

# BART Cancer: a web resource for transcriptional regulators in cancer genomes

Zachary V. Thomas<sup>1,2</sup>, Zhenjia Wang<sup>1</sup> and Chongzhi Zang<sup>1,2,3,\*</sup>

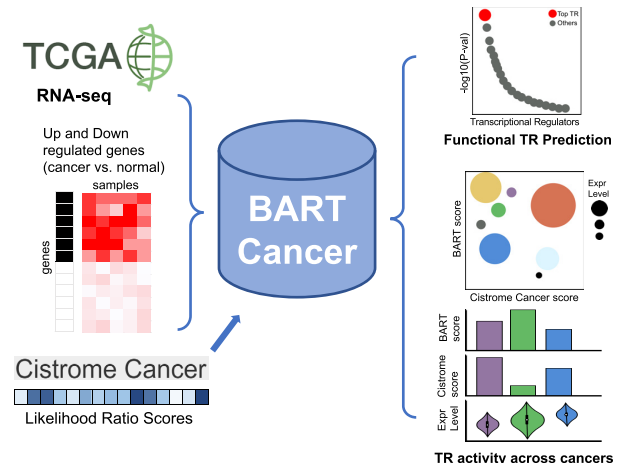
<sup>1</sup>Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA, <sup>2</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA and <sup>3</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA

Received December 17, 2020; Revised February 09, 2021; Editorial Decision February 25, 2021; Accepted March 01, 2021

## ABSTRACT

Dysregulation of gene expression plays an important role in cancer development. Identifying transcriptional regulators, including transcription factors and chromatin regulators, that drive the oncogenic gene expression program is a critical task in cancer research. Genomic profiles of active transcriptional regulators from primary cancer samples are limited in the public domain. Here we present BART Cancer ([bartcancer.org](http://bartcancer.org)), an interactive web resource database to display the putative transcriptional regulators that are responsible for differentially regulated genes in 15 different cancer types in The Cancer Genome Atlas (TCGA). BART Cancer integrates over 10000 gene expression profiling RNA-seq datasets from TCGA with over 7000 ChIP-seq datasets from the Cistrome Data Browser database and the Gene Expression Omnibus (GEO). BART Cancer uses Binding Analysis for Regulation of Transcription (BART) for predicting the transcriptional regulators from the differentially expressed genes in cancer samples compared to normal samples. BART Cancer also displays the activities of over 900 transcriptional regulators across cancer types, by integrating computational prediction results from BART and the Cistrome Cancer database. Focusing on transcriptional regulator activities in human cancers, BART Cancer can provide unique insights into epigenetics and transcriptional regulation in cancer, and is a useful data resource for genomics and cancer research communities.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Cancer and its progression arise from dysregulation of gene expression (1,2). The identification of transcriptional regulators (TRs), including transcription factors (TFs) and chromatin regulators (CRs), that control oncogenic gene expression program in each cancer type is an essential task for cancer research as these regulators could be targets for novel therapies. The Cancer Genome Atlas (TCGA) consortium has generated more than 10000 RNA-seq datasets for tumor and normal samples for over 30 cancer types, as one of the largest resources for gene expression profiles in human cancers (3). Knowing what TRs regulate the genes with differential expression between tumor and normal samples can help better understand functional gene regulatory networks in each cancer type. However, TCGA did not produce ChIP-seq data due to limited cell numbers for most primary tumor samples. Therefore, computational prediction of functional TRs from differential gene expression data from TCGA will be useful in understanding cancer transcriptional regulation. Such efforts include the Cistrome Cancer web resource, which uses a comprehensive computational modeling approach to integrate TCGA data with

\*To whom correspondence should be addressed. Email: zang@virginia.edu

publicly available chromatin profiling ChIP-seq data and reports the results on predicted TF targets and putative enhancer profiles in each cancer type (4). TF activities in Cistrome Cancer were inferred primarily based on the expression of the TF genes from TCGA and correlations with their putative target genes. The direct binding profiles of TFs from ChIP-seq data and their association with the putative enhancer profiles were not utilized in the model. Additionally, Cistrome Cancer focused on 575 factors, while growing ChIP-seq data from the public domain have covered more than 900 TRs (5–7), calling for an expansion of the existing database.

We previously developed Binding Analysis for Regulation of Transcription (BART), a computational method for inference of functional TRs from a target gene set (8). BART leverages a large collection of over 7000 human TR-binding profiles and 5000 mouse TR-binding profiles from the Cistrome Data Browser database (6,7) to build the inference model. When using gene list input, BART first uses collected ChIP-seq profiles of H3K27ac, an active enhancer histone mark, to predict a cis-regulatory profile that regulates the gene set. Each TR ChIP-seq dataset from the compendium is then mapped to the cis-regulatory profile and scored based on whether the TR-binding sites tend to overlap with high-ranking regions in the cis-regulatory profile, quantified by a receiver-operating characteristic (ROC) measure. The scores for all ChIP-seq datasets for the same TR are then integrated to generate TR-level quantifications using a statistical model (8). We demonstrated that BART has good performance in identifying the true TRs using differentially expressed genes generated from TR knock-down/knock-out experiments (9,10).

BART can be utilized for inferring functional TRs from differentially expressed genes derived from TCGA data, providing a new approach for integration of tumor molecular profiling data and public ChIP-seq data. By using BART on both up- and down-regulated genes in cancer compared to normal, over 900 TRs can be reported based on their ranks of predicted activity in each cancer type. This allows us to identify not only regulators that promote tumorigenesis (from up-regulated genes) but also regulators that may act as tumor suppressors (from down-regulated genes). Here, we present BART Cancer, an interactive web resource database to report the putative transcriptional regulators that are responsible for differentially expressed genes in 15 different cancer types in TCGA, and multiple activity measures of over 900 transcriptional regulators across these 15 cancer types. An overview of the workflow can be seen in Figure 1.

## MATERIALS AND METHODS

### TCGA data collection and processing

RNA-seq data from 31 cancer types were downloaded from TCGA. In order to obtain a robust set of gene expression profiles, normalized read count in FPKM data from over 10 000 cancer samples were clustered using *K*-means clustering. About 31 clusters were retained and the clusters from unambiguous cancer type annotation were kept with the TCGA annotation. Misclassified samples (<100) were removed from other cancer type clusters. Clusters with sim-

ilar expression patterns from multiple cancer types were merged, e.g., esophagus-stomach cancer (STES) was stomach adenocarcinoma (STAD) and esophageal carcinoma (ESCA) combined; colorectal cancer was colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) combined. Additionally, breast cancer (BRCA) was clustered into two separate clusters: BRCA\_1 for mainly luminal breast cancer (764 of 833 are luminal) and BRCA\_2 for mainly basal breast cancer (176 of 234 are basal), similar to the results in Cistrome Cancer (4). A cancer type was only included in BART Cancer if TCGA included both cancer genomic profiles and normal genomic profiles, leaving BART Cancer with 15 different reannotated cancer types. For each cancer type, both up- and down-regulated genes were identified using DESeq2 (11). Differentially expressed genes were defined as those having adjusted *P*-value <1e-7 and log<sub>2</sub>(fold change) >1 for up-regulated genes and < -1 for down-regulated genes. Differentially expressed genes provided by Cistrome Cancer (4) were used to cross-reference the newly identified cancer up-regulated genes. Because Cistrome Cancer does not provide down-regulated genes, newly identified down-regulated genes were not cross-referenced against any sources. These genes were then used as input to BART to get the putative transcriptional regulators responsible for regulating each gene set.

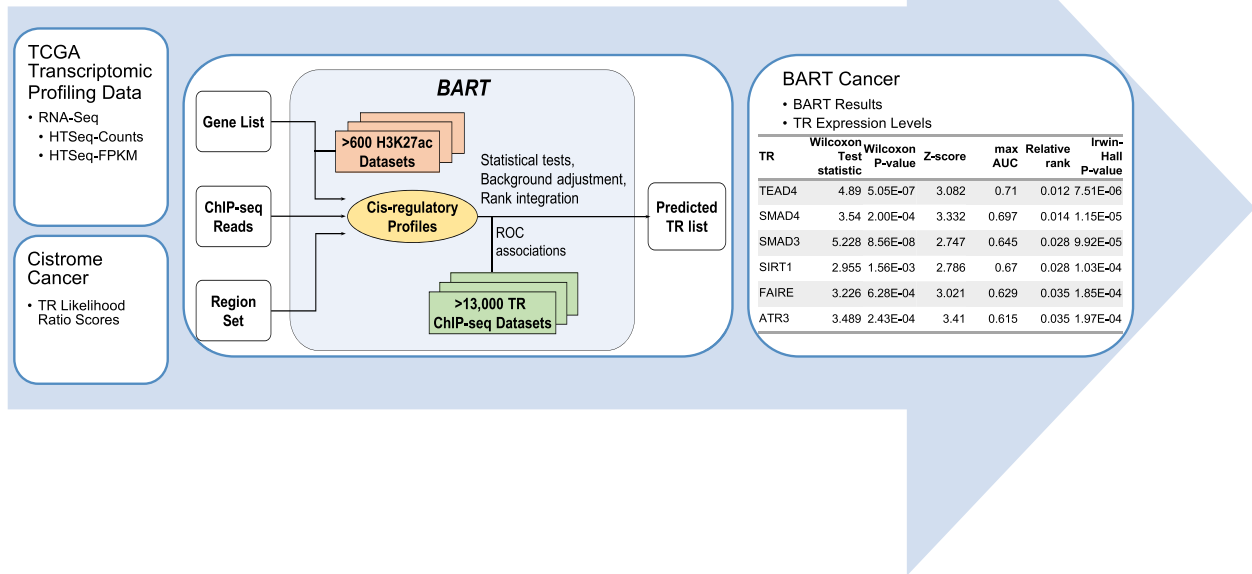
### BART analysis for TR prediction

BART is a bioinformatics tool for predicting functional transcriptional regulators that bind at cis-regulatory regions to regulate gene expression in human or mouse (8). BART version 2.0 was used to generate TR prediction results for both up-regulated and down-regulated gene lists derived from each cancer type. The main output of BART is a list of putative TRs with multiple quantification scores. The top ranked regulators are predicted putative regulators that are associated with the input dataset. The relative rank, calculated by the rank divided by the total number of TRs included in the analysis, was used as the BART prediction score for each TR in the data visualization. The Irwin-Hall *P*-value, calculated from the Irwin-Hall distribution as the null distribution for rank integration, was used as the final significance measure to rank all TRs.

### Database design and implementation

BART Cancer is designed using HTML5, CSS and JavaScript technologies. The Tabulator library (<http://tabulator.info/>) is used to display the BART result tables as well as list TRs to select plots to display. All TR plots were generated in R with ggplot2. The results, figures, and differentially expressed genes are all freely available for download.

The web-based user interface provides ways for researchers to view, visualize, and download the provided data. By selecting a cancer type and a type of differential expression, i.e., down- or up-regulated in cancer versus normal samples, the user is able to view the BART results of running BART on those genes for the specified cancer type. Directly below the display table include links to download the table as well as supplementary data files. Also provided



**Figure 1.** Schematic Representation of BART Cancer Workflow. RNA sequencing profiles were collected from TCGA and used to determine differentially expressed genes. These gene lists were then used as input to BART to get a putative transcriptional regulator (TR) list. This TR list, combined with Cistrome Cancer likelihood ratio scores, are displayed in BART Cancer.

is the link to the BARTweb (9) help page to better understand the meaning of the BART results (Figure 2).

Below the BART results table module, TR specific plots are displayed. Each plot displays the expression level, BART rank, and the likelihood ratio score of a TR across all cancer types. TR gene's average expression level was quantified as  $\log(\text{FPKM} + 0.1)$  from TCGA RNA-seq data and ranked across the 15 cancer types. BART rank was quantified as 1-relative rank from the BART result. The likelihood ratio score was obtained from the Cistrome Cancer database, measuring the regulatory association of ChIP-seq data for this TR to the target genes predicted in Cistrome Cancer (4). An additional figure is included that combines the data from the other three figures into one bubble plot for each TR. In these plots, the BART Rank is plotted against the Cistrome Cancer likelihood ratio and the bubble size corresponds to relative expression level ranked in the 15 cancer types (Figure 3). This results in cancer types that had the chosen TR ranked high in both BART and Cistrome Cancer likelihood appearing in the upper right quadrant. A large bubble of some color appearing in the upper right quadrant indicates that the TR is specifically expressed and activating its target genes in the cancer type represented by that color. This module uses a column of TR names which, when selected, change the figures displayed to the selected TR.

## RESULTS

BART Cancer web resource is created for users to browse putative TRs that are likely to regulate differential gene ex-

pression in a particular cancer and to visualize the relative activity of a TR across multiple cancer types. In order to validate that BART Cancer reports real functional TRs, we did a survey on the top three TRs predicted by BART from differentially expressed genes in each cancer type and cross-referenced with literature (Table 1).

### TRs for up-regulated genes

The TRs predicted from the up-regulated genes in cancer can be associated with oncogenic activation and promote tumorigenesis. Here, we explore some of the top ranked TRs including histone deacetylase 2 (HDAC2), estrogen receptor 1 (ESR1), TEA domain family member 4 (TEAD4), achaete-scute homolog 1 (ASCL1), and androgen receptor (AR) and yes associated protein 1 (YAP1) in several cancer types.

In bladder urothelial carcinoma (BLCA), HDAC2 ranked among the top three putative TRs. This class 1 deacetylase has been shown to be up-regulated in bladder tumors (12). Additionally, a 2016 study showed that the inhibition of HDAC2 successfully inhibited cell proliferation and suggested that the inhibition of HDAC2 may be a promising therapy for urothelial carcinoma (13).

The top ranked putative regulator in luminal breast invasive carcinoma (BRCA luminal) was ESR1, which has been previously shown as a predictive biomarker of breast cancer. These results agree with previous studies on breast cancer that estrogen receptors and ESR1 mutations are markers of poor prognosis (14).

In colorectal adenocarcinoma (COAD and READ), TEAD4 has been shown to be a biomarker and tumorigenic

## BART Results

Select a cancer type and differential expression type to view BART results

Select Cancer Type

- Bladder Urothelial Carcinoma (BLCA)
- Breast Invasive Carcinoma (Luminal) (BRCA\_1)
- Breast Invasive Carcinoma (Basal) (BRCA\_2)
- Colorectal Adenocarcinoma (COAD\_READ)
- Head and Neck Squamous Cell Carcinoma (HNSC)
- Kidney Chromophobe (KICH)
- Kidney Renal Clear Cell Carcinoma (KIRC)
- Kidney Renal Papillary Cell Carcinoma (KIRP)
- Liver Hepatocellular Carcinoma (LIHC)
- Lung Adenocarcinoma (LUAD)
- Lung Squamous Cell Carcinoma (LUSC)
- Prostate Adenocarcinoma (PRAD)
- Stomach and Esophageal Carcinoma (STES)
- Thyroid Carcinoma (THCA)
- Uterine Corpus Endometrial Carcinoma (UCEC)

Select Differential Expression Type

- Up regulated in Cancer versus Normal
- Down regulated in Cancer versus Normal

Submit

### Displaying BART Result for upregulated genes in BLCA

\*Hover over each column to display header definitions

TR	Wilcoxon ...	Wilcoxon ...	Z-score	max AUC	Relative r...	Irwin-Hal...
<a href="#">ESR1</a>	9.844	3.63e-23	1.861	0.733	0.031	<b>0.000133</b>
<a href="#">MBD2</a>	3.166	7.74e-4	2.074	0.721	0.04	<b>0.000294</b>
<a href="#">HDAC2</a>	3.558	1.87e-4	2.201	0.709	0.042	<b>0.000335</b>
<a href="#">ZNF143</a>	4.856	5.99e-7	1.831	0.716	0.044	<b>0.000389</b>
<a href="#">DROSHA</a>	2.412	7.94e-3	2.055	0.738	0.047	<b>0.00047</b>
<a href="#">KDM5B</a>	3.035	1.20e-3	1.964	0.718	0.048	<b>0.000492</b>
<a href="#">ZNF687</a>	2.395	8.31e-3	2.258	0.718	0.049	<b>0.000526</b>
<a href="#">GREB1</a>	2.987	1.41e-3	2.565	0.703	0.05	<b>0.000561</b>
<a href="#">TFDP1</a>	3.461	2.69e-4	1.482	0.739	0.052	<b>0.000651</b>
<a href="#">ASCL1</a>	3.3	4.83e-4	2.031	0.707	0.055	<b>0.000734</b>
<a href="#">ZNF592</a>	2.27	1.16e-2	2.87	0.708	0.055	<b>0.000749</b>
<a href="#">RAC3</a>	1.713	4.34e-2	2.596	0.725	0.062	<b>0.00106</b>
<a href="#">TET2</a>	2.535	5.63e-3	1.82	0.714	0.066	<b>0.00128</b>
<a href="#">NCAFG</a>	2.491	6.38e-3	3.014	0.694	0.067	<b>0.00132</b>
<a href="#">EZF6</a>	3.474	2.56e-4	1.887	0.694	0.073	<b>0.00175</b>

Right Click to Download...

- [Differentially Expressed Genes](#)
- [Adaptive Lasso File](#)
- [AUC File](#)
- [BART Results](#)

For a detailed explanation of BART results, please refer [here](#).

**Figure 2.** Example of BART Results Interface on BART Cancer. The radio button menus allow the user to select a reannotated TCGA cancer type and differential expression type. The BART result table then displays the list of putative transcriptional regulators for the cancer type and expression type selected. The table is displayed with the following columns: TR, name of transcriptional regulator; Wilcoxon test statistic, Wilcoxon rank-sum test comparing the set of association score from one TR with all the association scores. The higher the test score is, the more likely the TR regulates the input; Wilcoxon  $P$ -value, One-sided  $P$ -value of the Wilcoxon rank-sum test. The smaller the  $P$ -value is, the more significant the TR regulates the input; Z-score, Z-score indicates the deviation of the TR's Wilcoxon test score from its background model, which is generated from MSigDB gene sets. The background model contains the Wilcoxon test score for each TR and each gene set; Max AUC, Maximum association score of each TR in its datasets indicating the specific binding pattern of one dataset that has the highest correlation with the input; Relative Rank, Rank indicating average between Wilcoxon  $P$ -value, Z-score, and maximum association score and then dividing the absolute rank by the total number of TRs; and Irwin-Hall  $P$ -value, indicating the rank significance, the smaller the  $P$ -value is, the more significant the rank is. Below the table are links to download the tables and supplementary data files as well as a link to BARTweb for more information on BART results.

promoter and has been identified as a possible target for therapies (15). The BART results reflect this as TEAD4 is the highest ranked regulator (Figure 3A).

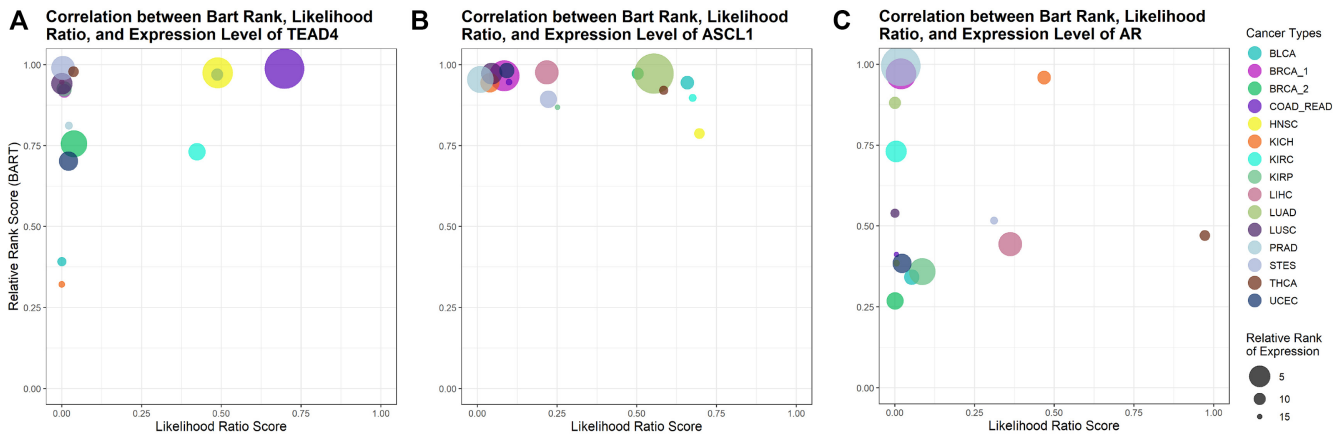
In both lung cancers, lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), ASCL1 ranked among the top five putative regulators (Figure 3B). It has been demonstrated that ASCL1 is required for tumorigenesis in pulmonary cells (16). Additionally, some ASCL1 targets were ranked among the top twenty regulators including SOX2 whose overexpression has been clearly described in lung cancers (17).

In prostate adenocarcinoma (PRAD), the top ranked factor is AR (Figure 3C). This nuclear receptor has been shown to promote prostate cell growth and survival when bound to dihydrotestosterone (DHT). Thus, when there is an AR mutation or overexpression, prostate cell growth is directly impacted (18). For this reason, androgen deprivation therapy has been used to suppress prostate tumorigenesis (19).

YAP1 is a transcriptional co-activator and its activity with TEAD transcription factors have proliferative effects (20). BART results for up-regulated genes in stomach and esophageal carcinoma (STES) rank YAP1 and TEAD4 among the top three ranked putative TRs. This is supported by a 2016 study that has shown YAP1 knockdown suppresses esophageal carcinoma and proposes YAP1 to be a target for gene therapy (21).

### TRs for down-regulated genes

Transcriptional regulators predicted from down-regulated genes can be associated with tumor suppressor functions. Here we highlight six factors from the top three ranked TRs predicted from the down-regulated genes in several cancer types: GATA-binding protein 4 (GATA4), repressor element 1 (RE-1)-silencing transcription factor (REST), forkhead box A1 (FOXA1), ESR1, hepatocyte nuclear factor 4 alpha (HNF4A) and AR.



**Figure 3.** Integrative visualization of BART rank, Cistrome Cancer likelihood ratio score, and relative expression level of transcriptional regulators across cancer types for three transcriptional regulators. The bubble plots combine information from BART rank, Cistrome Cancer likelihood ratio score, and expression level. Each cancer type is plotted as a bubble on the 1 - relative rank versus likelihood ratio axes. The top right quadrant contains cancer types in which the transcriptional regulators ranked high in likelihood score and high in BART rank. The size of the bubble corresponds to the rank (1–15) of its relative expression level using  $\log(\text{FPKM} + 0.1)$  from RNA-seq with the highest rank being the largest bubble, and the lowest rank being the smallest bubble. TEAD4 (A) and ASCL1 (B) show COAD\_READ and LUAD in the top right quadrant, respectively, with the largest bubble indicating that they are likely oncogenic regulators in the corresponding cancer type. AR (C) shows PRAD as the largest bubble and highest BART rank. Although it does not rank high in Cistrome Cancer likelihood ratio score, it is still likely to be a functional transcription factor in PRAD.

**Table 1.** The top three transcriptional regulators from BART results on up-regulated genes (left) and down-regulated genes (right). The regulators in bold are those that are discussed further as have literature-supported oncogenic activator or tumor suppressor functions. Cancer type abbreviations were following TCGA nomenclature.

Up-regulated Gene set BART Results				Down-regulated Gene set BART Results			
Cancer type	TR 1	TR 2	TR 3	Cancer type	TR 1	TR 2	TR 3
BLCA	ESR1	MBD2	<b>HDAC2</b>	BLCA	MEF2B	NEUROG2	MAML3
BRCA1	<b>ESR1</b>	GREB1	FOXA1	BRCA1	OTX2	SOX2	<b>GATA4</b>
BRCA2	MAX	POLR3D	TCF7L1	BRCA2	AR	<b>GATA4</b>	<b>REST</b>
COAD_READ	<b>TEAD4</b>	SMAD4	SMAD3	COAD_READ	<b>REST</b>	AR	MEF2B
HNSC	SMAD3	SIRT1	TEAD4	HNSC	AR	<b>FOXA1</b>	<b>ESR1</b>
KICH	ERG	ESR1	PR	KICH	<b>HNF4A</b>	RAD21	HNF1A
KIRC	SITR1	IKZF1	TAL1	KIRC	<b>AR</b>	HOXB13	FOXA1
KIRP	JUND	CEBPB	SMARCA4	KIRP	MAML3	RAD21	GATA4
LIHC	BRDU	ASCL1	POLR3D	LIHC	GATA2	TAL1	CEBPB
LUAD	SIRT1	<b>ASCL1</b>	HNF4A	LUAD	GATA2	TAL1	GATA4
LUSC	TP63	GRHL2	TP73	LUSC	GATA2	TAL1	LYL1
PRAD	<b>AR</b>	PIAS1	FOXA1	PRAD	OTX2	SIX2	TP63
STES	<b>TEAD4</b>	SMAD3	<b>YAP1</b>	STES	AR	REST	TBX5
THCA	BRDU	TEAD4	JUND	THCA	<b>AR</b>	OTX2	NANOG
UCEC	GRHL2	GREB1	ASCL1	UCEC	REST	CTCF	MAML3

In breast cancer, one of the top three ranked transcriptional regulators for both basal and luminal breast cancer’s down-regulated gene sets is GATA4. The dysregulation and aberrant expression of GATA4 has been correlated with tumor progression. While no direct tumor suppressive functions of GATA4 have been identified, it has been shown that restoration and even overexpression of GATA4 can impede the progression of breast tumors (22).

REST is ranked highly among many cancer’s BART results for down-regulated genes. This result is not surprising as REST has previously been identified as a tumor suppressor from genetic screens (23). In breast cancer specifically, REST was identified to have significantly decreased expression in tumor compared to normal samples and REST knockdown in MCF-7 cells resulted in an increase in cell proliferation and reduced sensitivity to anticancer drugs (24). Similar results were found in colorectal cancer cell

lines, further supporting its function as a tumor suppressor (25).

FOXA1 and ESR1, two of the top three ranked factors for head and neck cancer (HNSC) have been identified as tumor suppressors. First, second ranked FOXA1 was shown to suppress expression of miRNAs that were responsible for malignant behaviors in nasopharyngeal carcinoma, a subtype of head and neck squamous cell carcinoma (26). Second, third ranked ESR1 has been hypothesized to have tumor suppressive function in laryngeal squamous cell carcinoma (LSCC), another subcategory of head and neck cancer (27). It has also been shown that the removal of ESR1 results in a more aggressive LSCC and high ESR1 expression correlated to significantly improved survival (27).

HNF4A is identified as the top ranked regulator for kidney chromophobe (KICH). HNF4A has been associated

with diabetes and has more recently been associated with reduction of proliferation in kidney cells, specifically renal cell carcinoma (28). HNF4A works by positively regulating genes that are responsible for regulating proliferation, specifically ABAT and ALDH6A1 (29,30). By regulating these proliferation regulatory genes, HNF4A works as a tumor suppressor.

While AR was shown to be an oncogene, it has been shown to have the opposite effect in kidney chromophobe (KICH) and thyroid carcinoma (THCA). The loss of AR has been shown to correspond to increasing pathological grade of kidney cancers (31). Further, it was shown that high expression of AR was associated with better survival rates supporting a tumor suppressor role for AR in kidney cancer (32). Likewise, decreased expression of AR in thyroid carcinoma was associated with poor prognosis and more aggressive tumor behavior (33).

## DISCUSSION

BART Cancer web resource provides information about computational model-predicted transcriptional regulators in 15 different cancer types, by integrating TCGA gene expression profiling data with publicly available ChIP-seq data for over 900 factors. It provides insights to cancer transcriptional regulation in two aspects. First, for researchers interested in a particular cancer type, BART Cancer reports putative transcription factors that either activate oncogenic gene expression or regulate tumor-suppressive gene expression. Depending on up- or down-regulation of target genes and transcriptional activator or repressor function of the regulator, these putative regulators can have either oncogene or tumor-suppressor functions in this cancer. As many predicted transcription factors have been cross-referenced with literature, the novel factors in the prediction may guide further functional and mechanistic studies. Second, for researchers interested in a specific transcription factor or chromatin regulator, BART Cancer reports its relative activities across 15 cancer types and can give insights to which cancer types this factor might be functional in.

A few limitations of BART Cancer are worth noting. First, BART prediction was based on RNA-seq data from dozens to hundreds of TCGA samples and differentially expressed genes identified were shared across samples of each cancer type. The heterogeneity across samples were not taken into account. Also, cell heterogeneity could play a significant role especially in solid tumors, and the gene expression profiles could contain components from non-tumor cells such as tumor-infiltrating immune cells (4). As a result, BART predicted TRs might reflect cell composition changes rather than transcriptional regulation changes in cancer cells. Second, BART only utilizes publicly available ChIP-seq datasets for making predictions. There are over 1600 likely human transcription factors (34). Despite including more than 900 factors, there are still around 500 transcription factors that do not have high-quality ChIP-seq data and are absent in BART Cancer. The predicted TRs are far from complete. Third, discrepancies between BART rank and TR gene expression or likelihood ratio score were frequently found in TR plots. Possible reasons may include that differential transcription levels of a TR

gene are not necessarily correlated with different TR activities. Likelihood ratio scores were based on putative TR target genes, whose accuracy could be limited because of the co-expression assumption and various ChIP-seq data qualities. Users should be aware of such caveats when interpreting BART Cancer data. In summary, despite of having certain limitations, BART Cancer can provide useful information in cancer transcriptional regulation and can be a valuable data resource for the cancer research community.

## DATA AVAILABILITY

BART Cancer database resource is available at <http://bartcancer.org/>.

## ACKNOWLEDGEMENTS

The authors would like to thank Dr. Anindya Dutta for helpful discussions and Dr. Shenglin Mei for assistance on the Cistrome Cancer data.

*Author contributions:* C.Z. conceived the study. Z.V.T. performed the analysis, interpreted the results, curated the database, designed and implemented the web interface with supervision and assistance of Z.W. Z.V.T., Z.W. and C.Z. wrote the manuscript.

## FUNDING

National Institutes of Health [R35GM133712 to C.Z.].  
*Conflict of interest statement.* None declared.

## REFERENCES

1. Strausberg, R.L., Simpson, A.J.G., Old, L.J. and Riggins, G.J. (2004) Oncogenomics and the development of new cancer therapies. *Nature*, **429**, 469–474.
2. Bradner, J.E., Hnisz, D. and Young, R.A. (2017) Transcriptional addiction in cancer. *Cell*, **168**, 629–643.
3. Chang, K., Creighton, C.J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y.S.N. *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
4. Mei, S., Meyer, C.A., Zheng, R., Qin, Q., Wu, Q., Jiang, P., Li, B., Shi, X., Wang, B., Fan, J. *et al.* (2017) Cistrome Cancer: a web resource for integrative gene regulation modeling in cancer. *Cancer Res.*, **77**, e19–e22.
5. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. and Luscombe, N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
6. Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L. *et al.* (2017) Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.*, **45**, D658–D662.
7. Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.-H., Brown, M., Zhang, X., Meyer, C.A. *et al.* (2019) Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.*, **47**, gky1094.
8. Wang, Z., Civelek, M., Miller, C.L., Sheffield, N.C., Guertin, M.J. and Zang, C. (2018) BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics*, **34**, 2867–2869.
9. Ma, W., Wang, Z., Zhang, Y., Magee, N.E., Feng, Y., Shi, R., Chen, Y. and Zang, C. (2021) BARTweb: a web server for transcription factor association analysis. *NAR genom. bioinform.*, doi:10.1093/nargab/lqab022.
10. Feng, C., Song, C., Liu, Y., Qian, F., Gao, Y., Ning, Z., Wang, Q., Jiang, Y., Li, Y., Li, M. *et al.* (2019) KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Res.*, **48**, D93–D100.

11. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
12. Poyet, C., Jentsch, B., Hermanns, T., Schweckendiek, D., Seifert, H.-H., Schmidpeter, M., Sulser, T., Moch, H., Wild, P.J. and Kristiansen, G. (2014) Expression of histone deacetylases 1, 2 and 3 in urothelial bladder cancer. *BMC Clin. Pathol.*, **14**, 10.
13. Pinkerneil, M., Hoffmann, M.J., Deenen, R., Köhrer, K., Arent, T., Schulz, W.A. and Niegisch, G. (2016) Inhibition of class I histone deacetylases 1 and 2 promotes urothelial carcinoma cell death by various mechanisms. *Mol. Cancer Ther.*, **15**, 299–312.
14. Santo, I.D., McCartney, A., Migliaccio, I., Leo, A.D. and Malorni, L. (2019) The emerging role of ESR1 mutations in luminal breast cancer as a prognostic and predictive biomarker of response to endocrine therapy. *Cancer*, **11**, 1894.
15. Tang, J.-Y., Yu, C.-Y., Bao, Y.-J., Chen, L., Chen, J., Yang, S.-L., Chen, H.-Y., Hong, J. and Fang, J.-Y. (2018) TEAD4 promotes colorectal tumorigenesis via transcriptionally targeting YAP1. *Cell Cycle*, **17**, 1–23.
16. Borromeo, M.D., Savage, T.K., Kollipara, R.K., He, M., Augustyn, A., Osborne, J.K., Girard, L., Minna, J.D., Gazdar, A.F., Cobb, M.H. *et al.* (2016) ASCL1 and NEUROD1 reveal heterogeneity in pulmonary neuroendocrine tumors and regulate distinct genetic programs. *Cell Rep.*, **16**, 1259–1272.
17. Karachaliou, N., Rosell, R. and Viteri, S. (2013) The role of SOX2 in small cell lung cancer, lung adenocarcinoma and squamous cell carcinoma of the lung. *Transl. Lung Cancer Res.*, **2**, 172–179.
18. Tan, M.E., Li, J., Xu, H.E., Melcher, K. and Yong, E. (2015) Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol. Sin.*, **36**, 3–23.
19. Fujita, K. and Nonomura, N. (2018) Role of androgen receptor in prostate cancer: a review. *World J. Mens Heal.*, **36**, 288.
20. Sudol, M. (1994) Yes-associated protein (YAP65) is a proline-rich phosphoprotein that binds to the SH3 domain of the Yes proto-oncogene product. *Oncogene*, **9**, 2145–2152.
21. Zhao, S., Zhao, J., Li, X., Yang, Y., Zhu, D., Zhang, C., Liu, D. and Wu, K. (2016) Effect of YAP1 silencing on esophageal cancer. *Oncotarget Ther.*, **9**, 3137–3146.
22. Han, X., Tang, J., Chen, T. and Ren, G. (2019) Restoration of GATA4 expression impedes breast cancer progression by transcriptional repression of ReLA and inhibition of NF- $\kappa$ B signaling. *J. Cell. Biochem.*, **120**, 917–927.
23. Westbrook, T.F., Martin, E.S., Schlabach, M.R., Leng, Y., Liang, A.C., Feng, B., Zhao, J.J., Roberts, T.M., Mandel, G., Hannon, G.J. *et al.* (2005) A genetic screen for candidate tumor suppressors identifies REST. *Cell*, **121**, 837–848.
24. Lv, H., Pan, G., Zheng, G., Wu, X., Ren, H., Liu, Y. and Wen, J. (2010) Expression and functions of the repressor element 1 (RE-1)-silencing transcription factor (REST) in breast cancer. *J. Cell. Biochem.*, **110**, 968–974.
25. Coulson, J.M. (2005) Transcriptional regulation: cancer, neurons and the REST. *Curr. Biol.*, **15**, R665–R668.
26. Peng, Q., Zhang, L., Li, J., Wang, W., Cai, J., Ban, Y., Zhou, Y., Hu, M., Mei, Y., Zeng, Z. *et al.* (2020) FOXA1 suppresses the growth, migration, and invasion of nasopharyngeal carcinoma cells through repressing miR-100-5p and miR-125b-5p. *J. Cancer*, **11**, 2485–2495.
27. Verma, A., Schwartz, N., Cohen, D.J., Patel, V., Nageris, B., Bachar, G., Boyan, B.D. and Schwartz, Z. (2020) Loss of estrogen receptors is associated with increased tumor aggression in laryngeal squamous cell carcinoma. *Sci. Rep. UK*, **10**, 4227.
28. Lucas, B., Grigo, K., Erdmann, S., Lausen, J., Klein-Hitpass, L. and Ryffel, G.U. (2005) HNF4 $\alpha$  reduces proliferation of kidney cells and affects genes deregulated in renal cell carcinoma. *Oncogene*, **24**, 6418–6431.
29. Grigo, K., Wirsing, A., Lucas, B., Klein-Hitpass, L. and Ryffel, G.U. (2008) HNF4 $\alpha$  orchestrates a set of 14 genes to down-regulate cell proliferation in kidney cells. *Biol. Chem.*, **389**, 179–187.
30. Lu, J., Chen, Z., Zhao, H., Dong, H., Zhu, L., Zhang, Y., Wang, J., Zhu, H., Cui, Q., Qi, C. *et al.* (2020) ABAT and ALDH6A1, regulated by transcription factor HNF4A, suppress tumorigenic capability in clear cell renal cell carcinoma. *J. Transl. Med.*, **18**, 101.
31. Ha, Y.-S., Lee, G.T., Modi, P., Kwon, Y.S., Ahn, H., Kim, W.-J. and Kim, I.Y. (2015) Increased expression of androgen receptor mRNA in human renal cell carcinoma cells is associated with poor prognosis in patients with localized renal cell carcinoma. *J. Urol.*, **194**, 1441–1448.
32. Zhao, H., Leppert, J.T. and Peehl, D.M. (2016) A protective role for androgen receptor in clear cell renal cell carcinoma based on mining TCGA data. *PLoS One*, **11**, e0146505.
33. Chou, C.-K., Chi, S.-Y., Chou, F.-F., Huang, S.-C., Wang, J.-H., Chen, C.-C. and Kang, H.-Y. (2020) Aberrant expression of androgen receptor associated with high cancer risk and extrathyroidal extension in papillary thyroid carcinoma. *Cancer*, **12**, 1109.
34. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. and Weirauch, M.T. (2018) The human transcription factors. *Cell*, **172**, 650–665.