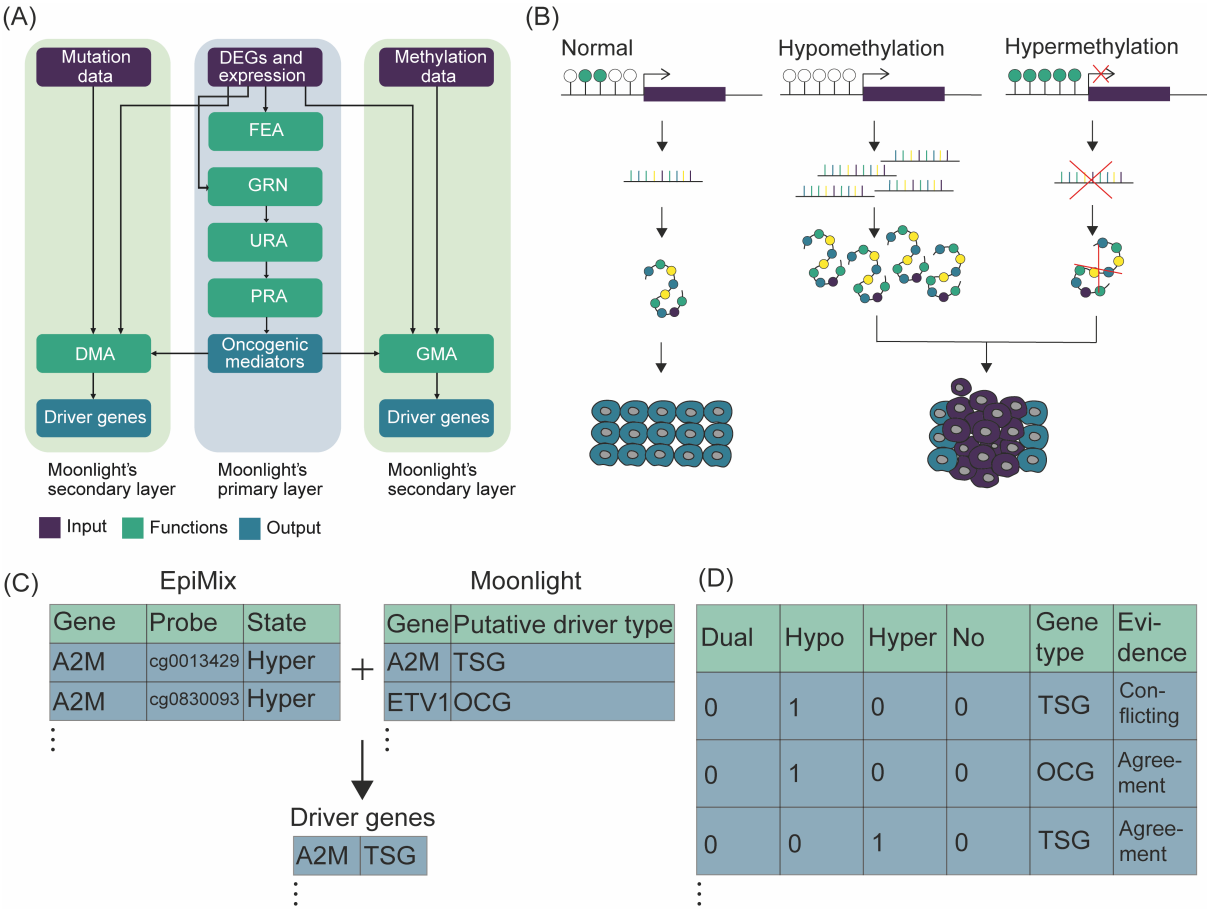

Fig 1. Overview of the Moonlight framework with new methylation functionality. (A) Moonlight consists of a primary layer requiring differentially expressed genes and gene expression data as input. The primary layer predicts oncogenic mediators through a series of functions called functional enrichment 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. Those oncogenic mediators labeled with an "agreement" evidence are retained as the final set of predicted driver genes.

Application of new functionality to three cancer (sub)types

Following implementation of the new functionality, GMA, in Moonlight2, we conducted a case study applying GMA to basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma data from TCGA to discover methylation-driven driver genes. Moreover, we compared these predicted drivers with mutation-driven drivers by applying our previously developed secondary mutational layer called Driver Mutation Analysis (DMA) [15]. Detailed methods behind this case study are included in S2 Text.

Results

Case study: Prediction of driver genes with differential methylation in three different cancer types using Moonlight2

To showcase the new functionality in Moonlight2 and predict driver genes driven by methylation changes, we applied Moonlight2 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 (Table 1). Following DEA, Moonlight's primary layer predicted 159, 1228, and 1598 oncogenic mediators in these three cancer

224 (sub)types, respectively (Table 1). Additionally, EpiMix alone identified 9483, 10018,
225 and 6142 functional gene-CpG pairs in these three cancer (sub)types, respectively.
226 These functional gene-CpG pairs represent differentially methylated CpG sites whose
227 DNA methylation state is associated with the expression of the corresponding genes
228 they map to. The number of hits discovered individually from EpiMix and Moonlight's
229 primary layer indicate a substantial volume of significant associations. Consequently,
230 integrating the results from EpiMix with the oncogenic mediators identified in Moon-
231 light's primary layer, as implemented in GMA presented here, help to narrow down the
232 most critical findings and yield the benefits of both approaches. From GMA, we found
233 that those oncogenic mediators in basal-like breast cancer that are associated with
234 differentially methylated CpGs include 38 hypomethylated CpGs, 165 hypermethyl-
235 ated CpGs, and 22 methylated CpGs with a dual status, meaning the CpG site was
236 found hypomethylated in cancer tissues from some patients, while hypermethylated in
237 other patients. Similarly, oncogenic mediators in lung adenocarcinoma that are asso-
238 ciated with differentially methylated CpGs include in total 218 hypomethylated CpGs,
239 625 hypermethylated CpGs, and 48 dual-methylated CpGs. Finally, oncogenic medi-
240 ators in thyroid carcinoma associated with differentially methylated CpGs contain in
241 total 945 hypomethylated CpGs, 305 hypermethylated CpGs, and 230 dual-methyl-
242 ated CpGs (Fig 2A).
243 Across all three cancer (sub)types, the number of differentially methylated CpG sites
244 mapped to the oncogenic mediators ranges between 0 and 28. The classifications of
245 methylation status in the oncogenic mediators in basal-like breast cancer are shown
246 in Fig 2B, generated with the plotGMA function. Next, we compared Moonlight's onco-
247 genic mediators with EpiMix' functional genes. For this, we included only those func-
248 tional genes that contained the same methylation state in all of its associated CpGs

249 and moreover, the dual states were excluded. In basal-like breast cancer, this com-
250 parison revealed 109 oncogenic mediators not associated with differentially methyl-
251 ated CpGs, 2754 functional genes not predicted as oncogenic mediators, 17 onco-
252 genic mediators with a “conflicting” evidence label, and 33 oncogenic mediators with
253 an “agreement” evidence label (Fig 2C). Consequently, these 33 oncogenic mediators
254 are retained as the final set of driver genes divided into 32 TSGs and 1 OCG (Table
255 1). Next, we visualized the effect of these predicted driver genes in basal-like breast
256 cancer on two well-known cancer hallmarks, apoptosis and proliferation of cells, using
257 the function plotMoonlightMet. These effects define the basis upon which the onco-
258 genic mediators are predicted from the PRA step in Moonlight’s primary layer, demon-
259 strating that the predicted OCGs have a positive effect on proliferation of cells and a
260 negative effect on apoptosis and vice versa for the predicted TSGs (Fig 2D). Similar
261 overviews for lung adenocarcinoma and thyroid carcinoma are shown in S1 Fig, which
262 resulted in a final prediction of 190 driver genes divided into 110 TSGs and 80 OCGs
263 in lung adenocarcinoma and 263 driver genes categorized into 5 TSGs and 258 OCGs
264 in thyroid carcinoma (Table 1). We did not discover any dual role genes across the
265 three cancer (sub)types, i.e. genes predicted as OCGs in one of the three cancer
266 (sub)types and as TSGs in another cancer (sub)type and vice versa.
267

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.

268

269 We then compared the predicted driver genes with the predicted oncogenic mediators
270 in each cancer (sub)type. We quantified these comparisons in terms of overlaps with
271 genes reported in the COSMIC Cancer Gene Census (CGC) [25]. Specifically, we
272 computed the precision as (TP/(TP+FP))*100 and sensitivity as (TP/(TP+FN))*100.
273 We defined the true positives (TP) as the overlap between the gene set (either the
274 driver genes or the oncogenic mediators) and the CGC, whereas the false positives
275 (FP) are those genes found in the gene set but are not included in CGC. In contrast,
276 the false negatives (FN) comprise those genes reported in CGC but are not predicted
277 in our gene set. For all three cancer (sub)types, we found that GMA had a greater
278 precision and lower sensitivity compared to using only Moonlight's primary layer (Fig
279 2E). A higher precision of GMA is desirable as it indicates that the predicted driver
280 gene sets have a higher fraction of genes from the CGC compared to the oncogenic
281 mediator sets. On the other hand, the higher sensitivity of using only Moonlight's pri-
282 mary layer compared to also using GMA might be attributed to the larger numbers of
283 oncogenic mediators. A larger number of oncogenic mediators results in a larger over-
284 lap between the CGC and the oncogenic mediators, thereby lowering the number of
285 FNs and increasing the sensitivity. In this case, prioritizing higher precision over sen-
286 sitivity is preferable since our aim is to find the most crucial driver genes among the
287 oncogenic mediators. Thus, a higher precision indicates a greater proportion of TPs,
288 corresponding with our objective. Next, we also evaluated the significance of associa-
289 tion between the gene sets and the CGC using a Fisher's exact test (Table 2). We only

290 found the oncogenic mediator and driver gene sets from basal-like breast cancer to
291 have a significant association with genes in the CGC (*p*-value = 0.000392 for the on-
292 cogenic mediators predicted using Moonlight's primary layer and *p*-value = 0.00228
293 for the driver genes predicted using GMA). However, in all three cancer (sub)types,
294 we found the driver genes to have a higher odds ratio than the oncogenic mediators,
295 demonstrating a greater association between the driver gene sets and the CGC com-
296 pared to the oncogenic mediators (Table 2).

297
298

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.

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

*p-value < 0.05.
Abbreviations: CGC, Cancer Gene Census; GMA, Gene Methylation Analysis.

299
300
301 While these results together demonstrate the added value of GMA, it is worth high-
302 lighting certain limitations. Notably, the driver genes reported in CGC are mainly based
303 on mutation evidence. 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 [26–37]. 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.

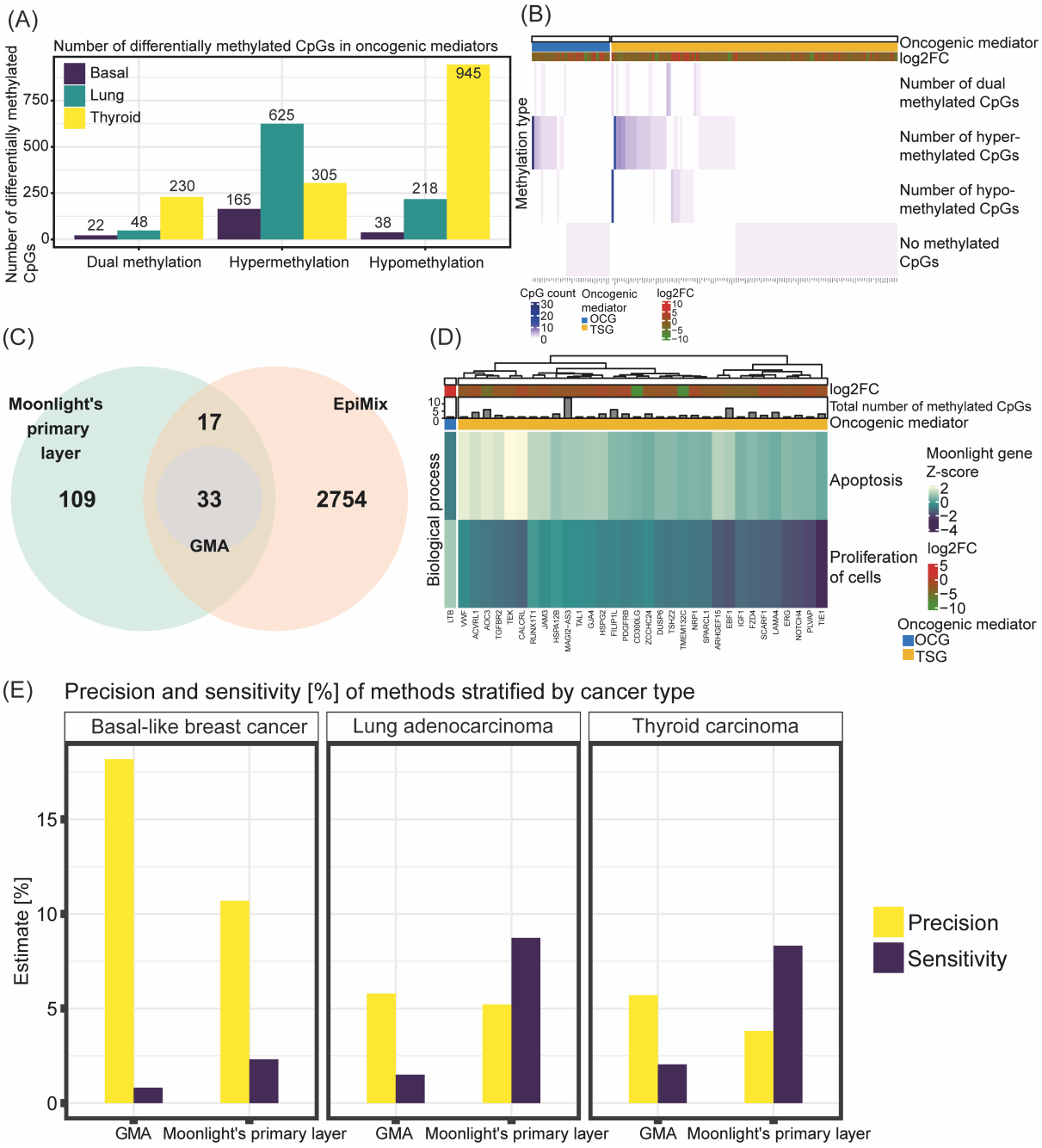

Fig 2. Integration of Moonlight and EpiMix for prediction of cancer driver genes. (A) Number of differentially methylated CpGs as found from EpiMix in oncogenic mediators predicted from Moonlight's primary layer. The differentially methylated CpGs are categorized into methylation status and stratified 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.

To investigate biological roles of the predicted driver genes, we performed enrichment analyses (Fig 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 [38]. Furthermore, IL−6/JAK/STAT3 signaling was significantly enriched among the predicted driver genes in basal-like breast cancer (Fig 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 [39]. Additionally, epithelial-mesenchymal transition (EMT) is a 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 [40–43].

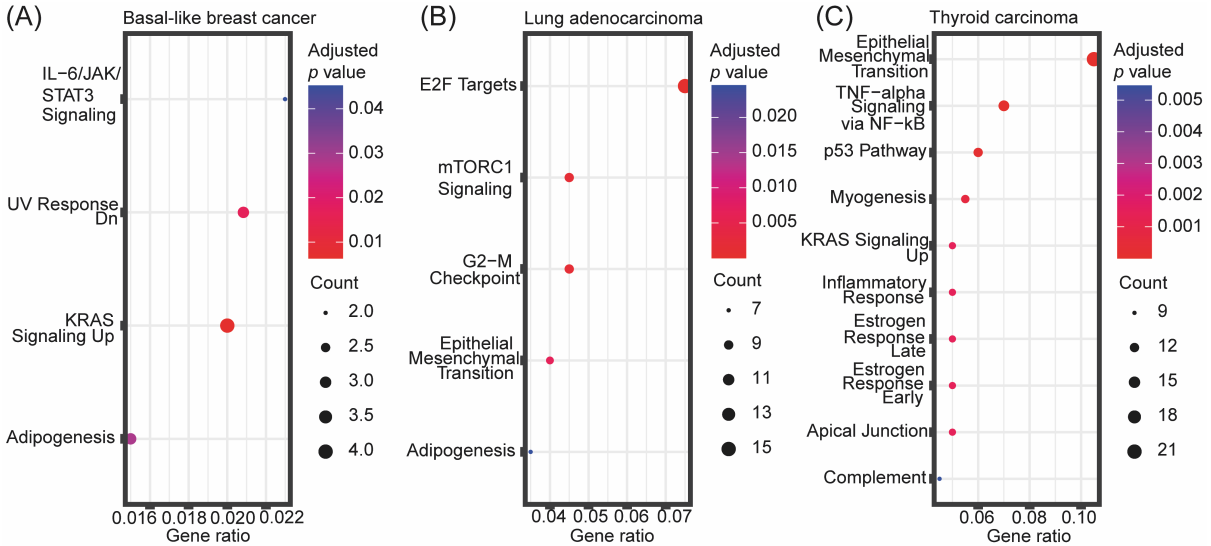

Fig 3. Enrichment analyses of predicted driver genes. Enrichment analysis of predicted driver genes in **(A)** basal-like breast cancer, **(B)** lung adenocarcinoma, and **(C)** thyroid carcinoma using the "MSigDB Hallmark 2020" database. The top 10 most significantly enriched terms (adjusted *p*-value < 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.

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

We performed survival analysis to evaluate the prognostic potential of the predicted driver genes. 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 (Fig 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

383 revealed a significant difference in survival between patients with high and low expres-
384 sion of 18 of the 20 prognostic OCGs in lung adenocarcinoma. The two OCGs that did
385 not show a significant difference were *RPL39L* and *GINS2*. On the other hand, we did
386 not observe a significant difference in survival between patients with high and low ex-
387 pression of the two predicted OCGs in thyroid carcinoma. These results together indi-
388 cate a greater prognostic potential of OCGs compared to TSGs and additionally, a
389 greater presence of prognostic OCGs in lung adenocarcinoma compared to basal-like
390 breast cancer and thyroid carcinoma. It is, however, worth mentioning that a smaller
391 subset of driver genes was predicted in basal-like breast cancer with only one pre-
392 dicted OCG, indicating a more limited search pool for prognostic OCGs.
393 To highlight a few examples, multivariate Cox regression analysis identified
394 *GNPNAT1*, *RRM2*, and *SLC2A1* as prognostic OCGs in lung adenocarcinoma with
395 hazard ratios of 1.4, 1.3, and 1.3, respectively. In all three cases, patients with high
396 expression of the OCG had a significantly lower survival probability compared to pa-
397 tients with low expression of these OCGs (Figs 4B-D) (p-values: < 0.0001, 0.0015 and
398 0.00055 for *GNPNAT1*, *RRM2* and *SLC2A1*, respectively). This aligns with the antici-
399 pated role of OCGs which are typically upregulated in cancer, indicating a worse prog-
400 nosis.
401

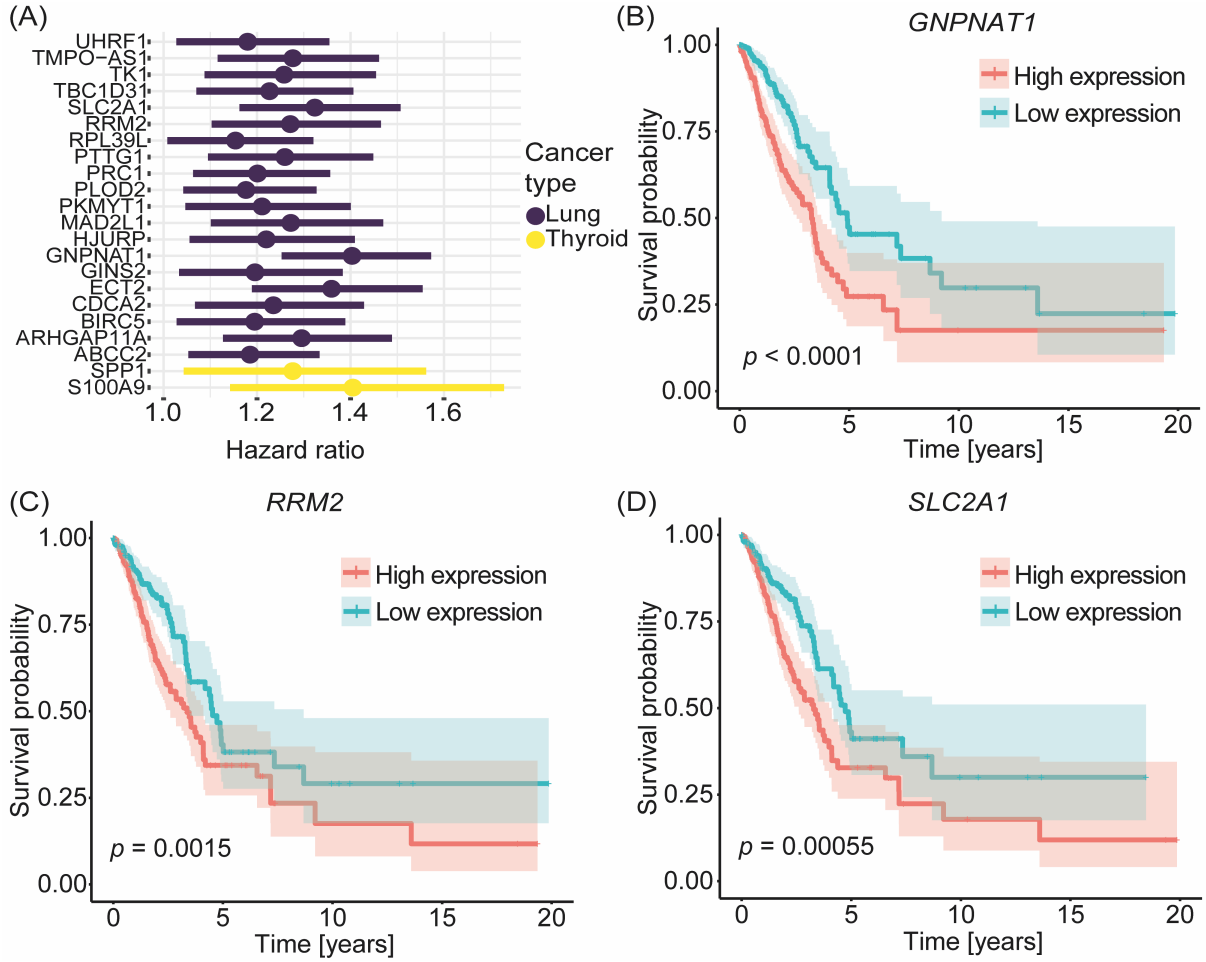

Fig 4. Survival analysis of predicted driver genes. (A) Hazard ratios from multivariate Cox proportional hazards regression of 20 of the predicted OCGs in lung adenocarcinoma and of two of the predicted OCGs in thyroid carcinoma. **(B-D)** Kaplan-Meier survival plots of three of the predicted OCGs in 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.

Predicted driver genes have therapeutic potential as drug targets

The potential of cancer driver genes as drug targets have previously been highlighted [44–46] and targeted therapies have been developed towards these genes. Thus, we next investigated the therapeutic potential of the predicted driver genes as drug targets by querying the Drug-Gene Interaction Database (DGIdb) [47] for driver gene-drug interactions using only cancer-specific data sources. In basal-like breast cancer, we

419 identified seven TSGs documented to interact with drugs in DGIdb. In lung adenocar-
420 cinoma, both OCGs and TSGs, numbering 12 each, were reported as drug targets.
421 Finally, in thyroid carcinoma, 23 OCGs were reported as interacting with drugs (S2
422 Fig). Across all three cancer (sub)types, the number of driver gene-drug interactions
423 varied between one and 55. Roughly half of all predicted driver genes interacted with
424 one drug while the other half interacted with two or more drugs (Fig 5A).
425 Next, we examined those driver gene-drug interactions for which the interaction type
426 was known. In basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma,
427 we found three (all TSGs), six (three OCGs and three TSGs), and five (all OCGs)
428 driver genes, respectively, for which the interaction type was known (Figs 5B-D). The
429 majority of the drugs were inhibitors. The two driver genes with the most interactions
430 were *PDGFRB* in basal-like breast cancer and *MET* in thyroid cancer. We predicted
431 *PDGFRB* as a TSG in basal-like breast cancer which is annotated to interact with 16
432 inhibitors and three drugs with antagonist or inhibitor interactions. These drugs exert
433 inhibitory mechanisms for targeting an OCG role of *PDGFRB*. As the gene-drug target
434 interactions are not specific for a certain cancer type, these results might suggest a
435 potential dual role of *PDGFRB*. On the other hand, *MET* predicted as an OCG in thy-
436 roid cancer interacted with 19 inhibitors, in accordance with the OCG role of *MET*.
437 Moreover, in lung adenocarcinoma, the predicted OCG *RRM2*, which we also identi-
438 fied as a prognostic gene above, interacted with one inhibitor, gemcitabine. Previously,
439 one study investigated the mRNA expression of RRM1 and RRM2 in tumors from pa-
440 tients with lung adenocarcinoma treated with docetaxel/gemcitabine. They found low
441 RRM2 mRNA expression to be associated with a higher response rate to treatment
442 compared to patients with high expression [48]. Similarly, in thyroid carcinoma, we
443 observed an interaction between *ERBB3*, a member of the epidermal growth factor

444 receptor (EGFR) family of receptor tyrosine kinases, and four inhibitors (sapitinib, pozi-
445 otinib, gefitinib, and dacomitinib). These inhibitors, all classified as tyrosine kinase in-
446 hibitors [49–56], align with *ERBB3*'s predicted role as an OCG. Another example is
447 the interaction between *EpCAM* and solitomab in lung adenocarcinoma. *EpCAM* is an
448 epithelial cell adhesion molecule which plays a role in cell proliferation, migration, and
449 signaling and is frequently overexpressed on the cell surface of several human carci-
450 nomas [57–59]. For instance, *EpCAM* was recently found to be upregulated in primary
451 lung cancer compared to normal lung tissues caused by gene amplification and pro-
452 moter hypomethylation [60]. Solitomab is a bispecific antibody binding to *EpCAM* and
453 *CD3* [57] which previously has shown preliminary signs of antitumor activity [61].

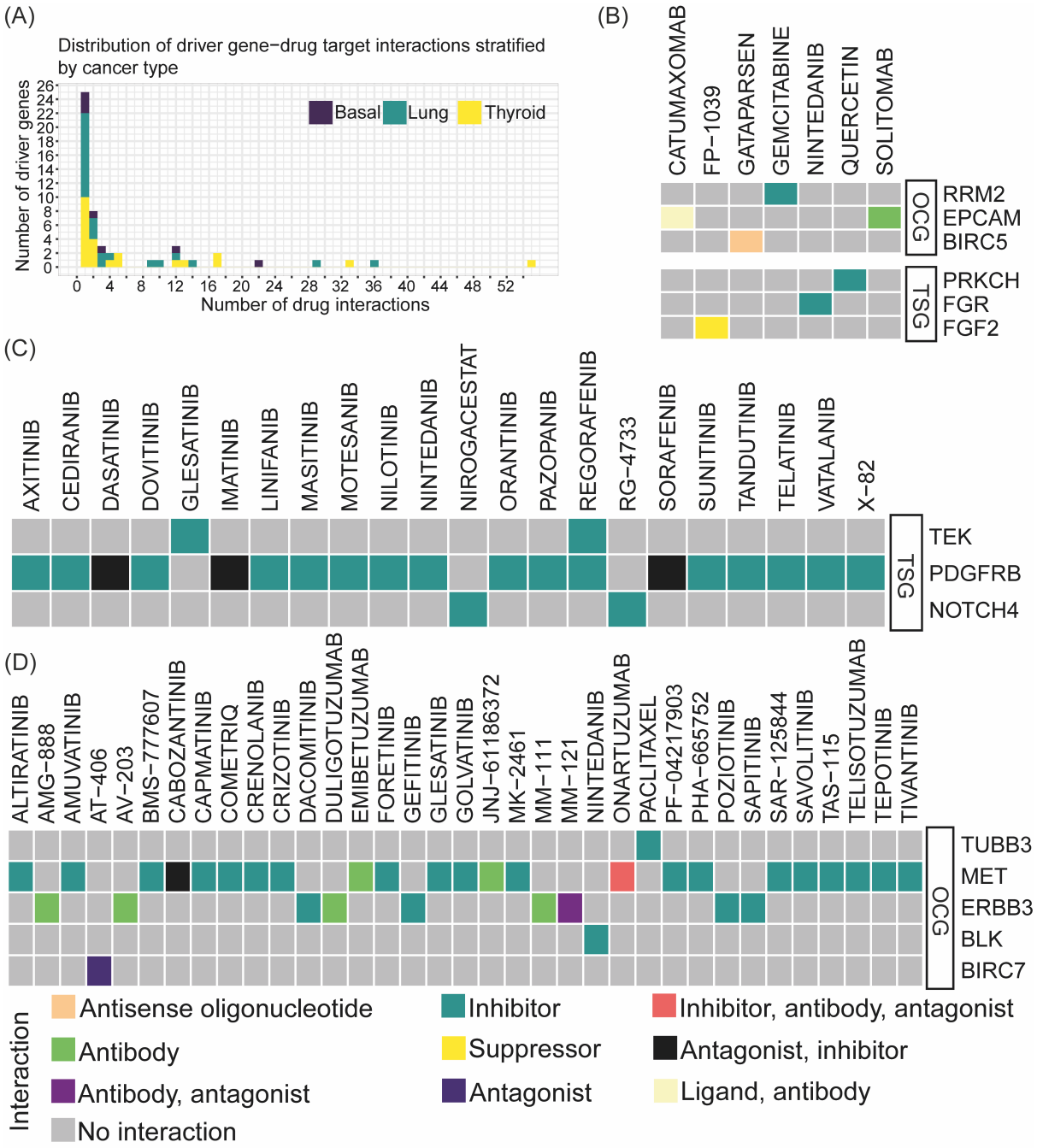

Fig 5. Exploration of predicted driver genes as drug targets. (A) Distribution of driver gene-drug interactions stratified by cancer type with the number of drug interactions on the *x* axis and number of driver genes on the *y* axis. **(B-D)** Heatmaps visualizing driver gene-drug interactions in **(B)** 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.

Integrating the results from DMA and GMA functions of Moonlight2

Next, we also applied the Moonlight2 DMA functionality [15] to the data used for the case studies above to show the potential of integrating different mechanistic indicators.

467 For basal-like breast cancer, DMA predicted 46 driver genes (10 OCGs and 36 TSGs),
468 while GMA predicted 33 driver genes (32 OCGs and 1 TSG) (Figs 6A-C). For lung
469 adenocarcinoma, DMA predicted 842 driver genes (490 OCGs and 352 TSGs), while
470 GMA predicted 190 (80 OCGs and 110 TSGs) (Figs 6D-F). Both secondary layers
471 predicted a larger number of driver genes in lung adenocarcinoma than basal-like
472 breast cancer (Table 1, Fig 6). This is likely a direct consequence of Moonight's pri-
473 mary layer, which identified a larger number of oncogenic mediators in lung adenocar-
474 cinoma than basal-like breast cancer. At the same time, DMA predicted a larger num-
475 ber of driver genes for both datasets than GMA, with a larger proportion in lung ade-
476 nocarcinoma than basal-like breast cancer (~4.5 times against ~1.4 times, respec-
477 tively). This observation aligns with previous reports suggesting that lung adenocarci-
478 noma exhibits a high mutation burden [62,63], suggesting that DMA was able to iden-
479 tify a larger number of driver mutations overall. In most cases, we found an overlap
480 between driver genes identified by DMA and GMA, which suggests multiple mecha-
481 nisms at play. In basal-like breast cancer, 13 driver genes were predicted by both DMA
482 and GMA, which were all TSGs (Figs 6A-C). In lung adenocarcinoma, 141 driver genes
483 (63 OCGs and 78 TSGs) were identified by both methods (Figs 6D-F). In the case of
484 lung adenocarcinoma, and more so than in basal-like breast cancer, the driver genes
485 predicted by GMA were in good part also predicted by DMA.
486 Next, we performed enrichment analysis of the DMA predicted driver genes in basal-
487 like breast cancer and lung adenocarcinoma to understand whether DMA and GMA
488 can identify distinct or overlapping biological mechanisms. The significantly enriched
489 terms (adjusted p-value < 0.05) among the DMA predicted driver genes in basal-like
490 breast cancer were angiogenesis, KRAS signaling up, epithelial mesenchymal transi-
491 tion, and IL-2/STAT5 signaling (Fig 7A) while among the GMA predicted drivers they

492 were IL-6/JAK/STAT3 signaling, UV response dn, KRAS signaling up and adipogene-
493 sis (Fig 3A). Thus, results from both GMA and DMA were enriched in the KRAS sig-
494 naling term only.
495 Both GMA and DMA identified *NRP1* as a driver gene involved in KRAS signaling.
496 *NRP1* has been shown to be highly expressed in different cancer types [64] and to-
497 gether with *FSTL1* is predicted to be driver for basal-like breast cancer by DMA. These
498 two genes are involved in angiogenesis, one of the cancer hallmarks [65–67], which
499 is also prognostic indicators of survival in breast cancer [68,69]. Additionally, for lung
500 adenocarcinoma, key enriched terms for both DMA and GMA predicted driver genes
501 included G2-M checkpoint, E2F targets and mTORC1 signaling (Figs 3B, 7B), sug-
502 gesting that the two mechanistic indicators identify at least partially overlapping bio-
503 logical processes. These processes are all important in cancer progression or metas-
504 tasis [70–73].
505 Finally, we also performed gene enrichment analysis of the driver genes identified by
506 both DMA and GMA. An enrichment analysis of the 141 overlapping driver genes be-
507 tween GMA and DMA in lung adenocarcinoma revealed E2F targets, G2-M checkpoint
508 and mTORC1 signaling to again be the most significant (Fig 7C), covering a vast ma-
509 jority of the overlapping genes. A similar enrichment analysis of the 13 overlapping
510 driver genes between GMA and DMA in basal-like breast cancer revealed no signifi-
511 cantly enriched terms.

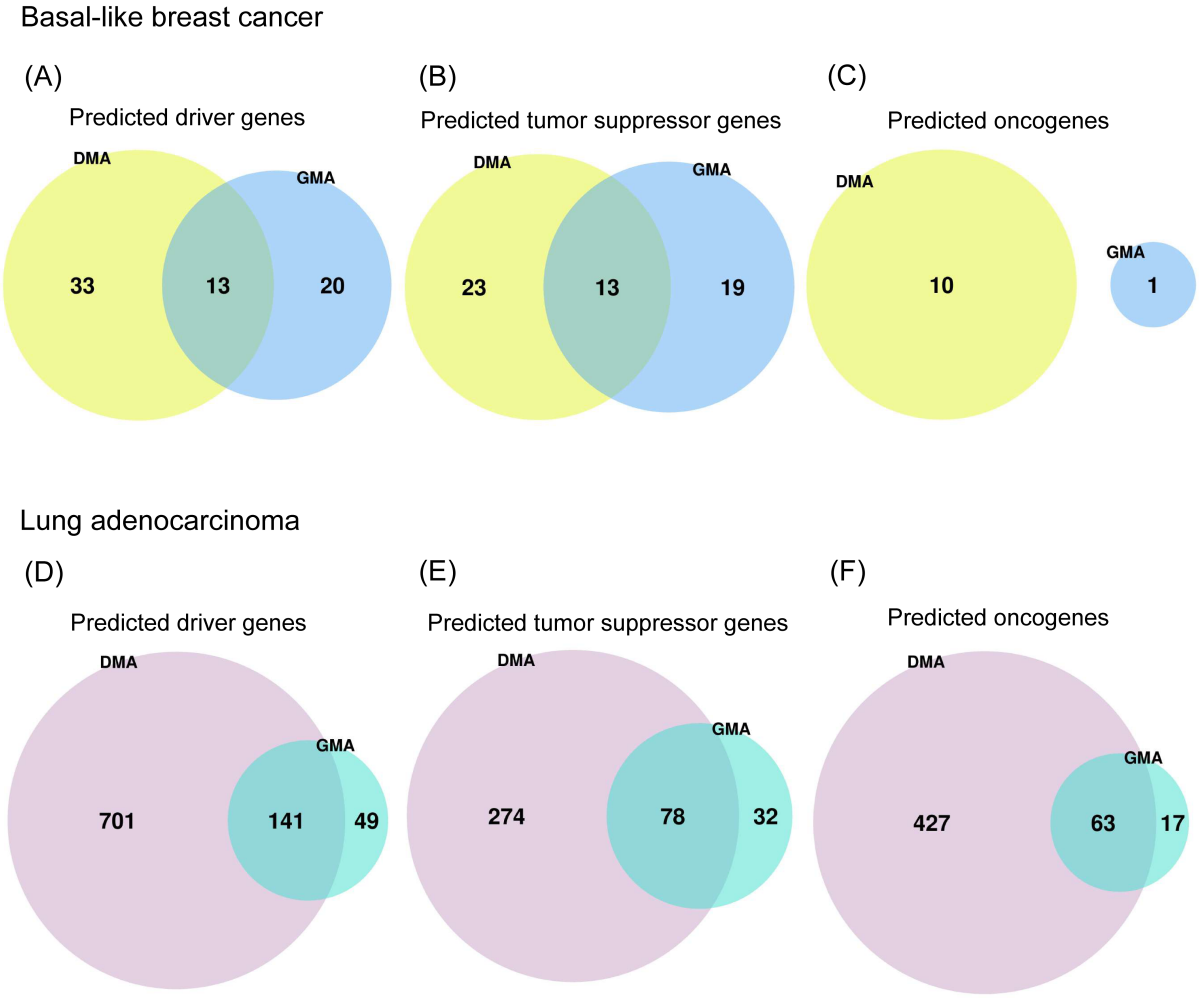

Fig 6. Comparison of number of mutation- and methylation-driven driver genes. Venn diagram comparing **(A, D)** the number of driver genes, **(B, E)** TSGs, and **(C, F)** OCGs predicted by the Driver Mutation Analysis (DMA) and Gene Methylation Analysis (GMA) functions of Moonlight2 for **(A-C)** basal-like breast cancer and **(D-F)** lung adenocarcinoma.

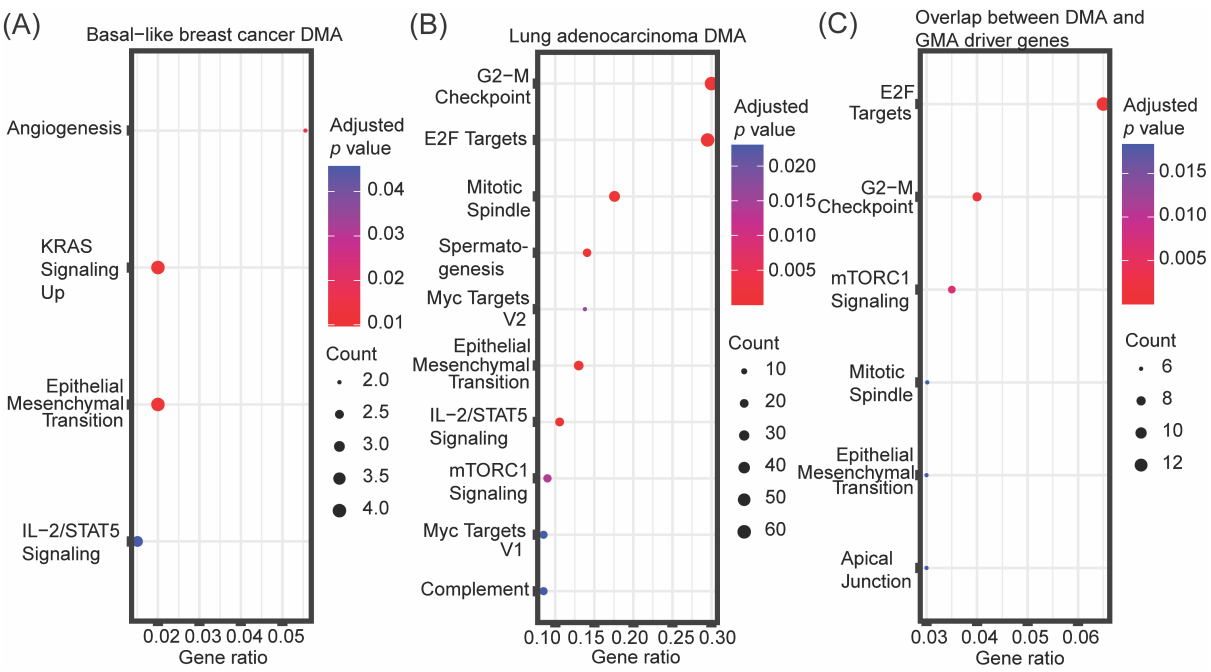

Fig 7. Enrichment analysis of predicted mutation-driven driver genes. Enrichment analysis of **(A)** mutation-driven driver genes predicted by Driver Mutation Analysis (DMA) in basal-like breast cancer, **(B)** mutation-driven driver genes predicted by Driver Mutation Analysis (DMA) in lung adenocarcinoma, and **(C)** driver genes predicted by both DMA and GMA in lung adenocarcinoma. The "MSigDB Hallmark 2020" database was used for the enrichment analyses. 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.

Availability and future directions

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/. Example data and vignette are available in the Moonlight2 R package. In the future, we envision incorporation of additional secondary -omics layers such as chromatin accessibility and copy number variation. Moreover, we would like to implement proteomics and single-cell RNA sequencing data as additional input data types.

Acknowledgements

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.

Author contributions

Conceptualization: Mona Nourbakhsh, Yuanning Zheng, Matteo Tiberti, Olivier Gevaert, Elena Papaleo

Data curation: Mona Nourbakhsh, Hongjin Chen, Subhayan Akhuli, Matteo Tiberti

Formal analysis: Mona Nourbakhsh, Yuanning Zheng, Humaira Noor, Hongjin Chen, Subhayan Akhuli, Matteo Tiberti, Olivier Gevaert, Elena Papaleo

Funding acquisition: Olivier Gevaert, Elena Papaleo

Investigation: Mona Nourbakhsh, Hongjin Chen, Subhayan Akhuli, Elena Papaleo

Methodology: Mona Nourbakhsh, Yuanning Zheng, Humaira Noor, Matteo Tiberti, Olivier Gevaert, Elena Papaleo

Project administration: Elena Papaleo

Resources: Elena Papaleo

Software: Mona Nourbakhsh, Hongjin Chen, Subhayan Akhuli

Supervision: Mona Nourbakhsh, Yuanning Zheng, Humaira Noor, Matteo Tiberti, Olivier Gevaert, Elena Papaleo

Validation: Mona Nourbakhsh, Matteo Tiberti

Visualization: Mona Nourbakhsh, Hongjin Chen, Subhayan Akhuli

Writing - original draft preparation: Mona Nourbakhsh, Hongjin Chen, Subhayan Akhuli, Elena Papaleo

Writing - review & editing: all the authors

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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. doi:10.3322/caac.21660
2. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458: 719–724. doi:10.1038/nature07943
3. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science. 2013;340: 1546–1558. doi:10.1126/science.1235122
4. Shen L, Shi Q, Wang W. Double agents: Genes with both oncogenic and tumor-suppressor functions. Oncogenesis. 2018;7: 1–14. doi:10.1038/s41389-018-0034-x
5. Datta N, Chakraborty S, Basu M, Ghosh MK. Tumor suppressors having oncogenic functions: The double agents. Cells. 2021;10: 1–26. doi:10.3390/cells10010046
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. doi:10.1016/j.cell.2011.02.013
8. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12: 31–46. doi:10.1158/2159-8290.CD-21-1059

602 9. Nourbakhsh M, Degn K, Saksager AB, Tiberti M, Papaleo E. Prediction of cancer
603 driver genes and mutations: the potential of integrative computational frameworks.
604 Brief Bioinform. 2024;25: 1–16. doi:10.1093/bib/bbad519
605 10. Konda P, Garinet S, Van Allen EM, Viswanathan SR. Genome-guided discovery of
606 cancer therapeutic targets. Cell Rep. 2023;42: 1–18.
607 doi:10.1016/j.celrep.2023.112978
608 11. Liu Y, Hu X, Han C, Wang L, Zhang X, He X, et al. Targeting tumor suppressor genes
609 for cancer therapy. BioEssays. 2015;37: 1277–1286. doi:10.1002/bies.201500093
610 12. Yu X, Zhao H, Wang R, Chen Y, Ouyang X, Li W, et al. Cancer epigenetics: from la-
611 boratory studies and clinical trials to precision medicine. Cell Death Discov. 2024;10:
612 1–12. doi:10.1038/s41420-024-01803-z
613 13. Iannuccelli M, Micarelli E, Surdo P Lo, Palma A, Perfetto L, Rozzo I, et al. Can-
614 cerGeneNet: Linking driver genes to cancer hallmarks. Nucleic Acids Res. 2020;48:
615 D416–D421. doi:10.1093/nar/gkz871
616 14. Colaprico A, Olsen C, Bailey MH, Odom GJ, Terkelsen T, Silva TC, et al. Interpreting
617 pathways to discover cancer driver genes with Moonlight. Nat Commun. 2020;11: 1–
618 17. doi:10.1038/s41467-019-13803-0
619 15. Nourbakhsh M, Saksager A, Tom N, Chen XS, Colaprico A, Olsen C, et al. A workflow
620 to study mechanistic indicators for driver gene prediction with Moonlight. Brief Bioin-
621 form. 2023;24: 1–13. doi:10.1093/bib/bbad274
622 16. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging land-
623 scape of oncogenic signatures across human cancers. Nat Genet. 2013;45: 1127–
624 1133. doi:10.1038/ng.2762
625 17. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009;1: 239–259.
626 doi:10.2217/EPI.09.33
627 18. Esteller M. Epigenetics in Cancer. N Engl J Med. 2008;358: 1148–1159.
628 doi:10.1056/NEJMra072067
629 19. Lakshminarasimhan R, Liang G. The role of DNA methylation in cancer. Adv Exp Med
630 Biol. 2016;945: 151–172. doi:10.1007/978-3-319-43624-1_7
631 20. Søes S, Daugaard IL, Singers Sørensen B, Carus A, Mattheisen M, Alsner J, et al.
632 Hypomethylation and increased expression of the putative oncogene ELMO3 are as-
633 sociated with lung cancer development and metastases formation. Oncoscience.
634 2014;1: 367–374. doi:10.18632/oncoscience.42
635 21. Zheng Y, Jun J, Brennan K, Gevaert O. EpiMix is an integrative tool for epigenomic
636 subtyping using DNA methylation. Cell Reports Methods. 2023;3: 1–23.
637 doi:10.1016/j.crmeth.2023.100515
638 22. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): An
639 immeasurable source of knowledge. Contemp Oncol. 2015;19: A68–A77.
640 doi:10.5114/wo.2014.47136
641 23. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond
642 Its Data. Cell. 2018;173: 283–285. doi:10.1016/j.cell.2018.03.042
643 24. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharma-
644 cology. 2013;38: 23–38. doi:10.1038/npp.2012.112
645 25. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Can-
646 cer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev
647 Cancer. 2018;18: 696–705. doi:10.1038/s41568-018-0060-1

26. 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. doi:10.1016/j.compbiomed.2019.103362
27. Akhavan-Safar M, Teimourpour B. KatzDriver: A network based method to cancer causal genes discovery in gene regulatory network. BioSystems. 2021;201: 1–13. doi:10.1016/j.biosystems.2020.104326
28. Wei PJ, Zhang D, Li HT, Xia J, Zheng CH. DriverFinder: A gene length-based network method to identify cancer driver genes. Complexity. 2017;2017: 1–11. doi:10.1155/2017/4826206
29. Pham VVH, Liu L, Bracken CP, Goodall GJ, Long Q, Li J, et al. CBNA: A control theory based method for identifying coding and non-coding cancer drivers. PLoS Comput Biol. 2019;15: 23. doi:10.1371/journal.pcbi.1007538
30. Dinstag G, Shamir R. PRODIGY: Personalized prioritization of driver genes. Bioinformatics. 2020;36: 1831–1839. doi:10.1093/bioinformatics/btz815
31. Wei PJ, Zhang D, Xia J, Zheng CH. LNDriver: Identifying driver genes by integrating mutation and expression data based on gene-gene interaction network. BMC Bioinformatics. 2016;17: 1–10. doi:10.1186/s12859-016-1332-y
32. Zhang D, Bin Y. DriverSubNet: A Novel Algorithm for Identifying Cancer Driver Genes by Subnetwork Enrichment Analysis. Front Genet. 2021;11: 1–10. doi:10.3389/fgene.2020.607798
33. Li A, Chapuy B, Varelas X, Sebastiani P, Monti S. Identification of candidate cancer drivers by integrative Epi-DNA and Gene Expression (iEDGE) data analysis. Sci Rep. 2019;9: 1–12. doi:10.1038/s41598-019-52886-z
34. Han Y, Yang J, Qian X, Cheng WC, Liu SH, Hua X, et al. DriverML: A machine learning algorithm for identifying driver genes in cancer sequencing studies. Nucleic Acids Res. 2019;47: 1–12. doi:10.1093/nar/gkz096
35. Gu H, Xu X, Qin P, Wang J. FI-Net: Identification of Cancer Driver Genes by Using Functional Impact Prediction Neural Network. Front Genet. 2020;11: 1–15. doi:10.3389/fgene.2020.564839
36. Hou Y, Gao B, Li G, Su Z. MaxMIF: A New Method for Identifying Cancer Driver Genes through Effective Data Integration. Advanced Science. 2018;5: 1–9. doi:10.1002/advs.201800640
37. Zapata L, Susak H, Drechsel O, Friedländer MR, Estivill X, Ossowski S. Signatures of positive selection reveal a universal role of chromatin modifiers as cancer driver genes. Sci Rep. 2017;7: 1–15. doi:10.1038/s41598-017-12888-1
38. Terekhanova N V., Karpova A, Liang WW, Strzalkowski A, Chen S, Li Y, et al. Epigenetic regulation during cancer transitions across 11 tumour types. Nature. 2023;623: 432–441. doi:10.1038/s41586-023-06682-5
39. D'Anello L, Sansone P, Storci G, Mitrugno V, D'Uva G, Chieco P, et al. Epigenetic control of the basal-like gene expression profile via Interleukin-6 in breast cancer cells. Mol Cancer. 2010;9: 1–13. doi:10.1186/1476-4598-9-300
40. 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. doi:10.1186/s12929-020-00632-3
41. Skrypek N, Goossens S, De Smedt E, Vandamme N, Berx G. Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity. Trends in Genetics. 2017;33: 943–959. doi:10.1016/j.tig.2017.08.004

42. Liu QL, Luo M, Huang C, Chen HN, Zhou ZG. Epigenetic Regulation of Epithelial to Mesenchymal Transition in the Cancer Metastatic Cascade: Implications for Cancer Therapy. Front Oncol. 2021;11: 1–13. doi:10.3389/fonc.2021.657546
43. Lu W, Kang Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. Dev Cell. 2019;49: 361–374. doi:10.1016/j.devcel.2019.04.010
44. Yang H, Gan L, Chen R, Li D, Zhang J, Wang Z. From multi-omics data to the cancer druggable gene discovery: A novel machine learning-based approach. Brief Bioinform. 2023;24: 1–11. doi:10.1093/bib/bbac528
45. Yoshimaru T, Nakamura Y, Katagiri T. Functional genomics for breast cancer drug target discovery. J Hum Genet. 2021;66: 927–935. doi:10.1038/s10038-021-00962-6
46. Zsákai L, Sipos A, Dobos J, Erős D, Szántai-Kis C, Bánhegyi P, et al. Targeted drug combination therapy design based on driver genes. Oncotarget. 2019;10: 5255–5266. doi:10.18632/oncotarget.26985
47. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, 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. doi:10.1093/nar/gkad1040
48. Souglakos J, Boukovinas I, Taron M, Mendez P, Mavroudis D, Tripaki 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. doi:10.1038/sj.bjc.6604344
49. Atwell B, Chen CY, Christofferson M, Montfort WR, Schroeder J. Sorting nexin-dependent therapeutic targeting of oncogenic epidermal growth factor receptor. Cancer Gene Ther. 2023;30: 267–276. doi:10.1038/s41417-022-00541-7
50. Robichaux JP, Le X, Vijayan RSK, Hicks JK, Heeke S, Elamin YY, et al. Structure-based classification predicts drug response in EGFR-mutant NSCLC. Nature. 2021;597: 732–737. doi:10.1038/s41586-021-03898-1
51. Robichaux JP, Elamin YY, Tan Z, Carter BW, Zhang S, Liu S, 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. doi:10.1038/s41591-018-0007-9
52. 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. doi:10.1038/s41573-021-00195-4
53. Poels KE, Schoenfeld AJ, Makhnin A, Tobi Y, Wang Y, Frisco-Cabanos H, 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. doi:10.1038/s41467-021-23912-4
54. van Alderwerelt van Rosenburgh IK, Lu DM, Grant MJ, Stayrook SE, Phadke M, Walther Z, et al. Biochemical and structural basis for differential inhibitor sensitivity of EGFR with distinct exon 19 mutations. Nat Commun. 2022;13: 1–16. doi:10.1038/s41467-022-34398-z
55. Pu X, Zhou Y, Kong Y, Chen B, Yang A, Li J, 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. doi:10.1186/s12885-023-11465-2
56. Momeny M, Zarrinrad G, Moghaddaskho F, Poursheikhani A, Sankanian G, Zaghal A, 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. doi:10.1038/s41598-017-04147-0

57. Brischwein K, Schlereth B, Guller B, Steiger C, Wolf A, Lutterbuese R, et al. MT110: A novel bispecific single-chain antibody construct with high efficacy in eradicating established tumors. Mol Immunol. 2006;43: 1129–1143. doi:10.1016/j.molimm.2005.07.034
58. Ferrari F, Bellone S, Black J, Schwab CL, Lopez S, Cocco E, 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. doi:10.1186/s13046-015-0241-7
59. Keller L, Werner S, Pantel K. Biology and clinical relevance of EpCAM. Cell Stress. 2019;3: 165–180. doi:10.15698/cst2019.06.188
60. Cui Y, Li J, Liu X, Gu L, Lyu M, Zhou J, 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. doi:10.3390/cancers14174121
61. Kebenko M, Goebeler ME, Wolf M, Hasenburg A, Seggewiss-Bernhardt R, Ritter B, 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. doi:10.1080/2162402X.2018.1450710
62. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007;446: 153–158. doi:10.1038/nature05610
63. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. 2017;9: 1–14. doi:10.1186/s13073-017-0424-2
64. Chen ZY, Gao H, Dong Z, Shen YN, Wang ZC, Wei W, et al. NRP1 regulates radiation-induced EMT via TGF-β/Smad signaling in lung adenocarcinoma cells. Int J Radiat Biol. 2020;96: 1281–1295. doi:10.1080/09553002.2020.1793015
65. Madu CO, Wang S, Madu CO, Lu Y. Angiogenesis in breast cancer progression, diagnosis, and treatment. J Cancer. 2020;11: 4474–4494. doi:10.7150/jca.44313
66. Teleanu RI, Chircov C, Grumezescu AM, Teleanu DM. Tumor angiogenesis and anti-angiogenic strategies for cancer treatment. J Clin Med. 2020;9: 1–21. doi:10.3390/jcm9010084
67. Badodekar N, Sharma A, Patil V, Telang G, Sharma R, Patil S, et al. Angiogenesis induction in breast cancer: A paracrine paradigm. Cell Biochem Funct. 2021;39: 860–873. doi:10.1002/cbf.3663
68. Vartanian RK, Weidner N. Correlation of Intratumoral Endothelial Cell Proliferation with Microvessel Density (Tumor Angiogenesis) and Tumor Cell Proliferation in Breast Carcinoma. Breast CA Endothelial Cell Proliferation. 1994;144: 1188–1194.
69. Weidner N, Semple JP, Welch WR, Folkman J. Tumor angiogenesis and metastasis - correlation in invasive breast carcinoma. N Engl J Med. 1991;324: 1–8. doi:10.1056/NEJM199101033240101
70. Gargalionis AN, Papavassiliou KA, Papavassiliou AG. Implication of mTOR Signaling in NSCLC: Mechanisms and Therapeutic Perspectives. Cells. 2023;12: 1–10. doi:10.3390/cells12152014
71. Nevins JR. The Rb/E2F pathway and cancer. Hum Mol Genet. 2001;10: 699–703.
72. Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer. Nat Rev Cancer. 2019;19: 326–338. doi:10.1038/s41568-019-0143-7
73. Stark GR, Taylor WR. Analyzing the G2/M Checkpoint. Checkpoint Controls and Cancer. New Jersey: Humana Press; 2004. pp. 051–082. doi:10.1385/1-59259-788-2:051

Supporting information

S1 Fig. Integration of Moonlight and EpiMix for prediction of cancer driver genes. (A) Heatmap showing number of differentially methylated CpGs and classifications of methylation status in the oncogenic mediators in lung adenocarcinoma. The heatmap was generated using the plotGMA function. **(B)** Venn diagram comparing oncogenic mediators predicted from Moonlight's primary layer with functional genes predicted from EpiMix in lung adenocarcinoma. 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. **(C)** Heatmap showing the effect of the predicted driver genes in lung adenocarcinoma 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. **(D)** Heatmap showing number of differentially methylated CpGs and classifications of methylation status in the oncogenic mediators in thyroid carcinoma. The heatmap was generated using the plotGMA function. **(E)** Venn diagram comparing oncogenic mediators predicted from Moonlight's primary layer with functional genes predicted from EpiMix in thyroid carcinoma. 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. **(F)** Heatmap showing the effect of the predicted driver genes in thyroid carcinoma 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.
(PDF)

S2 Fig. Number of driver gene-drug interactions. Number of driver gene-drug interactions in **(A)** basal-like breast cancer, **(B)** lung adenocarcinoma, and **(C)** thyroid carcinoma found by querying DGIdb. The driver genes are stratified into OCGs and TSGs. The number of drug interactions is shown on the *x* axis and the driver genes are shown on the *y* axis.
(PDF)

S1 Text. Evidence integration for defining driver genes.
(PDF)

S2 Text. Methods of case study: Prediction of driver genes with differential methylation in basal-like breast cancer, lung adenocarcinoma, and thyroid carcinoma using Moonlight.
(PDF)