

## ORIGINAL RESEARCH

# Integrated transcriptomics reveals master regulators of lung adenocarcinoma and novel repositioning of drug candidates

Marco Antônio De Bastiani<sup>1,2</sup>  | Fábio Klamt<sup>1,2</sup> 

<sup>1</sup>Laboratory of Cellular Biochemistry, Department of Biochemistry, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

<sup>2</sup>National Institute of Science and Technology for Translational Medicine (INCT-TM), Porto Alegre, RS, Brazil

**Correspondence**

Marco Antônio De Bastiani, Department of Biochemistry, UFRGS, Rua Ramiro Barcelos 2600, Porto Alegre (RS), Brazil 90035-003.

Email: marco.bastiani@ufrgs.br or tyrev@hotmail.com

**Funding information**

This study was supported by the Brazilian funds MCTI/CNPq INCTTM/CAPES/FAPESP (465458/2014-9), CNPq/MS/SCTIE/DECIT (466989/2014-8), and PRONEX/FAPERGS (16/2551-0000499-4). FK received a fellowship from MCT/CNPq (306439/2014-0).

**Abstract**

**Background:** Lung adenocarcinoma is the major cause of cancer-related deaths in the world. Given this, the importance of research on its pathophysiology and therapy remains a key health issue. To assist in this endeavor, recent oncology studies are adopting Systems Biology approaches and bioinformatics to analyze and understand *omics* data, bringing new insights about this disease and its treatment.

**Methods:** We used reverse engineering of transcriptomic data to reconstruct non-tumorous lung reference networks, focusing on transcription factors (TFs) and their inferred target genes, referred as regulatory units or regulons. Afterwards, we used 13 case-control studies to identify TFs acting as master regulators of the disease and their regulatory units. Furthermore, the inferred activation patterns of regulons were used to evaluate patient *survival* and search drug candidates for repositioning.

**Results:** The regulatory units under the influence of *ATOH8*, *DACHI*, *EPAS1*, *ETV5*, *FOXA2*, *FOXMI*, *HOXA4*, *SMAD6*, and *UHRF1* transcription factors were consistently associated with the pathological phenotype, suggesting that they may be master regulators of lung adenocarcinoma. We also observed that the inferred activity of *FOXA2*, *FOXMI*, and *UHRF1* was significantly associated with risk of death in patients. Finally, we obtained depropine, promazine, valproic acid, azacyclonol, methotrexate, and ChemBridge ID compound 5109870 as potential candidates to revert the molecular profile leading to decreased survival.

**Conclusion:** Using an integrated transcriptomics approach, we identified master regulator candidates involved with the development and prognostic of lung adenocarcinoma, as well as potential drugs for repurposing.

**KEYWORDS**

computational drug repositioning, connectivity map, lung cancer, master regulator, transcriptomic

## 1 | INTRODUCTION

Historically, lung cancer is one of the most insidious and lethal menaces of oncology, remaining the major cause of cancer

mortality among men and women for the past decades.<sup>1</sup> The two main types of lung cancer are small cell lung cancer (SCLC) and nonsmall cell lung cancer (NSCLC), the latter accounting for ~85% of diagnosed cases.<sup>2</sup> NSCLC collectively includes

numerous epithelial-derived tumors, of which adenocarcinoma and squamous cell carcinoma are the most frequent. Moreover, lung adenocarcinoma is the most common subtype in never smokers and its incidence is increasing steadily in the United States, Europe, and East Asia, including Japan.<sup>3,4</sup> Appropriate identification of histologic types is important because it affects prognosis, therapy selection, and considerations for molecular testing.<sup>5</sup> Therefore, type-specific molecular research is a reasonable standpoint to translate the biological diversity of lung cancer into clinically relevant subpopulation selection and subsequent customized therapy guidelines.

Although the search and study of single genes associated with cancer have been a commonplace in oncology research for many years, recent systems biology views emerged to introduce new complexity and understanding to the scenario.<sup>6</sup> In this context, methods evaluating RNA expression levels, such as microarray and RNA-seq, have become valuable instruments to assess the dynamic properties of biological systems in a fast, broad, and reliable way. Besides the technological aspect, the development of bioinformatics analytical methods and tools has also aided enormously our capacity to extract knowledge from *omics* technologies.<sup>7</sup>

One important area of systems biology involves the identification and understanding of gene regulatory networks (GRNs). For this, a commonly used approach is the so-called reverse engineering, a process of revealing the network structure of a biological system by reasoning backward from observed data.<sup>8</sup> The concept of master regulators (MRs) is inserted into the evaluation of GRNs and used to describe elements situated in higher positions of the biological network hierarchy, participating in the specification of cellular lineages by regulating multiple network elements, ultimately controlling the ability to specify/respecify cellular outcome.<sup>9</sup> In accordance with this concept of MR by default, we have the physiological role of transcription factors (TFs), which enable a relatively small number of molecules to generate a large diversity of cell types and phenotypic states.<sup>10,11</sup>

Additionally, following the momentum of big data basic research, clinical and translational communities are also incorporating this paradigm into their view of patient treatment and therapy. Recently, pharmacology research is attempting to break the “one disease, one target, one drug” model of drug discovery through the incorporation of a new archetype of drug research, namely systems pharmacology.<sup>12,13</sup> This approach merges systems view, drug repositioning, and bioinformatics to explore the possibility that drugs can alter the expression profiles of pathological network modules of complex biological systems. In a network, modules represent highly interconnected local regions, such as biochemical pathways for example.<sup>6</sup> Accordingly, modules-based drug repositioning has been proposed to retrieve drug candidates in breast cancer<sup>14</sup> and colorectal adenocarcinoma.<sup>15</sup> Both these approaches employed the molecular Connectivity Map (CMap) strategy of drug repurposing, which recognizes that biological elements

have several interdependencies and that attempts to disregard such a notion using single-targeted interventions are probably futile. Thus, the proposition is to modify the entire state of the pathological system toward the physiological scenario, through modulation of many targets simultaneously.<sup>16</sup>

In this study, 13 transcriptomic datasets were employed to identify potential MRs of lung adenocarcinoma using the ARACNe algorithm and GRNs centered on TFs and their targets. The inferred targets of nine MR candidates—*ATO8*, *DACHI*, *EPAS1*, *ETV5*, *FOXA2*, *FOXMI*, *HOXA4*, *SMAD6*, and *UHRF1*—were consistently enriched with differentially expressed genes overrepresented in the pathological phenotype. Furthermore, the inferred activity of *FOXA2*, *FOXMI*, and *UHRF1* were significantly associated with survival in several lung adenocarcinoma cohorts. Finally, we used a module-oriented CMap approach to query drugs able to revert the pathological gene expression profiles of these regulatory units. This strategy retrieved six repositioning candidate molecules—depropine, promazine, valproic acid, azacyclonol, methotrexate, and ChemBridge ID compound 5109870.

## 2 | METHODS

### 2.1 | High-throughput data acquisition

All expression profiles used in this study were obtained from Gene Expression Omnibus (GEO) database using the GEOquery package (version 2.48.0).<sup>17</sup> Available processed expression data were transformed to logarithmic scale prior to further analyses, when necessary. Two nontumorous lung datasets were used to infer the reference transcriptional networks centered on transcription factors. Thirteen datasets of case-control (cancer vs unaffected lung tissue) were acquired to compute altered regulatory units based on the reference networks. Finally, 11 different datasets of lung adenocarcinoma were employed to evaluate the survival associations of patients based on altered regulatory units. Details and summary descriptions of the expression data acquired are presented in Table 1.

### 2.2 | Reference transcriptional network inference

The transcriptional networks (TNs) centered on TFs and their predicted target genes were inferred using nontumorous lung tissue. Herein, the reference transcriptional network inferred using GSE23546 will be referred to as TN1 and the one using GSE71181 as TN2 (Table 1). In this study, the term *regulatory unit* or *regulon* is used to describe the groups of inferred genes and their associated TFs. Transcriptional networks were computed using the RTN package (version 2.4.6)<sup>18,19</sup> employing mutual information-based Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) method.<sup>20</sup> Interactions were evaluated using a MI threshold cutoff by permutation and

**TABLE 1** Gene expression datasets employed to retrieve master regulators of lung adenocarcinoma

Purpose	GEOID	Sample Size	Method
Nontumorous lung transcriptional network reconstruction (TN1)	GSE23546	904 (Laval+UBC)	Microarray
Nontumorous lung transcriptional network reconstruction (TN2)	GSE71181	284	Microarray
Master regulator inferences (case vs control studies)	GSE10072	104	Microarray
	GSE11969	95	Microarray
	GSE21933	32	Microarray
	GSE27262	50	Microarray
	GSE31552	47	Microarray
	GSE32665	179	Microarray
	GSE32863	116	Microarray
	GSE40275	84	Microarray
	GSE43458	110	Microarray
	GSE62113	25	Microarray
	GSE74706	28	Microarray
	GSE102511	31	RNA-seq
	GSE87340	27	RNA-seq
Survival	GSE14814	71	Microarray
	GSE26939	116	Microarray
	GSE29013	55	Microarray
	GSE11969	95	Microarray
	GSE37745	106	Microarray
	GSE41271	178	Microarray
	GSE42127	132	Microarray
	GSE50081	127	Microarray
	GSE87340	27	RNA-seq
TCGA	594	RNA-seq	

by bootstrap, creating a consensus network. Additionally, a final step employs a Data Processing Inequality algorithm with null tolerance to eliminate associations likely mediated by another TF. Our network reconstruction parameters were 1000 permutations, with a  $P$ -value cutoff of .001, and 100 bootstraps (all remaining parameter were kept default).

### 2.3 | Master regulator inference and two-tailed gene set enrichment analysis

For the MR inference, we employed the master regulator analysis (MRA) described by Carro and colleagues<sup>21</sup> to the regulatory units of each reference TN. The algorithm computes the statistical overrepresentation (enrichment) of differentially expressed genes (DEG), obtained from differential expression analysis, in the regulatory units of TN1 and TN2. Our criteria of differential expression were false discovery rate (FDR)-adjusted  $P$ -value  $<.05$  and absolute log fold change ( $\log_{2}FC$ )  $> 1$ . Regulons with statistical enrichment of DEG and with  $>50$  elements (regulon size  $>50$ ), in  $\geq 80\%$  of

the queried case-control studies, were considered altered in the lung adenocarcinoma, thus considered potential master regulators of the disease.

Two-tailed GSEA was also performed using the RTN package (version 2.4.6), with a  $P$ -value cutoff set to .05 and using 1000 permutations. Briefly, Pearson's correlation was used to split the regulatory units into two subgroups: positively associated targets (A) and negatively associated targets (B). Afterward, the phenotype association of each subgroup was tested by the GSEA<sup>22</sup> statistics, resulting in independent enrichment scores (ES) for each subgroup. An additional step was carried out to test the differential enrichment ( $ES_A - ES_B$ ), considering that a maximum deviation from zero near opposite extremes and a good separation of the two distributions are desirable for a clear association. Thus, a high negative differential score implies that the regulon is repressed in the disorder phenotype, while a high positive differential score indicates that the regulon is induced in the disorder phenotype.

The Bioconductor (version 3.7) package limma (version 3.36.5)<sup>17</sup> was employed to compute differential expression analysis of case-control studies and obtain DEG and  $\log_{2}FC$

for both MRA and GSEA inputs. Importantly, all analyses were implemented equally and independently to all case-control studies employed against each TN reference.

## 2.4 | Survival analyses

Multivariable Cox proportional hazards regressions were performed using *regulon's* ES, adjusted for clinical variables according to available information. Diagnostic analyses of the models were performed based on Martingale and weighted Schoenfeld residuals plots and tests<sup>23</sup>; clinical variables available were stratified to meet the proportional hazards assumption when necessary. The ES of queried regulatory units were firstly assessed in the studies selected for survival analyses (Table 1) through gene set variation analysis (GSVA) method.<sup>24</sup> This gene set enrichment method estimates group gene activity variation over a sample population in an unsupervised manner and provides the scores used to assess the influence of *regulons'* activity over the risk of death in patients. These scores were used to compute survival models for all selected studies in both TN references. Additionally, enrichment scores from all studies in a given TN reference were z-score standardized, merged and evaluated together for their association with risk. Finally, Kaplan-Meier curves were constructed by discretizing of ES distributions into three segments (low = lower quartile; mid = IQR; high = upper quartile). Risk differences between groups were assessed using the log-rank test.

Similarly, for gene expression association with survival risk, we employed multivariable Cox proportional hazards regression using expression data of each selected study in both TN reference contexts.

Cox models and Kaplan-Meier curves were computed and graphed using *survival* (version 2.43-3) and/or *survminer* (version 0.4.3) packages.<sup>25,26</sup> Importantly, all analyses were implemented equally and independently to all survival studies employed against each TN reference.

## 2.5 | Computational drug repositioning using connectivity map

Altered regulatory units in lung adenocarcinoma were queried for drug candidates capable of reverting their expression profiles using a module-oriented adaptation of classical connectivity maps (CMap) method of drug repositioning.<sup>27,28</sup> Briefly, for each case-control study in both TN references, inferred genes of several *regulons* were selected, combined, and their logFC values submitted to the CMap analysis workflow using *PharmacoGx* package (version 1.10.3).<sup>29</sup>

## 2.6 | Statistical software

All analyses were computed in R statistical environment (version 3.5.2),<sup>30</sup> employing the above-mentioned packages

for each analysis. Networks were computed using RTN package and network figures were constructed using *RedeR* package (version 1.28.0).<sup>31</sup> Other plots were created with *ggplot2* package (version 3.1.0).<sup>32</sup>

## 3 | RESULTS

### 3.1 | Human lung regulatory network reconstruction

Two large public transcriptional datasets of nontumorous lung tissue obtained from GEO were used for the regulatory network reconstruction. Our approach assumes that transcription factors' activities are valuable agents of phenotypic difference or modulation, and that their effects can be assessed through fluctuations in the expression of their target genes. Thus, the goal of the network reconstruction is to infer target genes of TFs through their coassociation and encapsulate them into lists of genes, which will be addressed as *regulatory units* or *regulons*.

The resulted reference networks, herein named TN1 and TN2, derived from GSE23546 and GSE71181, respectively, constitute the starting inquiry point of our analyses. TN1 represents a superseries of 904 nontumorous lung samples, whereas TN2 represents 284 noninvolved lung parenchyma from adenocarcinoma patients. Transcripts were classified as transcription factors when annotated in the Gene Ontology with the identifier GO:0003700 (transcription factor activity, sequence-specific DNA binding). Following the networks' reconstruction, all subsequent analyses were implemented over both TN or their intersected results. Figure S1 summarizes the workflow employed in our study for master regulators of lung adenocarcinoma, their influence in patient survival and their potential to identify drug repositioning candidates.

### 3.2 | Inference and activity of lung adenocarcinoma master regulators

Nontumorous lung network reconstruction yielded sets of regulatory units centered on TFs for each TN as outputs. We next questioned which of these regulatory units were altered in lung adenocarcinoma. To address this issue, we collected 13 different transcriptomic datasets containing tumoral and nontumoral tissue information and used the MRA method proposed by Carro and collaborators<sup>21</sup> to infer consistently altered *regulons*. We found 19 *regulons* significantly enriched with differentially expressed genes (FDR-adjusted *P*-value <.05 and absolute logFC > 1) in ≥80% of the 13 studies for TN1 and 17 *regulons* for TN2. We termed the transcription factors modulating these regulatory units *common* MR of the disease. Figure 1A shows the subnetwork representation of these impaired

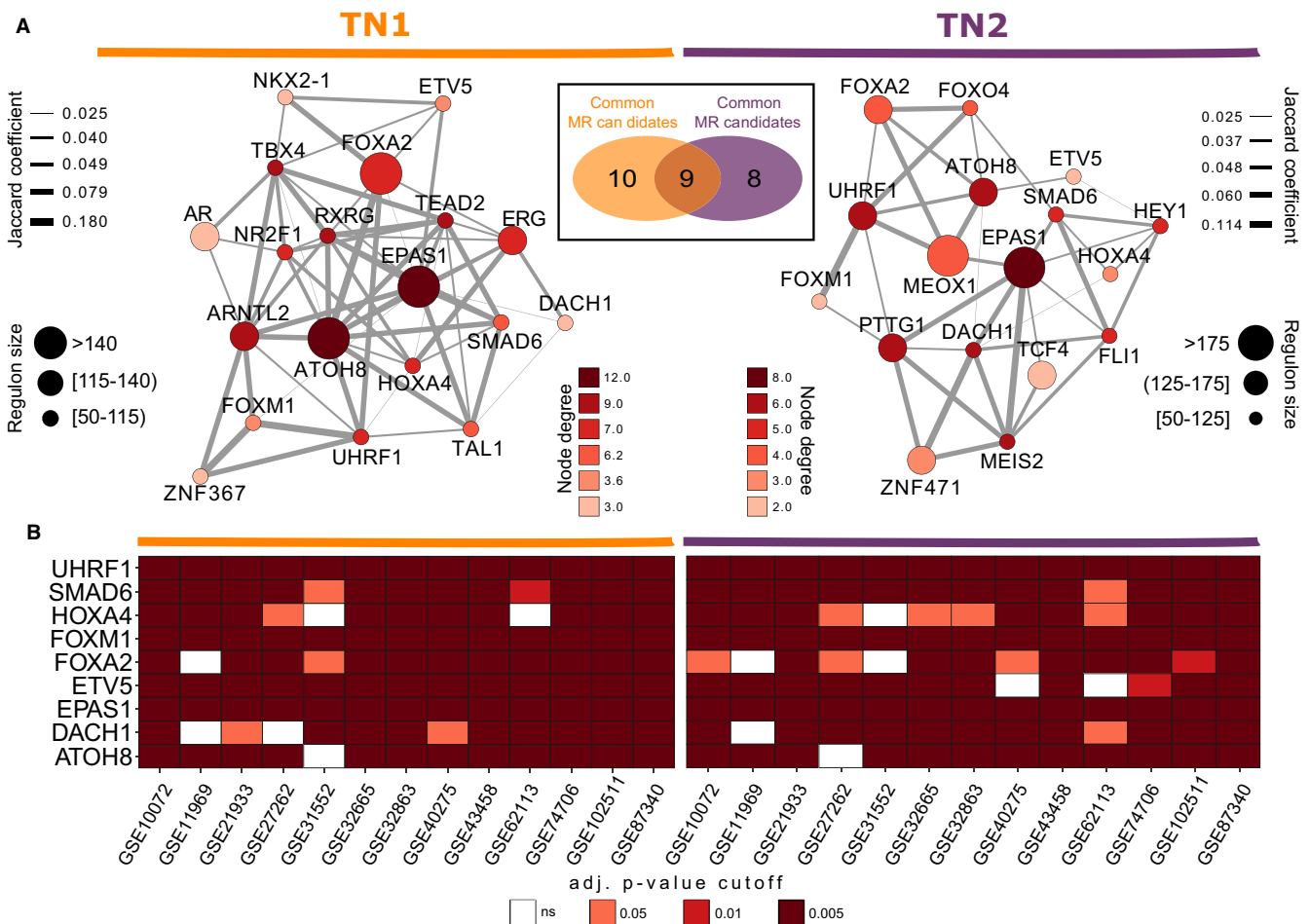
regulons and their associations with each other in both TNs. Table S4 shows additional information about the subnetwork structures, as well as TN1 and TN2 reconstruction overlaps. Additionally, we observed nine *common* MRs consistently dysregulated in both reference TN (Figure 1A *insert*), which were referred to as *consensus* MR. These *consensus* MRs were the regulons under the influence of ATOH8, DACH1, EPAS1, ETV5, FOXA2, FOXM1, HOXA4, SMAD6, and UHRF1 transcription factors (Figure 1B).

For each regulatory unit in both TNs, the network reconstruction also infers the transcription factors' modes of action on targets through their expression correlation. With this information, we queried the state of activation of the *consensus* MRs in TN1 and TN2 using a two-tailed variation of the GSEA method. Figure 2A shows the inferred activation states observed for the nine MR candidates in

both TNs. We observed repression of six MRs—*ATOH8*, *DACH1*, *EPAS1*, *ETV5*, *FOXA2*, *HOXA4*, and *SMAD6*—meaning that the negative targets of these transcription factors are upregulated in the disease phenotype, whereas their positive targets are downregulated (FDR-adjusted  $P$ -value  $< .05$ ). On the other hand, *FOXM1* and *UHRF1* were activated in lung adenocarcinoma, with increased expression of their positive targets and/or decreased expression of their negative targets.

### 3.3 | Master regulators of survival in lung adenocarcinoma

Besides investigating the association of transcription factors and their regulatory units with the diseased phenotype by identifying master regulator candidates, we asked how variations in the activity state of these regulons would affect patient outcome. To



**FIGURE 1** *Common and Consensus Master Regulator of Lung Adenocarcinoma*. A, Subnetwork of inferred *common* master regulators and their association for each reference transcription network. Node sizes map the number of genes inferred for a given transcription factor, the number of connections (degree) was mapped to the color of the nodes and edge widths represent the number of overlapped genes shared by pairs of TF. *Insert* shows a Venn diagram of the regulons in TN1 and TN2 networks (intersection regulons were termed *consensus* master regulators). B, Master regulator analysis showing the statistical overrepresentation (enrichment) of differentially expressed genes in each *consensus* master regulator, for all case-control datasets and for both reference TN. Our criteria of differential expression were false discovery rate (FDR)-adjusted  $P$ -value  $< .05$  and absolute log fold change ( $\log_{2}FC$ )  $> 1$

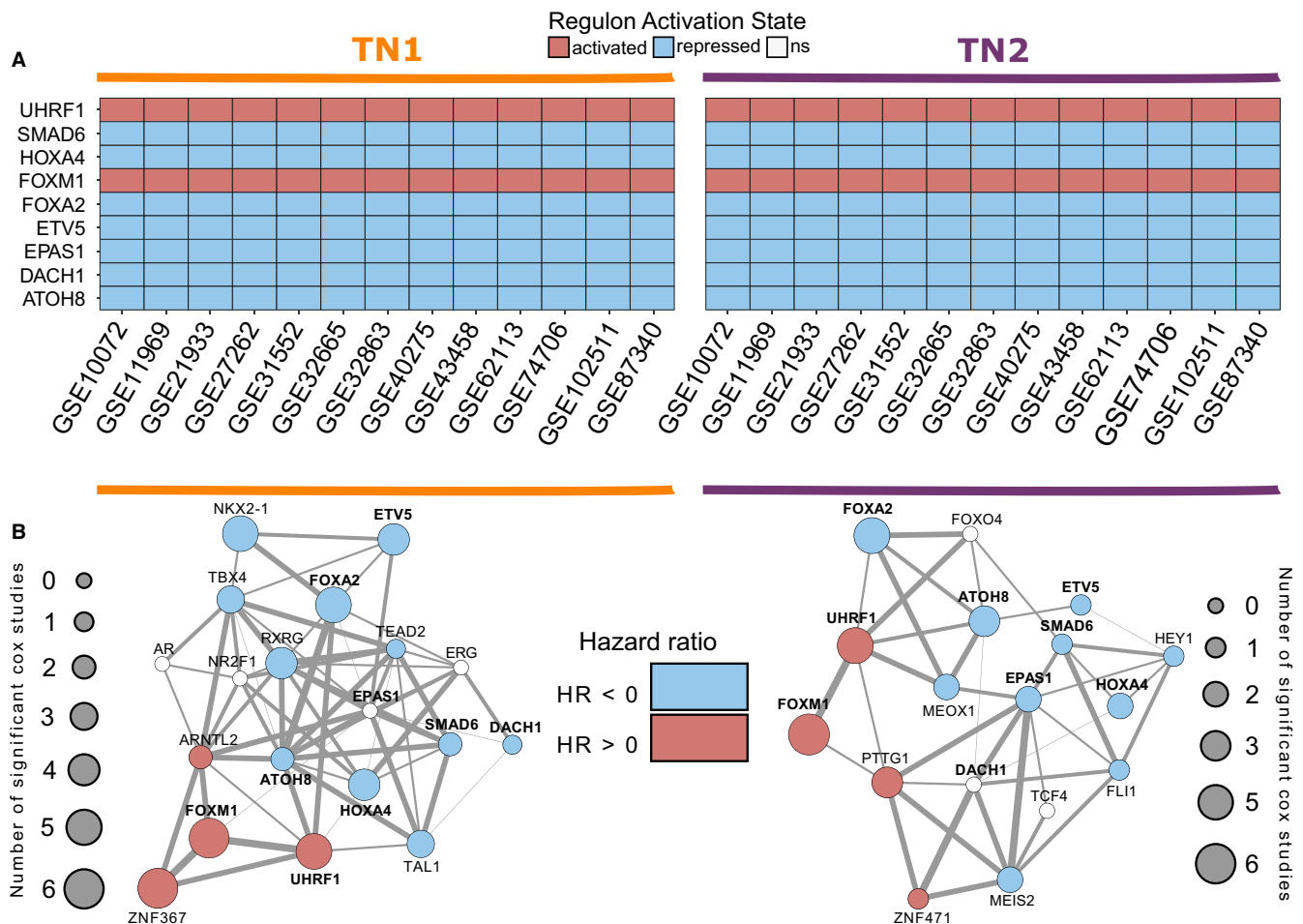


answer this, we used multiple Cox proportional hazards analyses of 10 different datasets of transcriptomic information (Table 1). Regulon activity was estimated in an unsupervised manner through gene set variant analysis (GSVA), providing ES of each patient with which survival models were fit.

Firstly, we modeled the influence of each *common* MR enrichment score on patient risk in all 10 studies from both reference TNs. Figure 2B shows the count result of these analyses in a survival network. Node size maps the number of transcriptomic datasets in which the *common* MR enrichment score showed significant association with patient risk, adjusted for the clinical variables available in each dataset. Node colors represent the direction of the hazard ratio (HR) for the studies showing significant survival risk. HR > 0 (pink nodes) means that increased ES (regulon activation) is associated with an increase in patient risk of death, whereas HR < 0 (blue nodes) means that increased ES is associated with a decrease in patient risk of death. White nodes map

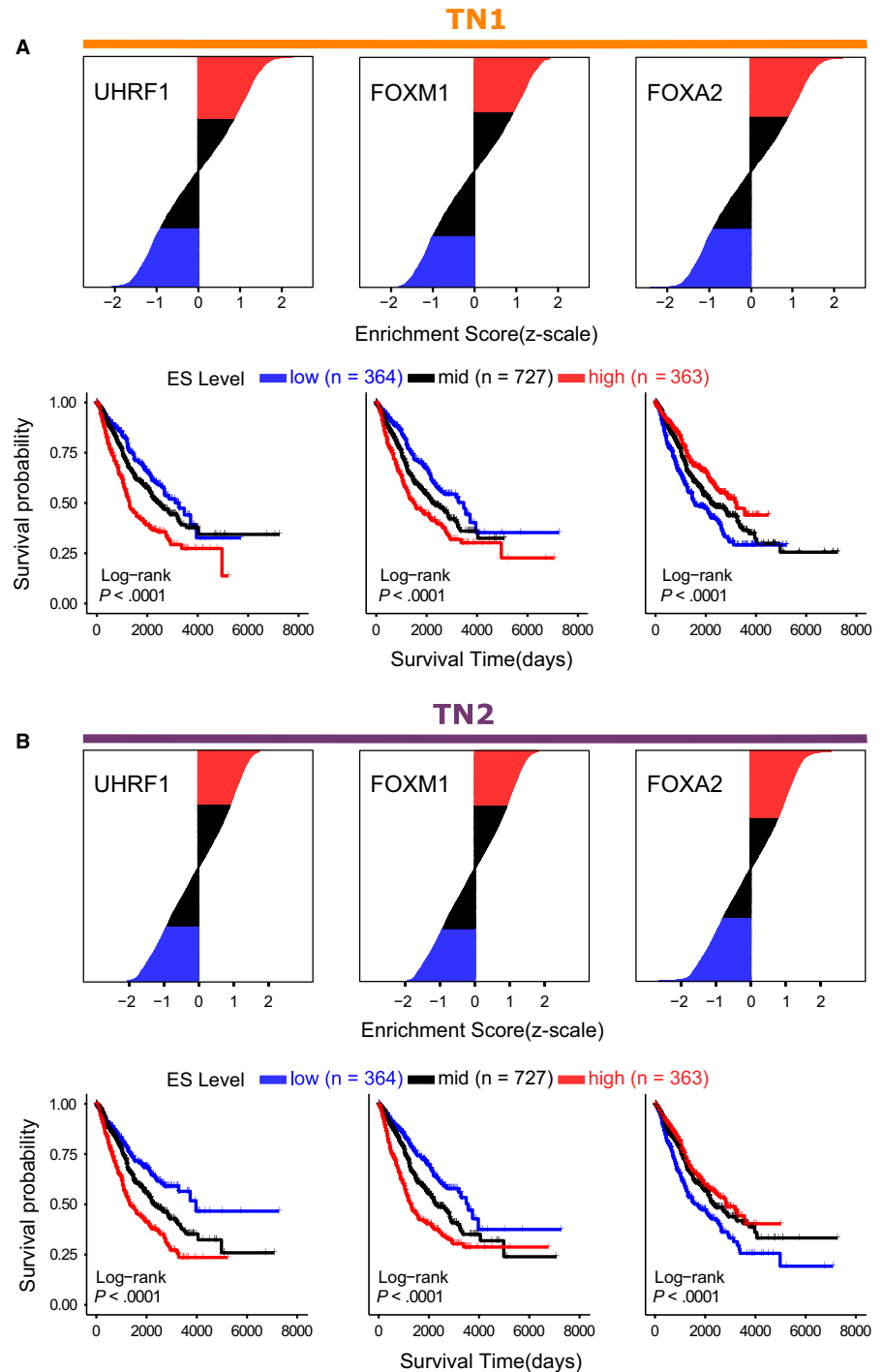
the regulatory units which showed no significant association with survival, or inconsistency in HR direction.

Interestingly, we observed three *consensus* MRs—*UHRF1*, *FOXM1*, and *FOXA2*—significantly associated with survival in  $\geq 50\%$  of the studies queried, for both reference TN (Figure S3). Next, we evaluated the z-scaled, combined ES distribution of these regulatory units using Kaplan-Meier survival curves (Figure 3). We observed that the discretized, standardized ES distributions of all three MRs, especially in the first and third quartiles, significantly separate patient risk of death (log-rank  $P$ -value < .001). Additionally, to decompose the net influence of each target gene in the regulons, we investigated the importance of their expression to patient survival in all three consensus MRs (Figure S4). Indeed, we can observe an overrepresentation of targets whose expressions are significantly associated with risk of death in  $\geq 50\%$  of the datasets investigated, including proliferation markers such as the marker of proliferation Ki-67 (*MKI67*) and checkpoint



**FIGURE 2** Activation State and Survival Network of *Consensus* Master Regulators. A, Two-tailed gene set enrichment analysis was used to query the activation state of *consensus* master regulators in both reference TNs. B, Regulon enrichment was used to investigate the altered regulatory units' association with survival risk using Cox proportional hazards regression in 10 transcriptomic studies. The results were mapped over the *common* MR networks in each transcription network. Node size represents the number of studies in which a significant association of regulons' activity with survival was observed. Node color shows the hazard ratio of these associations

**FIGURE 3** Kaplan-Meier Survival Curves of Master Regulators. Enrichment scores of *consensus* master regulators consistently associated with patient survival were standardized (z-score), merged and their distributions were discretized. The discretized quartile segments (low = first quartile; mid = second quartile; high = third quartile) were then evaluated using Kaplan-Meier curves and log-rank test (lower segments) in both (A) TN1 and (B) TN2 reference networks



kinase 1 (*CHEK1*). Table S1 shows detailed information of each target gene coefficient, adjusted for clinical variables, in all datasets for both reference TNs.

### 3.4 | Master regulator connectivity map (MRCMap)

After retrieving several possible MRs of lung adenocarcinoma, their regulatory units, and these units' association with patient risk, we queried repositioning drug candidates that

would revert the expression profiles of these regulons using a modular variation of the connectivity map method. We employed three different gene list queries, using all case-control studies and both reference TNs.

In the first query, we merged the *common* MRs' regulatory units into a single gene list input. Seventy-one drugs emerged consistently reverting this gene list expression profile in all 13 connectivity maps of case-control (connectivity score  $< 0$ ; FDR-adjusted  $P$ -value  $< .05$ ) studies for TN1; whereas 57 drugs were observed for TN2. Next, the merged regulons of

*consensus* MR were used in the same way, resulting in 59 drug candidates for TN1 and 43 for TN2. Finally, querying the list of genes acquired by merging the three regulatory units associated with patient risk yielded 89 drugs for TN1 and 18 drugs for TN2. Table S2 and Table S3 summarize the results obtained using the master regulators connectivity map pipeline for TN1 and TN2, respectively.

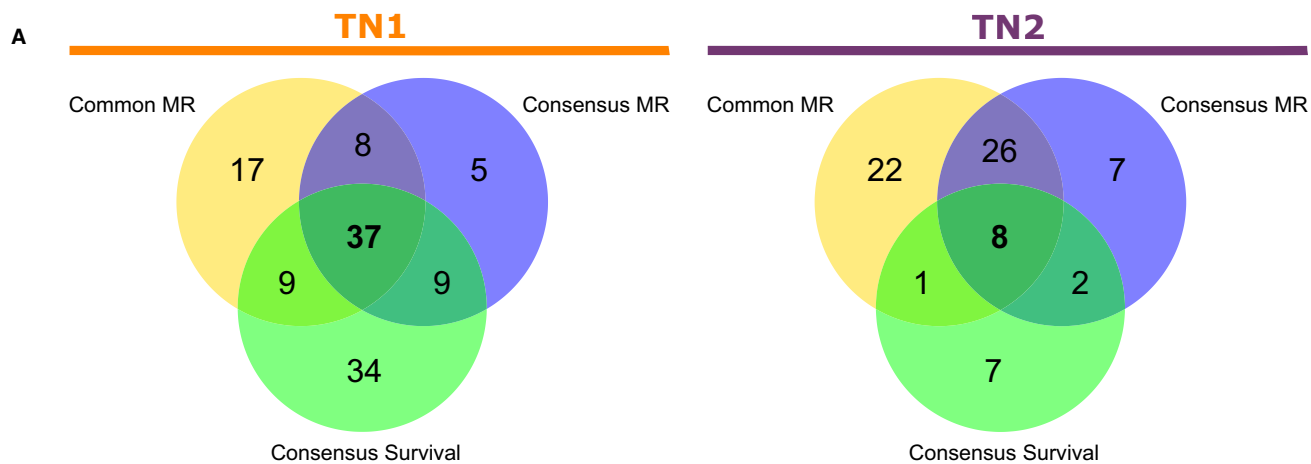
Figure 4A shows the Venn diagrams of the number of drug candidates for repositioning and their intersections, given each gene list input, for both reference TNs. Of these intersections, six molecules (Figure 4B) were consistently present in both reference transcription network MRCMap, namely depropine, promazine, valproic acid, azacyclonol, methotrexate, and ChemBridge ID compound 5109870 (<https://www.hit2lead.com/screening-compounds/5109870>).

## 4 | DISCUSSION

Molecular aberrations in important genes have been the standard of oncology research for many years, especially when searching for cancer biomarkers. However, we are past the phase of single-gene paradigms and have entered the era

of Big Data and systems thinking. Indeed, systems and network biology expanded our perception of complex diseases beyond the actors of the pathological narrative. In this study, we have used a gene regulatory reconstruction approach centered on TFs to study lung adenocarcinoma and employed a systems pharmacology strategy to prospect drugs for repositioning, observing nine TFs acting as master regulators of the disease, and six drug candidates for repurposing.

We used the term GO:0003700 (transcription factor activity, sequence-specific DNA binding) to center the regulatory network reconstruction of nontumorous data. Gene Ontology is the leading information hub on biological knowledge about gene and their products across genomic resources. The ontology is expanded as new findings rise and new annotations occur to keep up with literature. However, despite the efforts, at times the association of ontology and terms with genes is incomplete. This inherent incompleteness hinders the evaluation of computational methods and should be acknowledged when using the database.<sup>33</sup> In our case, this means that some elements under GO:0003700 could include genes not specifically classified as TFs. For example, recently Lambert and collaborators compiled a review on human transcription factors, highlighting the difficulties regarding the identification



### B

Drug	Description	ATC level 1	ATC level 3	CAS	PUBCHEM
depropine	Depropine (INN)	R=RESPIRATORY_SYSTEM	R06A=ANTIHIISTAMINES_FOR_SYSTEMIC_USE	604-51-3	51091740
promazine	Promazine (INN); Combelen [veterinary] (TN)	N=NERVOUS_SYSTEM	N05A=ANTIPSYCHOTICS	58-40-2	96025116
valproic acid	Valproic acid (USP); Depakene (TN)	N=NERVOUS_SYSTEM	N03A=ANTIPILEPTICS	99-66-1	7847465
azacyclonol	NA	NA	NA	NA	NA
methotrexate	Methotrexate (JP17/USP/INN); Otrexup (TN); Xatmep (TN)	L=ANTINEOPLASTIC_AND_IMMUNOMODULATING_AGENTS	L01B=ANTIMETABOLITES	59-05-2	7847210
5109870	NA	NA	NA	NA	NA

**FIGURE 4** Master Regulator Connectivity Maps. A, Venn diagrams of drug candidates obtained from the connectivity maps method. Genes from the *common* master regulator regulons, the eight *consensus* master regulator regulons, and the three master regulators of survival regulons were each combined to form the different sets of gene lists. We only counted molecules in which connectivity score reverted the expression profiles of each gene list; and had FDR-adjusted *P*-value <.05 in all 13 adenocarcinoma case-control studies. Left diagram shows the count of drug candidates found for TN1 and right diagram shows counts of candidates found for TN2 reference network. B, Table showing the six drugs observed in the intersections of the connectivity maps of both reference TNs, their anatomical therapeutic chemical (ATC) classification, and chemical abstracts service (CAS) registry number



of TF-target association and the methods employed to evaluate these interactions. Additionally, they also manually curated the current TF collection, combining putative lists from several sources, including GO. The final count encompassed 1639 known or likely human TFs.<sup>34</sup> According to their curated list, two of the MR candidates we identified (*SMAD6* and *UHRF1*) are not considered true TFs.

Although the master regulator concept is not exclusive to transcription factors, these considerations are important when evaluating our results. Nevertheless, we believe the approach employed here can bring new hints about the pathological scenario of adenocarcinoma. For example, *ATOH8* (atonal Homolog 8) belongs to a group of basic-helix-loop-helix (bHLH) transcription factors involved in the regulation of cardiovascular development, hematopoiesis, skeletal muscle development, neurogenesis, and embryogenesis.<sup>35</sup> Dysregulation of this TF was identified with malignant phenotype in several types of cancer.<sup>35,36</sup> However, its exact role in cancer development remains unclear. In our study, the inferred regulatory unit under *ATOH8* influence was associated with malignant phenotype and repressed in the pathology. To our knowledge, this is the first time *ATOH8* is reported associated with lung adenocarcinoma.

On the other hand, many of the MR candidates obtained by our analyses have been previously reported in lung cancer. For example, *ETV5* (Ets variant 5) is part of the ETS family of transcription factors, which deregulation can alter the expression of proteins involved in stem cell development, cell senescence, proliferation, migration, apoptosis, and tumorigenesis.<sup>37</sup> In lung adenocarcinoma, *ETV5* was shown to inhibit N-cadherin-dependent adhesion in cooperation with its cotranscriptional factor LPP (lipoma-preferred partner), favoring epithelial-to-mesenchymal transition (EMT) and metastatic potential of NSCLC cell line PC14PE6 *in vivo*.<sup>38</sup> On the other hand, loss of *ETV5* seems to impair lung recovery from drug-induced damage, triggering tumor initiation by oncogenic Kras.<sup>39</sup> In our study, inferred *ETV5* activity was observed repressed in lung adenocarcinoma using two-tailed GSEA analysis. Thus, it remains speculative whether these events are tumor specific or depend on cellular-specific circumstances, and what is the exact role of this transcription factor in this disease.

*SMAD6*, *HOXA2*, and *UHRF1* are also poorly explored potential regulators of NSCLC. *SMAD6* negatively regulates BMP, TGF-beta, and activin signaling pathways, which control growth, differentiation, apoptosis of cells, and angiogenesis.<sup>40</sup> Hints to its potential tumor-suppressing activity are reported in several cancers.<sup>41-43</sup> Yet, in lung cancer, Jeon and collaborators initially proposed that *SMAD6* reduction inhibits cancer cell growth and induces apoptosis.<sup>44</sup> Meanwhile, however, other studies reinforce the suppressor potential observed in other cancers.<sup>45-47</sup> Our findings also support these latter observations, since our analysis indicated that *SMAD6* regulon activity was repressed in the lung adenocarcinoma case-control studies utilized.

*HOXA4* (Homeobox A4) belongs to the Homeobox gene family of transcription factors associated with cell differentiation and embryonic development control.<sup>48</sup> *HOXA4* regulation seems to associate with lung cancer cell proliferation, migration, and invasion *in vitro* and *in vivo*, besides poor prognosis in patients.<sup>49,50</sup> We identified that the inferred regulatory unit of *HOXA4* is enriched with differentially expressed genes and its activity was suppressed in tumor tissues compared to control tissues. Therefore, our evaluation also supports the role of *HOXA4* as a potential tumor suppressor in lung adenocarcinoma.

*UHRF1* (Ubiquitin Like with PHD and Ring Finger Domains 1) is essential for cell proliferation and DNA methylation maintenance.<sup>51</sup> It possibly exerts its effects on expression via epigenetic alteration, by interacting with DNMT1 (cytosine-5 DNA methyltransferases) and HDAC1 (histone deacetylase 1), thus propagating or maintaining epigenetic patterns. These interplays may lead to tumor initiation, progression, metastasis, and relapse.<sup>51-53</sup> *UHRF1* is reported as a tumor epigenetic modulator toward malignancy in lung cancer and as a diagnostic biomarker.<sup>54-56</sup> Our results are in agreement with this line of evidence, since its inferred activity was upregulated in the two-tailed GSEA analysis and associated with poor prognosis in survival analyses.

The four remaining transcription factors retrieved as potential MRs—*DACH1*, *EPAS1*, *FOXA2*, and *FOXMI*—have been extensively reported in associations with cancers in literature. *DACH1* is a key member of the retinal determination gene network, a group which affects cell cycle regulation and cancer cell growth, EMT, invasion, and migration in a context-specific manner.<sup>57</sup> Its reduced expression is observed in several types of cancer, mostly associated with poor prognosis.<sup>58</sup> Our results also point that these TFs' inferred activity is repressed in lung adenocarcinoma, similarly to previous lung cancer studies.<sup>59-61</sup> Interestingly, *DACH1* activity did not show a consistently remarkable association with risk in our survival evaluation. This may be because it acts as a regulator of targeted genes also through interaction with other transcription factors.<sup>58</sup> For example, it can directly bind and enhance several functions of p53 in NSCLC,<sup>59</sup> whereas it also can antagonize *FOXMI* transcriptional modulation through competitive binding of DNA segments.<sup>62</sup>

*EPAS1* (endothelial periodic acid-Schiff domain protein 1), also known as HIF2-alpha, is expressed in type II pneumocytes and pulmonary endothelial cells.<sup>63</sup> Several studies evaluating the role of *EPAS1* in lung cancer demonstrated that its activity is puzzling and ambiguous.<sup>64,65</sup> Collectively, the literature suggests inhibition of overexpressed *EPAS1* in lung cancer might have beneficial effects but reductions below a critical threshold favor tumorigenesis and are associated with poor prognosis. In our evaluation of 13 transcriptomic studies, the inferred activity of *EPAS1* was consistently repressed in tumor tissues compared to unaffected lung. Although the

exact mechanisms directing the pro- or antitumoral role remain elusive, it is likely that genetic polymorphisms and epigenetic modifications of *EPAS1* may lead to varied gene expression and outcomes.<sup>65,66</sup>

The two remaining MRs belong to the forkhead box (FOX) gene superfamily, which is described controlling several important biological processes,<sup>67</sup> including airway epithelial differentiation and branching.<sup>68</sup> Several studies observed the tumor suppressor potential of *FOXA2* (forkhead box transcription factor A2) in lung cancer through inhibition of EMT, metastasis, and proliferation.<sup>69-72</sup> Accordingly, our study shows that the inferred activity of *FOXA2* is repressed in tumor samples. Moreover, overall survival evaluation of *FOXA2* activity showed that populations with lower activity were significantly associated with higher risk of death. The exact molecular mechanisms associated with *FOXA2* expression suppression are pending, but evidences suggest that epigenetic alterations are mediating this phenomena.<sup>70</sup> On the other hand, given the complex physiology of TFs, it is hard to dissect the effect of other actors influencing *FOXA2* activity. For example, *NKX2-1* is another master regulator of pulmonary differentiation downregulated in poorly differentiated lung adenocarcinoma, which interacts with *FOXA1/FOXA2* in human cell lines, regulates global Foxa1/Foxa2 binding in murine adenocarcinoma and cooperates to inhibit metastasis.<sup>71,73</sup> Interestingly, our master regulator analysis also retrieved *NKX2-1* as a common MR of lung adenocarcinoma, sharing several targets with *FOXA2* in TN1.

*FOXM1* (forkhead box M1) plays an important role in cell proliferation, cycle regulation, and is expressed only in proliferating cells.<sup>74</sup> In fact, the oncogenic property of its overexpression is well recognized in many types of solid tumors,<sup>75-77</sup> associated with tumor EMT, growth, migration, metastasis, multidrug resistance, and radioresistance.<sup>75,78,79</sup> Recently, Gentles and collaborators observed that an integrated protein-protein association network of *FOXMI* is significantly enriched with adversely prognostic genes in a pan-cancer TCGA study, suggesting that this TF is a major driver of inferior survival regardless of cancer type.<sup>80</sup> We also identified *FOXMI* as a master regulator of lung adenocarcinoma in all 13 transcriptomic case-control studies employed, for both reference TNs reconstructed. Furthermore, its inferred activity was also observed associated with poor prognosis.

Beside the search for master regulators of the pathology and their effects on prognosis, we also explored a systems pharmacology, module-oriented implementation of drug repositioning using the connectivity maps paradigm. Unsurprisingly, this approach retrieved molecules already used or proposed in lung cancer therapy, such as methotrexate and valproic acid. The first drug, methotrexate, is one of the oldest, most common and efficient antineoplastic drugs. It acts by inhibiting dihydrofolate reductase to deplete intracellular tetrahydrofolate, impairing thymidylate production and leading to apoptosis or autophagy

of highly proliferative cell populations. Unfortunately, its clinical use is associated with various toxicities.<sup>81,82</sup> Valproic acid is reported to be a class I histone deacetylase (HDAC) inhibitor commonly used to treat mood disorders and epilepsy. Its anticancer effect has been reported in several cancers in recent years, including lung adenocarcinoma,<sup>83-85</sup> through modulations of epigenetic patterns via HDAC inhibition and gene expression modification. Furthermore, many of the MR candidates retrieved by our analyses have reported or suggested epigenetic mechanisms of modulation, which further supports the study and search for pharmacological epigenetic controllers in cancer therapy research.

The remaining drugs have few to no associations with lung cancer therapy in literature, and occasional suggestions in other types of cancer. For example, a previous CMap study also retrieved azacyclonol as a potential inhibitor of transcription factor activity using lung cancer expression profiles,<sup>86</sup> albeit using a different querying strategy. Promazine, a phenothiazine derivate antipsychotic, had an antitumoral activity in breast cancer cell lines<sup>87</sup> and was able to induce apoptosis and tumor growth of diffuse large B-cell lymphoma in vivo and in vitro.<sup>88</sup> Recently, depropine, a H1-histamine and muscarinic receptor antagonist, was reported inhibiting cell viability and mammosphere formation of breast cancer stem cells, but it did not inhibit the self-renewal capacities of MDA-MB-231 cells.<sup>89</sup> Finally, regarding ChemBridge ID compound 5109870, not much is known about its behavior in biological systems besides its iron chelating effect<sup>90,91</sup> and this is the first study suggesting its potential as an anticancer drug.

Finally, it is important to highlight one aspect of Systems Biology that is difficult to assess using the many network analyses usually employed, that is, the dynamical changes in association patterns between genes during phenotypic transitions.<sup>92</sup> Indeed, most reverse-engineering methods model gene networks as static processes, in which interaction changes among elements in the network are not accounted across different conditions.<sup>93</sup> To address this issue, recent studies have been working toward the development of such methods using a differential coregulation framework.<sup>92,94</sup> However, the addition of dynamic approaches is no trivial endeavor and these methods can be hindered by network reconstruction stability problems,<sup>95,96</sup> since network topology is of critical importance in these scenarios.<sup>95</sup> This can be especially troublesome in multidataset studies such as ours. Thus, we opted to evaluate each TN independently in our study and to highlight the consistencies between results obtained from each network.

The importance of OMICS technologies and Systems Biology to the adoption of a holistic paradigm of biology is undeniable, enhancing our understanding of intricate organizations and uncovering molecular signatures underlying complex cellular phenotypes. Our computational approach retrieved nine TFs potentially acting as master regulators of lung adenocarcinoma phenotype. Among them, several are

already described as actors of tumorigenesis. However, this is the first study implicating *ATOH8* with lung cancer development. Evidently, further experimental studies are required to fully address its role in the pathology. We also identified six drug candidates for repositioning which could revert the pathological transcription profiles resulted from deregulated activity patterns the three MR candidates robustly associated with risk of death—*FOXA2*, *FOXMI*, and *UHRF1*. We believe the thorough evaluation of these molecules in cellular and animal models, besides randomized clinical trials can potentially lead to the development of new therapies and therapeutic strategies for treatment.

## CONFLICT OF INTEREST

The authors declare no potential conflicts of interest.

## AUTHOR CONTRIBUTIONS

Marco Antônio De Bastiani was involved in conceptualization, formal analyses, methodology, project administration, writing, reviewing, and editing original draft. Fábio Klamt was involved in resource and funding acquisition, supervision, review, and editing.

## DATA AVAILABILITY STATEMENT

Datasets used in this study can be accessed via NCBI GEO portal (<https://www.ncbi.nlm.nih.gov/geo/>). Further intermediate data and codes generated are available from the corresponding author on request.

## ORCID

Marco Antônio De Bastiani  <https://orcid.org/0000-0003-3924-5278>

Fábio Klamt  <https://orcid.org/0000-0001-9152-4229>

## REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
- Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med*. 2008;359(13):1367-1380.
- Nakamura H, Saji H. Worldwide trend of increasing primary adenocarcinoma of the lung. *Surg Today*. 2014;44(6):1004-1012.
- Bhimji SS, Wallen JM. *Cancer, Lung Adenocarcinoma*; StatPearls Publishing, 2018. <https://www.ncbi.nlm.nih.gov/books/NBK519578/>.
- Scagliotti GV, Parikh P, von Pawel J, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol*. 2008;26(21):3543-3551. <https://doi.org/10.1200/JCO.2007.15.0375>.
- Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12(1):56-68.
- McWilliam H, Li W, Uludag M, et al. Tool web services from the EMBL-EBI. *Nucleic Acids Res*. 2013;41(W1):W597-W600.
- Linde J, Schulze S, Henkel SG, Guthke R. Data- and knowledge-based modeling of gene regulatory networks: an update. *EXCLI J*. 2015;14:346-378.
- Kin Chan SS, Chan SS-K, Kyba M, Kin Chan SS. What is a master regulator? *J Stem Cell Res Ther*. 2013;03(02):10-13.
- Yeh JE, Toniolo PA, Frank DA. Targeting transcription factors: promising new strategies for cancer therapy. *Curr Opin Oncol*. 2013;25(6):652-658.
- Bhagwat AS, Vakoc CR. Targeting transcription factors in cancer. *Trends Cancer*. 2015;1(1):53-65.
- Benson N, van der Graaf PH. The rise of systems pharmacology in drug discovery and development. *Future Med Chem*. 2014;6(16):1731-1734.
- Van Hasselt J, Van Der Graaf PH. Towards integrative systems pharmacology models in oncology drug development. *Drug Discov Today Technol*. 2015;15:1-8.
- Jadamba E, Shin M. A systematic framework for drug repositioning from integrated omics and drug phenotype profiles using pathway-drug network. *Biomed Res Int*. 2016;2016:1-17.
- Chung F-H, Chiang Y-R, Tseng A-L, et al. Module Connectivity Map (FMCM): a framework for searching repurposed drug compounds for systems treatment of cancer and an application to colorectal adenocarcinoma. *PLoS ONE*. 2014;9(1):e86299.
- Lamb J. The connectivity map: a new tool for biomedical research. *Nat Rev Cancer*. 2007;7(1):54-60.
- Ritchie ME, Phipson B, Wu DI, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- Fletcher M, Castro M, Wang X, et al. Master regulators of FGFR2 signalling and breast cancer risk. *Nat Commun*. 2013;4(1):1-12.
- Castro M, de Santiago I, Campbell TM, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat Genet*. 2015;48(1):12-21.
- Margolin AA, Nemenman I, Basso K, et al. An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*. 2006;7(S1):1-15.
- Carro MS, Lim WK, Alvarez MJ, et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature*. 2014;463(7279):318-325.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102(43):15545-15550.
- Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-526.
- Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*. 2013;14(1):7.
- Kassambara A, Kosinski M, Biecek P. survminer: drawing survival curves using 'ggplot2'. *R Packag version 03*. 2017;1.
- Therneau TM, Lumley T. Package 'survival'. *R Top Doc*. 2015;128.
- Lamb J, Crawford ED, Peck D, et al. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;313(5795):1929-1935.



28. De Bastiani MA, Pfaffenseller B, Klamt F. Master regulators connectivity map: a transcription factors-centered approach to drug repositioning. *Front Pharmacol.* 2018;9:1-9.
29. Smirnov P, Safikhani Z, El-Hachem N, et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics.* 2016;32(8):1244-1246.
30. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. <https://www.R-project.org/>. Accessed August 14, 2019.
31. Castro M, Wang X, Fletcher M, Meyer KB, Markowitz F. RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. *Genome Biol.* 2012;13(4):R29.
32. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag; 2009.
33. Gaudet P, Dessimoz C. Gene ontology: pitfalls, biases, and remedies. *Methods Mol Biol.* 2017;1446:189-205.
34. Lambert SA, Jolma A, Campitelli LF, et al. The human transcription factors. *Cell.* 2018;172(4):650-665.
35. Song Y, Pan G, Chen L, et al. Loss of ATOH8 increases stem cell features of hepatocellular carcinoma cells. *Gastroenterology.* 2015;149(4):1068-1081.
36. Wang Z, Xie J, Yan M, et al. Downregulation of ATOH8 induced by EBV-encoded LMP1 contributes to the malignant phenotype of nasopharyngeal carcinoma. *Oncotarget.* 2016;7(18): <https://doi.org/10.18632/oncotarget.8503>.
37. Huang W, Waknitz M. ETS gene fusions and prostate cancer. *Am J Transl Res.* 2009;1(4):341-351.
38. Kuriyama S, Yoshida M, Yano S, et al. LPP inhibits collective cell migration during lung cancer dissemination. *Oncogene.* 2016;35(8):952-964.
39. Zhang Z, Newton K, Kummerfeld SK, et al. Transcription factor ETV5 is essential for the maintenance of alveolar type II cells. *Proc Natl Acad Sci.* 2017;114(15):3903-3908.
40. Pickup MW, Owens P, Moses HL. TGF- $\beta$ , bone morphogenetic protein, and activin signaling and the tumor microenvironment. *Cold Spring Harb Perspect Biol.* 2017;9(5):a022285.
41. de Boeck M, Cui C, Mulder AA, Jost CR, Ikeno S, Ten Dijke P. Smad6 determines BMP-regulated invasive behaviour of breast cancer cells in a zebrafish xenograft model. *Sci Rep.* 2016;6:24968.
42. Osawa H, Nakajima M, Kato H, Fukuchi M, Kuwano H. Prognostic value of the expression of Smad6 and Smad7, as inhibitory Smads of the TGF-beta superfamily, in esophageal squamous cell carcinoma. *Anticancer Res.* 2004;24(6):3703-3709.
43. Singh P, Srinivasan R, Wig J. The Smad family and its role in pancreatic cancer. *Indian J Cancer.* 2011;48(3):351.
44. Jeon H-SS, Dracheva T, Yang S-HH, et al. SMAD6 contributes to patient survival in non-small cell lung cancer and its knockdown reestablishes TGF- $\beta$  homeostasis in lung cancer cells. *Cancer Res.* 2008;68(23):9686-9692.
45. Frullanti E, Colombo F, Falvella FS, et al. Association of lung adenocarcinoma clinical stage with gene expression pattern in noninvolved lung tissue. *Int J Cancer.* 2012;131(5):E643-E648.
46. Mullapudi N, Ye B, Suzuki M, et al. Genome wide methylome alterations in lung cancer. *PLoS ONE.* 2015;10(12):1-21.
47. Yang D, Ma M, Zhou W, Yang B, Xiao C. Inhibition of miR-32 activity promoted EMT induced by PM2.5 exposure through the modulation of the Smad1-mediated signaling pathways in lung cancer cells. *Chemosphere.* 2017;184:289-298.
48. Pick L. Hox genes, evo-devo, and the case of the ftz gene. *Chromosoma.* 2016;125(3):535-551.
49. Cheng S, Qian F, Huang Q, Wei L, Fu Y, Du Y. HOXA4, down-regulated in lung cancer, inhibits the growth, motility and invasion of lung cancer cells. *Cell Death Dis.* 2018;9(5):465.
50. Tang X, Jiang J, Zhu J, He N, Tan J. HOXA4-regulated miR-138 suppresses proliferation and gefitinib resistance in non-small cell lung cancer. *Mol Genet Genomics.* 2019;294(1):85-93.
51. Patnaik D, Estève P-O, Pradhan S. Targeting the SET and RING-associated (SRA) domain of ubiquitin-like, PHD and ring finger-containing 1 (UHRF1) for anti-cancer drug development. *Oncotarget.* 2018;9(40):26243-26258.
52. Sidhu H, Capalash N. UHRF1: the key regulator of epigenetics and molecular target for cancer therapeutics. *Tumour Biol.* 2017;39(2):1010428317692205. <https://doi.org/10.1177/1010428317692205>.
53. Jin W, Liu Y, Xu S-G, et al. UHRF1 inhibits MDR1 gene transcription and sensitizes breast cancer cells to anticancer drugs. *Breast Cancer Res Treat.* 2010;124(1):39-48.
54. Unoki M, Brunet J, Mousli M. Drug discovery targeting epigenetic codes: the great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. *Biochem Pharmacol.* 2009;78(10):1279-1288.
55. Unoki M, Daigo Y, Koinuma J, Tsuchiya E, Hamamoto R, Nakamura Y. UHRF1 is a novel diagnostic marker of lung cancer. *Br J Cancer.* 2010;103(2):217-222.
56. Huang P, Cheng C-L, Chang Y-H, et al. Molecular gene signature and prognosis of non-small cell lung cancer. *Oncotarget.* 2016;7(32):51898-51907.
57. Liu YU, Han NA, Zhou SI, et al. The DACH/EYA/SIX gene network and its role in tumor initiation and progression. *Int J Cancer.* 2016;138(5):1067-1075.
58. Kong D, Liu YU, Liu Q, et al. The retinal determination gene network: from developmental regulator to cancer therapeutic target. *Oncotarget.* 2016;7(31):50755-50765.
59. Chen K, Wu K, Cai S, et al. Dachshund binds p53 to block the growth of lung adenocarcinoma cells. *Cancer Res.* 2013;73(11):3262-3274.
60. Han NA, Yuan X, Wu H, et al. DACH1 inhibits lung adenocarcinoma invasion and tumor growth by repressing CXCL5 signaling. *Oncotarget.* 2015;6(8):5877-5888.
61. Zhu JI, Wu C, Li H, et al. DACH1 inhibits the proliferation and invasion of lung adenocarcinoma through the downregulation of peroxiredoxin 3. *Tumor Biol.* 2016;37(7):9781-9788.
62. Zhou J, Wang C, Wang Z, et al. Attenuation of Forkhead signaling by the retinal determination factor DACH1. *Proc Natl Acad Sci USA.* 2010;107(15):6864-6869.
63. Wiesener MS, Jürgensen JS, Rosenberger C, et al. Widespread hypoxia-inducible expression of HIF-2 $\alpha$  in distinct cell populations of different organs. *FASEB J.* 2003;17(2):271-273.
64. Mazumdar J, Hickey MM, Pant DK, et al. HIF-2 $\alpha$  deletion promotes Kras-driven lung tumor development. *PNAS.* 2010;107(32):14182-14187.
65. Xu X-H, Bao Y, Wang X, et al. Hypoxic-stabilized EPAS1 proteins transactivate DNMT1 and cause promoter hypermethylation and transcription inhibition of EPAS1 in non-small cell lung cancer. *FASEB J.* 2018;32(12):6694-6705. <https://doi.org/10.1096/fj.201700715>.

66. Wang Z, Wei Y, Zhang R, et al. Multi-omics analysis reveals a HIF network and hub gene EPAS1 associated with lung adenocarcinoma. *EBioMedicine*. 2018;32:93-101.
67. Friedman JR, Kaestner KH. The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci*. 2006;63(19–20):2317-2328.
68. Wan H, Dingle S, Xu Y, et al. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem*. 2005;280(14):13809-13816.
69. Tang Y, Shu G, Yuan X, Jing N, Song J. FOXA2 functions as a suppressor of tumor metastasis by inhibition of epithelial-to-mesenchymal transition in human lung cancers. *Cell Res*. 2011;21(2):316-326.
70. Basseres DS, D'Alò F, Yeap BY, et al. Frequent downregulation of the transcription factor Foxa2 in lung cancer through epigenetic silencing. *Lung Cancer*. 2012;77(1):31-37.
71. Li C-C, Gocheva V, Oudin MJ, et al. Foxa2 and Cdx2 cooperate with Nkx2-1 to inhibit lung adenocarcinoma metastasis. *Genes Dev*. 2015;29(17):1850-1862.
72. Huang C, Liu J, Xiong B, Yonemura Y, Yang X. Expression and prognosis analyses of forkhead box A (FOXA) family in human lung cancer. *Gene*. 2019;685:202-210.
73. Snyder E, Watanabe H, Magendantz M, et al. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol Cell*. 2013;50(2):185-199.
74. Wang I-C, Chen Y-J, Hughes D, et al. Forkhead box M1 regulates the transcriptional network of genes essential for mitotic progression and genes encoding the SCF (Skp2-Cks1) ubiquitin ligase. *Mol Cell Biol*. 2005;25(24):10875-10894.
75. Raychaudhuri P, Park HJ. FoxM1: a master regulator of tumor metastasis. *Cancer Res*. 2011;71(13):4329-4333.
76. Dai J, Yang L, Wang J, Xiao Y, Ruan Q. Prognostic value of FOXM1 in patients with malignant solid tumor: a meta-analysis and system review. *Dis Markers*. 2015;2015:352478.
77. Gartel AL. FOXM1 in cancer: interactions and vulnerabilities. *Cancer Res*. 2017;77(12):3135-3139.
78. Yao S, Fan L, Lam E. The FOXO3-FOXM1 axis: a key cancer drug target and a modulator of cancer drug resistance. *Semin Cancer Biol*. 2017;2018(50):77-89.
79. Xiu G, Sui X, Wang Y, Zhang Z. FOXM1 regulates radiosensitivity of lung cancer cell partly by upregulating KIF20A. *Eur J Pharmacol*. 2018;833:79-85.
80. Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*. 2015;21(8):938-945.
81. Khan ZA, Tripathi R, Mishra B. Methotrexate: a detailed review on drug delivery and clinical aspects. *Expert Opin Drug Deliv*. 2012;9(2):151-169.
82. Hess JA, Khasawneh MK. Cancer metabolism and oxidative stress: Insights into carcinogenesis and chemotherapy via the non-dihydrofolate reductase effects of methotrexate. *BBA Clin*. 2015;3:152-161.
83. Zhuo W, Zhang L, Zhu Y, Xie Q, Zhu B, Chen Z. Valproic acid, an inhibitor of class I histone deacetylases, reverses acquired Erlotinib-resistance of lung adenocarcinoma cells: A Connectivity Mapping analysis and an experimental study. *Am J Cancer Res*. 2015;5(7):2202-2211.
84. Chen J-H, Zheng Y-L, Xu C-Q, et al. Valproic acid (VPA) enhances cisplatin sensitivity of non-small cell lung cancer cells via HDAC2 mediated down regulation of ABCA1. *Biol Chem*. 2017;398(7):785-792.
85. Wang L-L, Zhou L-B, Shu J, et al. Up-regulation of IRF-3 expression through GATA-1 acetylation by histone deacetylase inhibitor in lung adenocarcinoma A549 cells. *Oncotarget*. 2017;8(44):75943-75951.
86. Xue D, Lu M, Gao B, Qiao X, Zhang Y. Screening for transcription factors and their regulatory small molecules involved in regulating the functions of CL1-5 cancer cells under the effects of macrophage-conditioned medium. *Oncol Rep*. 2014;31(3):1323-1333.
87. Kuzma-Richeret A, Saczko J, Choromańska A, et al. The influence of phenothiazine derivatives on doxorubicin treatment in sensitive and resistant human breast adenocarcinoma cells. *Folia Biol (Praha)*. 2011;57(6):261-267.
88. Nagel D, Spranger S, Vincendeau M, et al. Pharmacologic inhibition of MALT1 protease by phenothiazines as a therapeutic approach for the treatment of aggressive ABC-DLBCL. *Cancer Cell*. 2012;22(6):825-837.
89. Cui J, Hollmén M, Li L, et al. New use of an old drug: inhibition of breast cancer stem cells by benzotropine mesylate. *Oncotarget*. 2016;8(1):1007-1022.
90. Coombs GS, Schmitt AA, Canning CA, et al. Modulation of Wnt/ $\beta$ -catenin signaling and proliferation by a ferrous iron chelator with therapeutic efficacy in genetically engineered mouse models of cancer. *Oncogene*. 2012;31(2):213-225.
91. Fryknäs M, Zhang X, Bremberg U, et al. Iron chelators target both proliferating and quiescent cancer cells. *Sci Rep*. 2016;6:38343.
92. Lichtblau Y, Zimmermann K, Haldemann B, Lenze D, Hummel M, Leser U. Comparative assessment of differential network analysis methods. *Brief Bioinform*. 2017;18(5):837-850.
93. Ideker T, Krogan NJ. Differential network biology. *Mol Syst Biol*. 2012;8:565.
94. Gambardella G, Moretti MN, de Cegli R, Cardone L, Peron A, di Bernardo D. Differential network analysis for the identification of condition-specific pathway activity and regulation. *Bioinformatics*. 2013;29(14):1776-1785.
95. Jurman G, Filosi M, Visintainer R, Riccadonna S, Furlanello C. Stability in GRN inference. *Methods Mol Biol*. 2019;1883:323-346.
96. Filosi M, Visintainer R, Riccadonna S, Jurman G, Furlanello C. Stability indicators in network reconstruction. *PLoS ONE*. 2014;9(2):e89815.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** De Bastiani MA, Klamt F. Integrated transcriptomics reveals master regulators of lung adenocarcinoma and novel repositioning of drug candidates. *Cancer Med*. 2019;8:6717–6729. <https://doi.org/10.1002/cam4.2493>