A Computational Pipeline for Studying Functional Genome Organization

By

Alexandra Hickman, Undergraduate Department of Biomedical Engineering
Zachary Thomas, Undergraduate Department of Biomedical Engineering
Chongzhi Zang, Center for Public Health Genomics, University of Virginia
Zhenjia Wang, Center for Public Health Genomics, University of Virginia

Word Count: 2534
Number of Figures: 3
Number of Tables: 3
Number of Equations: 0
Number of Supplements: 0
Number of References: 41

Approved: Chongzhi Zang 4 Digitally signed by Chongzhi Zang 4 Date: 2021.05.07 17:48:07 -04'00' Date:_____

Chongzhi Zang, Center for Public Health Genomics, University of Virginia

A Computational Pipeline for Studying Functional Genome Organization

Alex Z. Hickman[a,b], Zachary V. Thomas[a,b], Zhenjia Wang[b], Chongzhi Zang[a,b,1]

[a] University of Virginia Biomedical Engineering
[b] University of Virginia Center for Public Health Genomics
[1] Correspondence: zang@virginia.edu

Abstract

The number and complexity of transcriptional regulators (TR) make identification of these proteins that function within cancer difficult. However, identification of TR can lead to better targets for cancer therapies and provide a more concrete understanding of cancer genomics for other researchers. Current methods to measure genome-wide TF have limitations due to accuracy and scalability and pipelines, such as Chicdiff and Selfish, are only effective when analyzing binding patterns of a chosen transcriptional regulator. To improve on current identification methods, we developed a computational pipeline for analyzing tumor Hi-C data by incorporating existing and novel computational methods, BART3D and HiC-Pro. We applied the pipeline to colorectal cancer (CRC) as the epigenetics of CRC are still not well understood and it is the third most common cancer diagnosed in the US. The computational method HiC-Pro was utilized to produce normalized contact maps from raw, paired-end fastq data. These contact maps were used within BART3D, which leverages over 7,000 transcriptional regulator binding profiles from the public domain to infer those regulators that are associated with genome-wide differential chromatin interactions. The top 20 TRs associated with either increasing or decreasing chromatin interactions were identified in CRC. Literature review was conducted on the ranked list of TRs and revealed supporting evidence of these TRs within colorectal and other cancers.

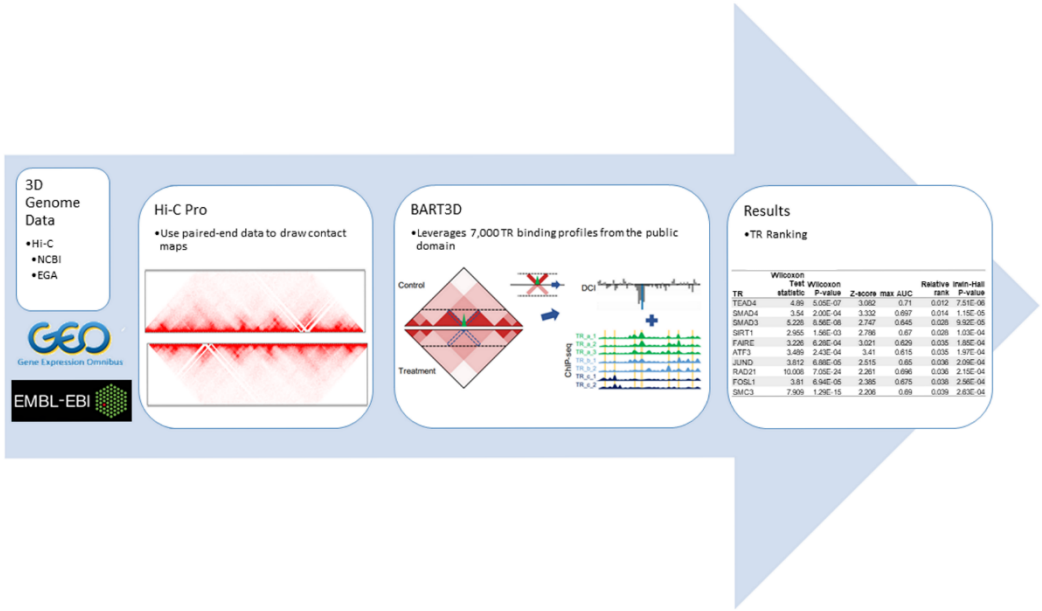
Keywords: Hi-c, Pipeline, Colorectal Cancer, BART3D, HiC-Pro, Transcriptional Regulator

Introduction

Cancer is a complex disease involving multiple layers of dysregulation in the genome, epigenome, and higher-order genome organization. A key regulator of genome organization is CCCTC-binding factor (CTCF), which has critical functions in 3-dimensional (3D) genome organization, especially looping[1]. It has been shown that CTCF binding can be induced by other transcription factors to regulate oncogenic gene expression[2,3]. A number of computational frameworks have been developed to study functional regulators like CTCF at these different levels. Previously, Binding Analysis for Regulation of Transcription (BART) and BART3D have been developed[4,5]. These computational tools rely on publicly available ChIP-seq data sets to rank putative transcriptional regulators that control differential expression (BART) or differential chromatin interactions (BART3D) in a treatment group versus control. Where BART takes in a gene list, scored region set, or ChIP-seq reads, BART3D relies on interaction matrices as input, like those produced by HiC-Pro. HiC-Pro is a computational tool capable of taking in raw fastq files from chromatin conformation capture methods and processing them to raw and normalized contact maps.

Currently, pipelines exist that perform differential analysis of Hi-C data. These include Chicdiff and Selfish. Chicdiff is a powerful computational tool for analyzing Hi-C data and the detection of differential interactions. The authors validated Chicdiff on human monocytes and confirmed associations between promoter regions and gene expression. Much like Chicdiff, Selfish is a novel tool for measuring reproducibility of Hi-C replicates and differential chromatin interactions. Chicdiff and Selfish both accomplish differential analysis in a locus-to-locus manner. While these methods are effective at analyzing

A



B

TR	STATISTIC	PVALUE	ZSCORE	MAX_AUC	RE_RANK	IRWIN_HALL_PVALUE
ZNF486	1.868	3.085e-02	67.681	0.508	0.108	**0.005644**
PHOX2B	1.660	4.842e-02	40.715	0.526	0.079	**0.002226**
ZNF808	1.601	5.471e-02	39.593	0.519	0.095	**0.003846**
ZNF490	1.088	1.382e-01	23.641	0.498	0.199	0.035350
HOTAIR	2.299	1.074e-02	22.749	0.504	0.112	**0.006288**
CHAT	1.342	8.986e-02	21.948	0.496	0.188	0.029900
SATB1	0.757	2.244e-01	21.803	0.500	0.211	0.042280
ZNF586	1.650	4.948e-02	18.293	0.524	0.084	**0.002678**
KDM1B	1.472	7.050e-02	18.018	0.511	0.123	**0.008289**
ZNF485	0.905	1.827e-01	17.762	0.493	0.226	0.052270

C

```
bash-4.2$./hic3d.sh -h

Usage: computational pipeline for analyzing tumor Hi-C data incorporating existing and novel computational methods
 (HiC-Pro and BART3D) to predict putative oncogenic transcriptional regulators
This tool requires HiC-Pro 3.0.0 and BART3D

usage: hic3d -T TREATMENT -C CONTROL -t TREATMENT OUTPUT -c CONTROL OUTPUT -F TREATMENT CONFIG FILE -f CONTROL CON
FIG FILE -s SPECIES [-b] [-h]

OPTIONS INPUTS

  -T|--treatment TREATMENT INPUT : input data folder; Must contain a folder per sample with input files
  -C|--control  CONTROL INPUT : input data folder; Must contain a folder per sample with input files
  -t|--tout TREATMENT OUTPUT : output folder for treatment results
  -c|--cout CONTROL OUTPUT : output folder for control results
  -F|--tconfig TREATMENT CONFIG FILE : configuration file for HiC-Pro for treatment files
  -f|--cconfig CONTROL CONFIG FILE : configuration file for HiC-Pro for control files
  -s|--species SPECIES : specify species for BART3D. Please choose from  hg38 or mm10
  [-b|--bout]: BART3D output folder. Default bart3d_output
  [-h|--help]: help
```

Fig. 1. Pipeline Overview. (A) Schematic representation of the pipeline with steps that include data selection, contact map creation, and transcriptional regulator prediction. (B) Sample HTML output. This is the main output of the pipeline. It consists of a table of transcriptional regulators and their corresponding statistics. Each column is sortable. In this figure, the table is sorted by decreasing Z-score. The most important statistic is the Irwin-Hall P-value (IHPV). Those with IHPVs less than 0.01 are marked in bold and the entire row is highlighted to indicate significance. (C) Command line help menu for the pipeline.

short, specific regions of the genome, they are only effective when analyzing binding patterns of a chosen transcriptional regulator. In order to identify potential transcriptional regulators for experimental confirmation, the differential analysis would need to be genome-wide while maintaining high resolution. To this end, we developed hic3d, a computational pipeline that combines the function of constructing interaction matrices, whole genome differential interaction identification, and putative transcriptional regulator prediction. We applied hic3d to

3D colorectal cancer genome data to predict putative functional transcriptional regulators involved in colorectal tumorigenesis. An overview of the pipeline can be seen in Figure 1.

Results

Pipeline Overview

The usage of the pipeline is twofold. First, to generate interaction matrices of both a treatment and control group. Second, to use the previously generated matrices to calculate differential interaction and predict those transcriptional regulators that are responsible for areas of increased and decreased interaction in treatment compared to normal. In order to achieve the first goal, we utilized HiC-Pro and to achieve the second goal we utilized BART3D. Details of these computational tools can be found in the Materials and Methods section.

The main input of the pipeline is two folders: one containing raw fastq reads for the treatment condition and one containing raw fastq reads for the control condition. There are several other parameters that need to be specified but they all directly relate to the two main folder inputs. After the pipeline processes the data, the main output is a ranked transcriptional regulator table. There is one table ranking the putative regulators responsible for increased genome interaction in treatment versus control and one table ranking the putative regulators responsible for decreased genome interaction in treatment versus control. These tables come in two forms, text file and HTML file. The two different file types contain the same information. The text file is provided for ease of use in other projects and extraction of data whereas the HTML file is used for viewing. A sample HTML output is shown in Figure 2.

Colorectal Cancer Application

To demonstrate the power of the pipeline, we applied it to colorectal cancer. Colorectal cancer is the third leading cause of cancer death in the world[6]. It is estimated that there will be 104,610 new cases of colon cancer and 43,340 cases of rectal cancer diagnosed in 2020[7].

Data Collection

Hi-C data from three colorectal cancer samples and two normal colon samples were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and the Encyclopedia of DNA Elements (ENCODE). A summary of the data sets collected can be found in Table 1.

Table 1. Summary of collected colorectal cancer data. Data collected from the public domain that were used to validate the pipeline.

Principle Investigator	Sample Type	Sample Source	Restriction Enzyme	Accession Number
Suhas Rao	Cancer	HCT116	MboI	GSE104334
Suhas Rao	Cancer	HCT116	MboI	GSE104334
Seolkyoung Jung	Cancer	HCT116	DpnII	GSE98119
Bing Ren	Normal	Transverse colon	MboI	ENCSR504OTV
Bing Ren	Normal	Transverse colon	MboI	ENCSR295BDK

Replicate Testing

While the three cancer datasets weren't exactly replicates, the sample source for all three were the same HCT116. Thus, we expected that they would yield similar results. Figure 2A displays Venn diagrams comparing the top ranked transcriptional regulators (Irwin-Hall $p < 0.01$) among the cancer types for each normal sample. These Venn diagrams illustrate significant overlap between the first Rao sample and the Jung sample as well as some overlap between the two Rao samples. Figure 2B was drawn to further investigate the correlation between the cancer samples. Here, it is further evident that the first Rao and Jung samples are significantly correlated whereas the second Rao samples didn't seem to display any significant correlation. This correlation supports the eventual combination of the first Rao and Jung samples. While the second Rao sample didn't align with the first Rao sample in any metric, the two datasets were marked as replicates in GEO and will be used as such.

To explore whether or not the two normal samples should be combined in replicate, the correlation between the two for each cancer sample was measured. Figure 2C displays these correlations. For each cancer sample both the increased and decreased interaction rankings appear to be significantly correlated. For this reason and the fact that they were marked as replicates in ENCODE, they will be used as replicates in the analysis.

Finally, two different combinations of replicates were tested. First, just the Rao samples were tested in replicate versus the normal samples in replicate. Then all of the cancer samples (Rao and Jung) were tested in replicate versus the normal samples in replicate. The increased interaction correlation between just the Rao samples and all of the HCT116 replicates, as seen in Figure 2D, had a correlation coefficient of 0.993. The decreased interaction correlation of these samples also had a strong correlation of 0.997. Since the pipeline ranked the transcriptional regulators nearly the same in each replicate combination, and therefore had high correlation values, we will use all of the collected cancer samples in replicate and all of the collected normal samples in replicate for further analysis.

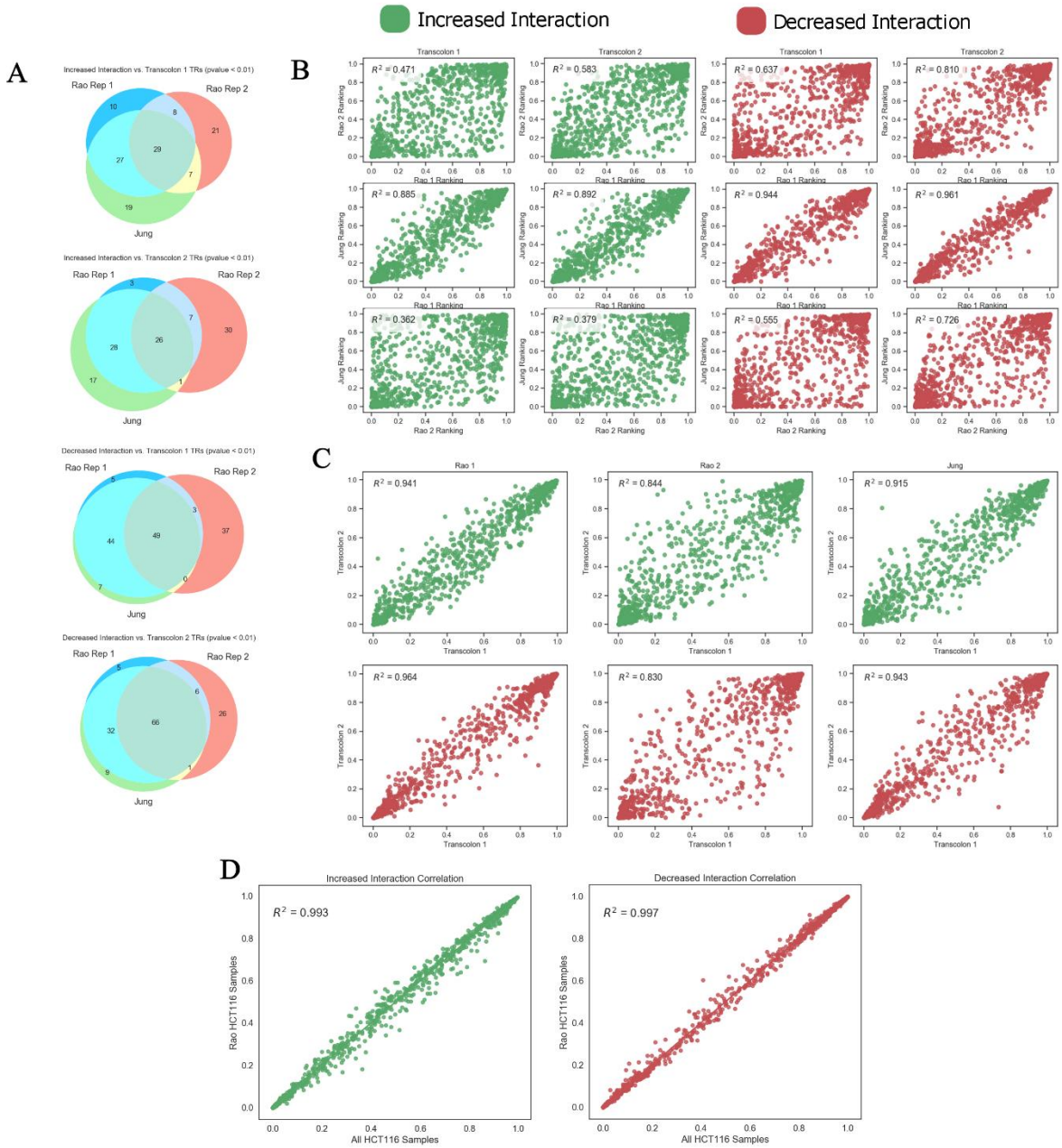


Fig. 2. Replicate Testing. (A) Common transcriptional regulators. After the pipeline was applied to the individual datasets, the top regulators with Irwin-Hall p-value less than 0.01 were collected. Displayed here are the number of TRs shared among each group. (B) Correlation between cancer group rankings. These scatter plots were drawn to assess the correlation between the cancer groups' complete rankings. Rao 1 and Jung appear to be the best correlated. (C) Correlation between normal group rankings. These scatter plots were drawn to assess the correlation between the normal groups' complete rankings. All of these plots serve as supporting evidence for the combination of the normal samples to be used in replicate. (D) Replicate configuration correlations.

Colorectal Cancer Results

Table 2 displays the top twenty ranked transcriptional regulators using all of the cancer samples in replicate and all of the normal samples in replicate as input to the pipeline. Using a p-value of less than 0.01 to indicate significance, 68 transcriptional regulators were marked as significant predictions in increased interaction and 102 transcriptional regulators were marked as significant predictions in decreased interaction. In the following analysis, only the top twenty ranked transcriptional regulators of each type, increased and decreased, were used for literature validation.

Table 2. Colorectal Cancer Results. Displayed here at the top twenty transcriptional regulators in increased interaction and decreased interaction categories. Additionally, those that had supporting literature have their references noted. If a reference is in bold, it means that reference specifically studied colorectal cancer while those that are not bolded have literature supporting their role in cancer in general.

Increased Interaction			Decreased Interaction		
Transcription Factor	Irwin-Hall P-value	Literature Citation	Transcription Factor	Irwin-Hall P-value	Literature Citation
EZH2	0.00004515	8	AR	0.000008963	**24**
SIRT1	0.00005741	**9-10**	HOXB13	0.00001059	**25-28**
SUZ12	0.00008471	11	FOXA1	0.00003478	29-30
GTF2I	0.0001111	**12**	GATA2	0.00003871	31
SP1	0.0001153	**13-14**	FOXA2	0.00004079	32
MBD2	0.0001474	**15**	TRIM28	0.00004515	33
ZTA	0.0001681	16	GATA3	0.00006573	31
KLF9	0.0002351	17	TLE3	0.00006867	**34**
RNF2	0.000242	**18**	DUX4	0.00006867	35
JARID2	0.0002707	19	PIAS1	0.00008819	**36**
TET2	0.0003796		MAML3	0.00009543	
KLF1	0.0006113		ATF4	0.0001195	
YBX1	0.000664	20	GATA6	0.0001239	31
BRDU	0.0009053		GATA4	0.0001575	**31**
REST	0.001082		OTX2	0.0002633	37
INTS11	0.001324		CEBPB	0.0002707	
SP2	0.001503	**21**	CDX2	0.0003435	**38**
ZNF436	0.001775	22	NUP98	0.0006113	
CNOT3	0.001827	23	HOXA9	0.0006113	
TRIM25	0.001854		EP300	0.0006373	

Discussion

Literature Validation

Increased Interaction

Several of the ranked increased interaction transcription factors (Figure 3A) had supporting literature evidence of the promotion of colorectal cancer specifically. SIRT1, a NAD+ dependent class III deacetylase, has been shown to enhance tumorigenesis in colorectal carcinoma patients[9]. GTF2I promotes cell cycle progression by the downregulation of TGFβR2[12]. The zinc finger protein SP1 plays an important role in the regulation of gene expression[14]. Another member of the SP protein family, SP2 is also involved in the regulation of proliferation, apoptosis, and differentiation[39]. Increase in MBD2 has been linked to a promotion in tumor growth and metastasis in colorectal cancer[15]. RNF2, a critical component of polycomb repressive complex 1, is upregulated in many kinds of cancers and an elevation of this protein is associated with poor cancer prognosis[18].

Decreased Interaction

The ranked decreased interaction transcriptional regulators with supporting literature research for colorectal cancer were AR, HOXB13, TLE3, PIAS2, GATA4, and CDX2 (Figure 3B). ARs are involved in the differentiation and progression of colorectal cancer tissues and are one of the most important sex hormones[24]. Decreased HOXB13 expression is associated with poorer differentiation and lymph node metastasis in gastric cancer[26]. TLE has been shown to suppress colorectal cancer proliferation through the inhibition of MAPK and AKT signaling pathways[34]. PIAS1 regulates the interferon-gamma signaling pathway which affects tumor development and biology[36]. GATA, a zinc finger protein controls the development of tissues through the activation and suppression of transcription[31]. CDX2 has been shown to suppress Wnt/β-catenin signaling and thereby inhibit colon cancer proliferation[41].

Limitations

Despite including over 900 transcriptional regulators, there are approximately 1600 proteins presumed to function as transcriptional regulators in the human genome. Additionally, the pipeline only utilizes ChIP-seq data sets that are publicly available. For these reasons, the predictions are potentially less accurate as there are many regulators that do not have publicly available ChIP-seq data.

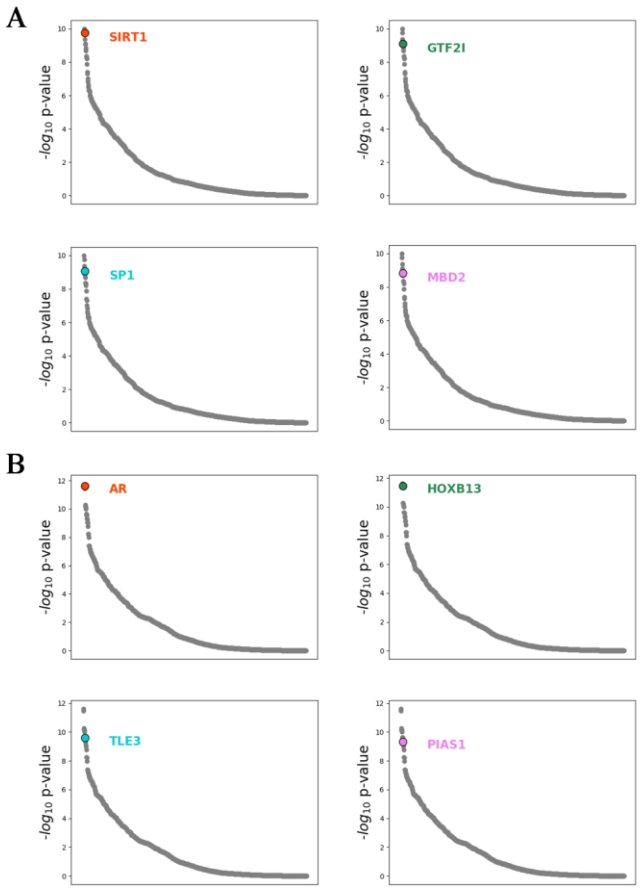


Fig. 3. Top ranked regulators. (A) Increased interaction regulators. Those highlighted are discussed as having supporting evidence. (B) Decreased interaction regulators. Those highlighted are discussed as having supporting evidence.

Future Perspectives

Currently, this pipeline has only been applied to colorectal cancer using the publicly available datasets. However, there are many Hi-C datasets publicly available for various disease states. For example, Table 3 summarizes which cancer types have both tumor and corresponding normal Hi-C samples presently available in the Gene Expression Omnibus. Additionally, this pipeline may be applied to various disease states where data for normal conditions and disease conditions are available.

Materials and Methods

Data Collection

Hi-C data from three human colorectal cancer HCT116 and two normal cells lines from the transverse colon were collected from NCBI GEO. All cancer samples

Table 3. Summary of Available Data.

Tissue	Cancer and Normal Available?	Normal	Tumor
Blood	Yes	Peripheral Blood	HL-60, HAP-1
Bone	No		
Brain	No		
Breast	Yes	MCF10A	MCF7, T47D
CD4	No		
Cervix	No		
Colon	Yes	Patient Data	HCT116
Fibroblast	No		
Kidney	Yes	HEK293	ACHN, CAKI-2
Liver	Yes	Patient Data	HepG2
Lung	Yes	IMR90	HCI-H460, A549
Muscle	No		
Pancreas	Yes	Patient Data	PANC1
Prostate	Yes	RWPE1	LNCaP
Skin	Yes	NHEK	SK-MEL-5

are accessible through the National Center for Biotechnology Information's Gene Expression Omnibus (NCBI GEO). Fastq-dump, as provided by the SRA Toolkit, was used to download the data. The normal data was stored in the ENCODE database and was downloaded using the command line utility wget. After running quality control metrics using FastQC to ensure acceptability of the published data, the raw fastq files were used as input to the proposed pipeline. Between all six transcriptional regulator ranking results from preliminary data, there was substantial agreement among the twenty top ranked regulators. This indicates that the samples could be combined into two groups, cancer and normal. Detailed information regarding accession numbers of collected samples can be found in Table 1.

HiC-Pro

The collected raw data in paired-end fastq format was converted to normalized contact matrices with a resolution of 5k base pairs using HiC-Pro[40]. During the workflow of HiC-Pro, reads only uniquely aligned with the reference genome are assigned to a restriction fragment. Each read pair was assigned to a restriction fragment. Invalid ligation products were filtered and single genome-wide interaction maps were created. Normalization within HiC-Pro was applied to eliminate Hi-C systematic bias. The HiC-Pro pipeline is optimized to run on a single computer or cluster, which was ideal for a general 3D analysis pipeline. The resulting chromosomal contact maps were used within BART3D to infer transcriptional regulators.

BART3D

Computational method BART3D was utilized to infer transcriptional regulators associated with genome-wide differential chromatin interactions. Hi-C contact maps from normal and cancerous data sets were used as input.

Chromosomes were scanned to generate differential chromatin interaction profiles (DCI) by comparing interactions between the two inputs[4]. The BART algorithm was applied to associate public ChIP-seq data with the DCI profiles and two ranked lists of transcriptional regulators associated with either increasing or decreasing chromatin interactions were generated.

Pipeline Design and Construction

The pipeline was written using bash script as BART3D and HiC-Pro were both provided as command line tools. Additionally, python was used to implement a text file to HTML file conversion. Both text files and HTML files were provided for ease of use and ease of visualization, respectively.

Literature Review

The top twenty ranked transcriptional regulators were the focus of the literature searches. The following search terms were used to find *in vivo* colorectal cancer: 1) "TRANSCRIPTIONAL REGULATOR colorectal cancer", 2) "TRANSCRIPTIONAL REGULATOR tumorigenesis colorectal cancer", and 3) "TRANSCRIPTIONAL REGULATOR tumor suppressor colorectal cancer".

End Matter

Author Contributions and Notes

C.Z. and Z.W designed research, A.Z.H. and Z.V.T. performed research, Z.V.T. wrote software, A.Z.H. and Z.V.T. analyzed data; and A.Z.H. and Z.V.T. wrote the paper.

The authors declare no conflict of interest.

References

1. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. 15, 234–246 (2014).
2. Fang, C. et al. Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation. Genome Biol 21, (2020).
3. Schwartz, M. et al. Genomic retargeting of p53 and CTCF is associated with transcriptional changes during oncogenic HRas-induced transformation. Communications Biology 3, 696 (2020)
4. Wang, Z. et al. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. Bioinformatics 34, 2867–2869 (2018).
5. Wang, Z., Zhang, Y. & Zang, C. BART3D: Inferring transcriptional regulators associated with differential chromatin interactions from Hi-C data. bioRxiv 2020.08.19.258095 (2020)Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 14, 89–103 (2019).
6. Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 14, 89–103 (2019).
7. Siegel, R. L. et al. Colorectal cancer statistics, 2020. CA: A Cancer Journal for Clinicians 70, 145–164 (2020).
8. Yamagishi, M. & Uchimaru, K. Targeting EZH2 in cancer therapy. Current Opinion in Oncology 29, 375–381 (2017).
9. Chen, X. et al. High levels of SIRT1 expression enhance tumorigenesis and associate with a poor prognosis of colorectal carcinoma patients. Scientific Reports 4, 7481 (2014).
10. Shen, T. et al. Ube2v1-mediated ubiquitination and degradation of Sirt1 promotes metastasis of colorectal cancer by epigenetically suppressing autophagy. J Hematol Oncol 11, 95 (2018).
11. San, K., Horita, M., Ganapathy, A., Chinnadurai, G. & Ezekiel, U. R. Deregulated expression of microRNA-200b/c and SUZ12, a Polycomb repressive complex 2 subunit, in chemoresistant colorectal cancer cells. Genes Cancer 8, 673–681 (2017).
12. Nambara, S. et al. GTF2IRD1 on chromosome 7 is a novel oncogene regulating the tumor‐suppressor gene TGFβR2 in colorectal cancer. Cancer Sci 111, 343‐355 (2020).
13. Song, B. et al. Targeting FOXA1-mediated repression of TGF-β signaling suppresses castration-resistant prostate cancer progression. J Clin Invest 129, 569–582 (2019).
14. Guo, Z. et al. Sp1 upregulates the four and half lim 2 (FHL2) expression in gastrointestinal cancers through transcription regulation. Mol Carcinog 49, 826–836 (2010).
15. Yuan, K. et al. Decreased Levels of miR-224 and the Passenger Strand of miR-221 Increase MBD2, Suppressing Maspin and Promoting Colorectal Tumor Growth and Metastasis in Mice. Gastroenterology 145, 853-864.e9 (2013).
16. Lan, Y.-Y. et al. Epstein-Barr Virus Zta Upregulates Matrix Metalloproteinases 3 and 9 That Synergistically Promote Cell Invasion In Vitro. PLOS ONE 8, e56121 (2013).
17. Bagati, A. et al. KLF9-dependent ROS regulate melanoma progression in stage-specific manner. Oncogene 38, 3585–3597 (2019).
18. Wei, F. et al. Ring finger protein 2 promotes colorectal cancer progression by suppressing early growth response 1. Aging (Albany NY) 12, 26199–26220 (2020).
19. Zhu, X.-X. et al. Jarid2 is essential for the maintenance of tumor initiating cells in bladder cancer. Oncotarget 8, 24483–24490 (2017).
20. Roßner, F. et al. Uncoupling of EGFR–RAS signaling and nuclear localization of YBX1 in colorectal cancer. Oncogenesis 5, e187 (2016).
21. Phan, D. et al. Identification of Sp2 as a Transcriptional Repressor of Carcinoembryonic Antigen-Related Cell Adhesion Molecule 1 in Tumorigenesis. CAN. RES. 64, 3072–3078 (2004).
22. Shang, Y., Li, Y., Zhang, Y. & Wang, J. ZNF436 promotes tumor cell proliferation through transcriptional activation of BCL10 in glioma. Biochemical and Biophysical Research Communications 515, 572–578 (2019).
23. Shirai, Y.-T. et al. CNOT3 targets negative cell cycle regulators in non-small cell lung cancer development. Oncogene 38, 2580–2594 (2019).
24. Xia, T. et al. Androgen receptor gene methylation related to colorectal cancer risk. Endocrine Connections 8, 979–987 (2019).
25. Ewing, C. M. et al. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med 366, 141–149 (2012).
26. Sui, B.-Q., Zhang, C.-D., Liu, J.-C., Wang, L. & Dai, D.-Q. HOXB13 expression and promoter methylation as a candidate biomarker in gastric cancer. Oncology Letters 15, 8833–8840 (2018).

27. Jung, C. et al. HOXB13 is downregulated in colorectal cancer to confer TCF4-mediated transactivation. Br J Cancer 92, 2233–2239 (2005).
28. Edwards, S. et al. Expression analysis onto microarrays of randomly selected cDNA clones highlights HOXB13 as a marker of human prostate cancer. British Journal of Cancer 92, 376–381 (2005).
29. Jin, H.-J., Zhao, J. C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. Nature Communications 5, 3972 (2014).
30. Kim, J. et al. FOXA1 inhibits prostate cancer neuroendocrine differentiation. Oncogene 36, 4072–4080 (2017).
31. Zheng, R. & Blobel, G. A. GATA Transcription Factors and Cancer. Genes Cancer 1, 1178–1188 (2010).
32. Basseres, D. S. et al. Frequent downregulation of the transcription factor Foxa2 in lung cancer through epigenetic silencing. Lung Cancer 77, 31–37 (2012).
33. Cui, J. et al. TRIM28 protects CARM1 from proteasome-mediated degradation to prevent colorectal cancer metastasis. Science Bulletin 64, 986–997 (2019).
34. Yang, R.-W. et al. TLE3 represses colorectal cancer proliferation by inhibiting MAPK and AKT signaling pathways. J Exp Clin Cancer Res 35, (2016).
35. Bury, M. et al. NFE2L3 Controls Colon Cancer Cell Growth through Regulation of DUX4, a CDK1 Inhibitor. Cell Reports 29, 1469-1481.e9 (2019).
36. Coppola, D., Parikh, V., Boulware, D. & Blanck, G. Substantially reduced expression of PIAS1 is associated with colon cancer development. J Cancer Res Clin Oncol 135, 1287–1291 (2009).
37. Luo, J. et al. DMBX1 promotes tumor proliferation and regulates cell cycle progression via repressing OTX2-mediated transcription of p21 in lung adenocarcinoma cell. Cancer Letters 453, 45–56 (2019).
38. Baba, Y. et al. Relationship of CDX2 Loss with Molecular Features and Prognosis in Colorectal Cancer: Implications for Clinical and Pathology Practice. Clin Cancer Res 15, 4665–4673 (2009).
39. Zhu, Y. et al. Sp2 promotes invasion and metastasis of hepatocellular carcinoma by targeting TRIB3 protein. Cancer Med 9, 3592–3603 (2020).
40. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology* 16, 259 (2015).
41. Yu, J. et al. CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/β-catenin signaling via transactivation of GSK-3β and Axin2 expression. Cell Death Dis 10, 26 (2019).