

ELMER v2

An R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles

Tiago Chedraoui Silva
RECOMB-CBB

2018-04-19



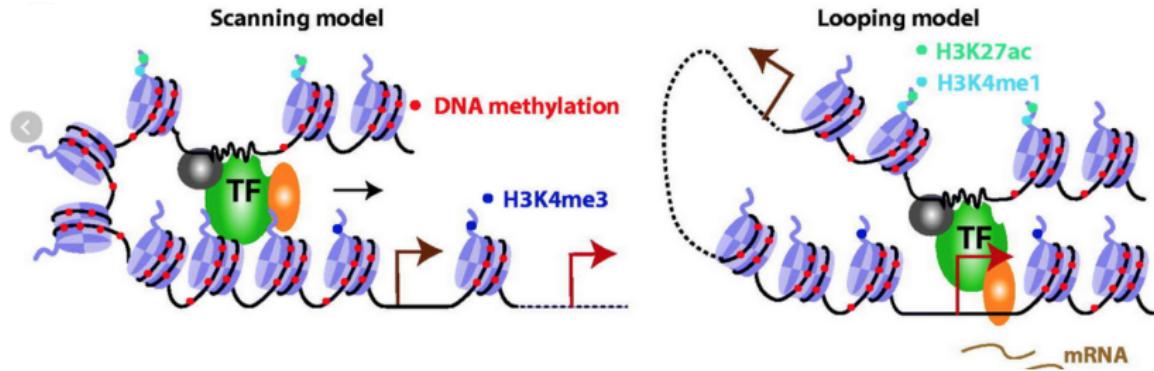
CEDARS-SINAI®



Overview

- ① Introduction
- ② Methods: ELMER algorithm
- ③ Case of study: TCGA-BRCA
- ④ Conclusion

Enhancer-mediated gene regulation

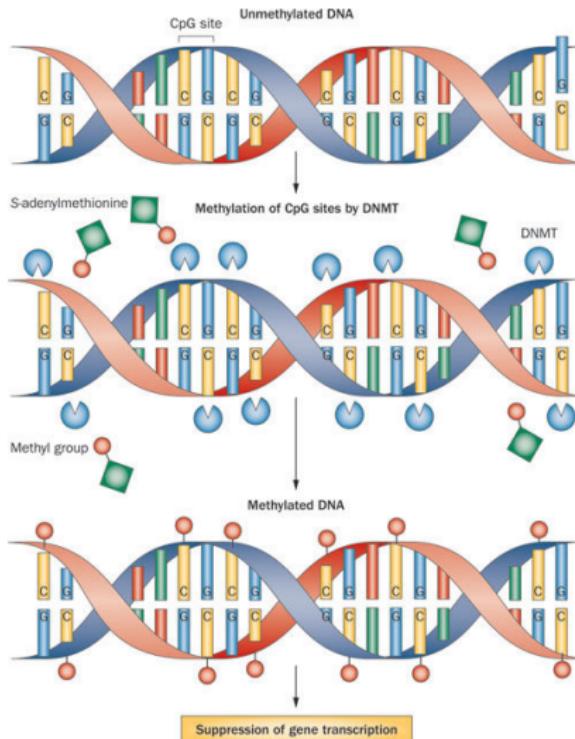


Source: Yao et al. Genome Biology (2015)

Enhancer-mediated gene regulation

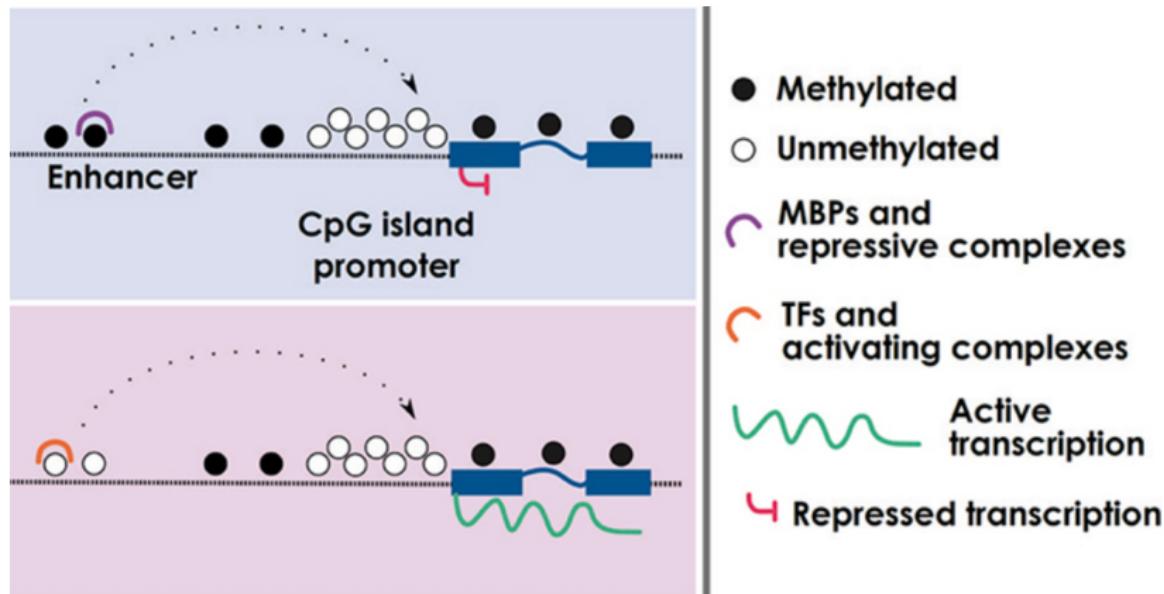
- 73% of the tested distal elements do not link to the nearest gene (Sanyal et al., 2012)
 - 40% of the enhancers involved in loops do not interact with the TSS of the nearest gene (Li et al., 2012),
 - one-third of the distal interactions were not directed to the promoter of the nearest gene (Mifsud et al., 2015),
 - 85% of tumor-specific enhancers that could be linked to the expression of a nearby gene skipped the nearest gene (Yao et al., 2015).

Epigenetics alterations - DNA methylation



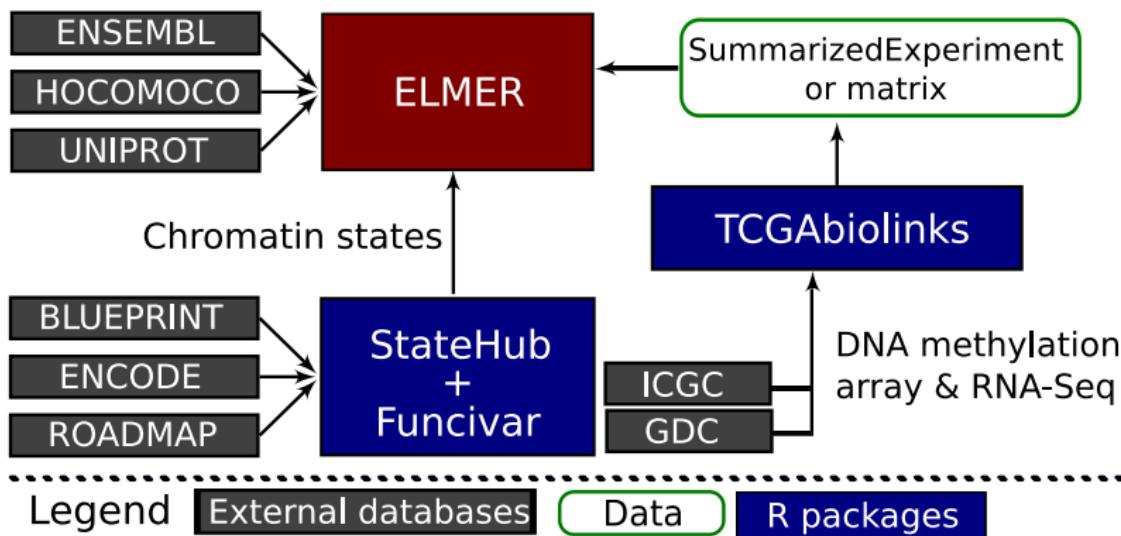
Source: Koch MW, et al. Nat Rev Neurol. (2013)

Enhancer-mediated gene regulation



Source: Carrio et al. Frontiers in aging neuroscience (2015)

ELMER v.2: Enhancer Linking by Methylation/Expression Relationship

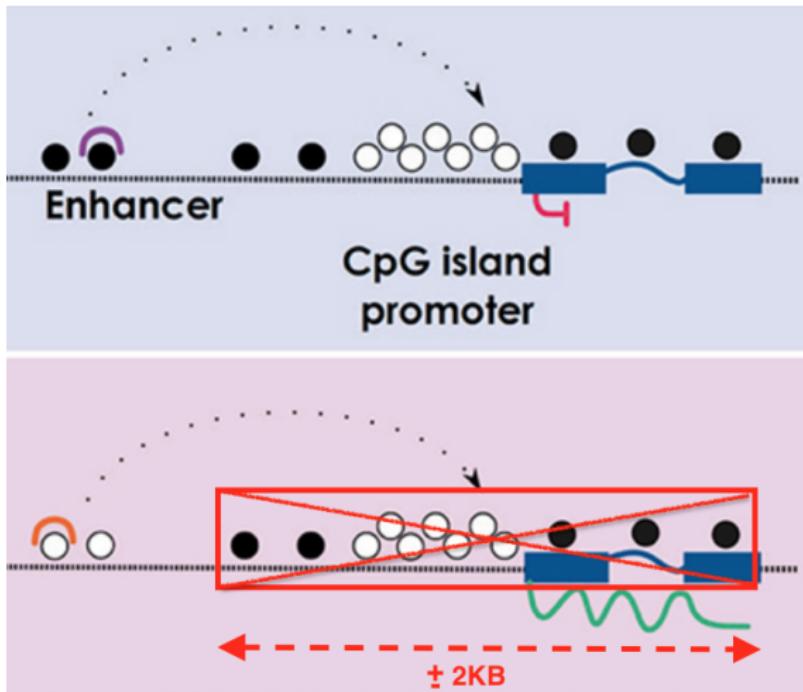


Algorithm

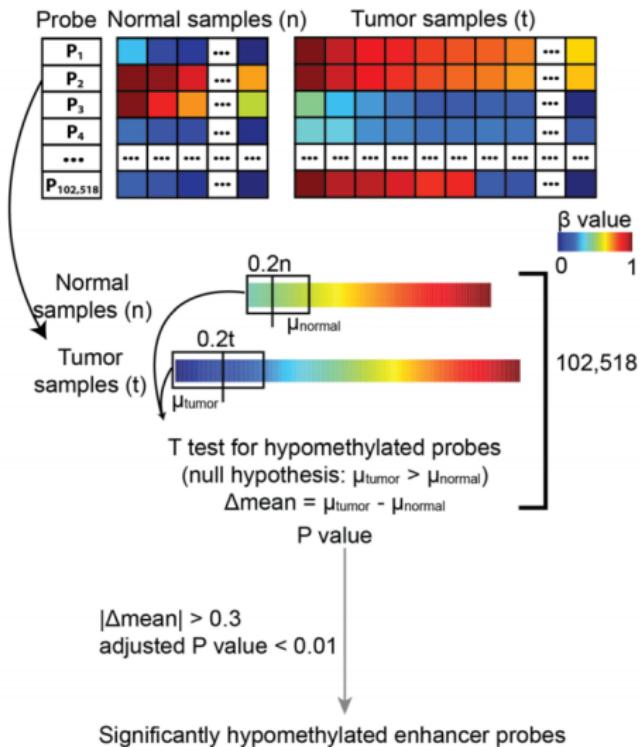
Steps

- ① Identify distal probes on HM450K/EPIC.
 - ② Identify distal probes with significantly different DNA methylation level in group 1 compared to group 2.
 - ③ Identify putative target genes for differentially methylated distal enhancer probes.
 - ④ Identify enriched motifs for the distal probes which are significantly differentially methylated and linked to a putative target gene.
 - ⑤ Identify regulatory TFs whose expression associate with DNA methylation at motifs.

Step 1: Identify distal probes



Step 2: Differentially methylated distal probes



Source: Yao et al. Genome Biology (2015)

Groups U and M definition in (*un*)supervised mode

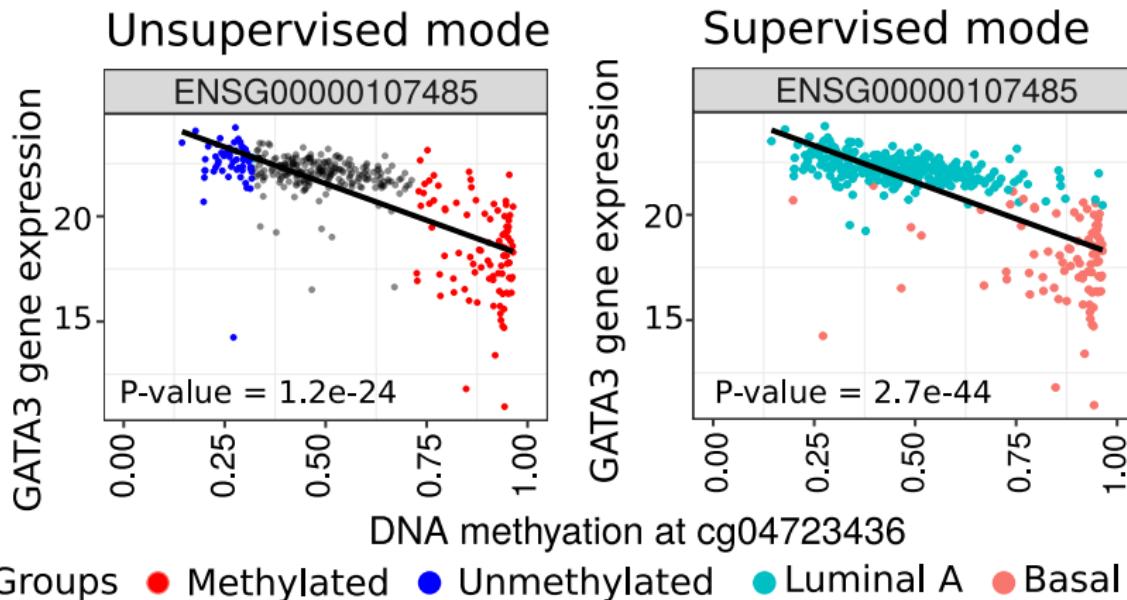
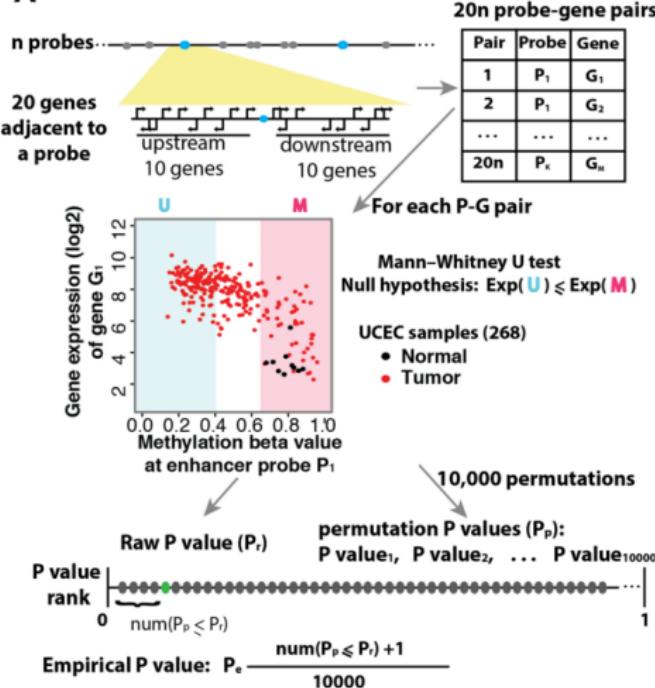
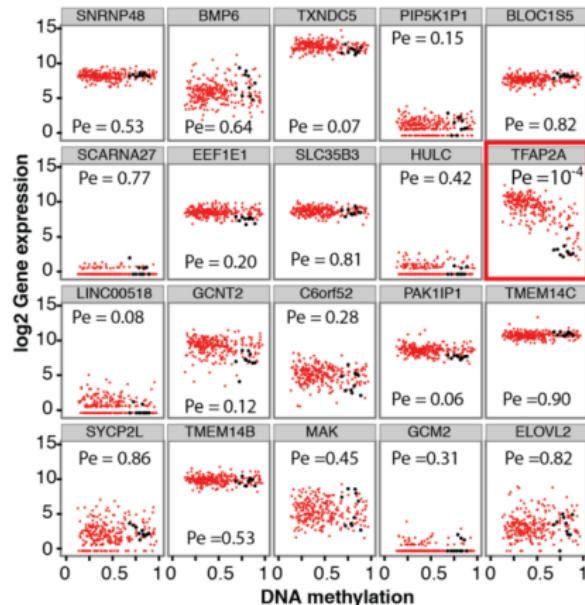


Figure: A: *unsupervised mode*; when `minSubgroupFrac` argument is set to 40%, the methylated group is defined as the highest quintile and the unmethylated group as the lowest quintile; B: *supervised mode*; methylated and unmethylated group are defined as one of the known molecular subtypes.

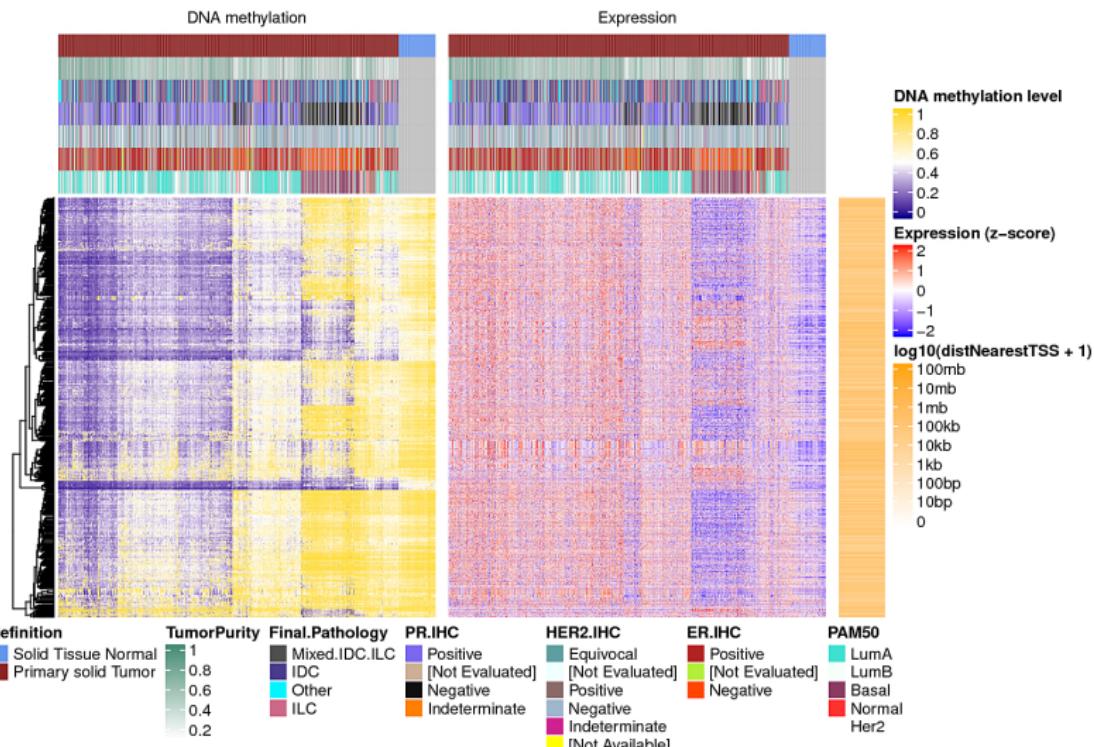
Step 3: Identification of putative target gene(s)

A**B**

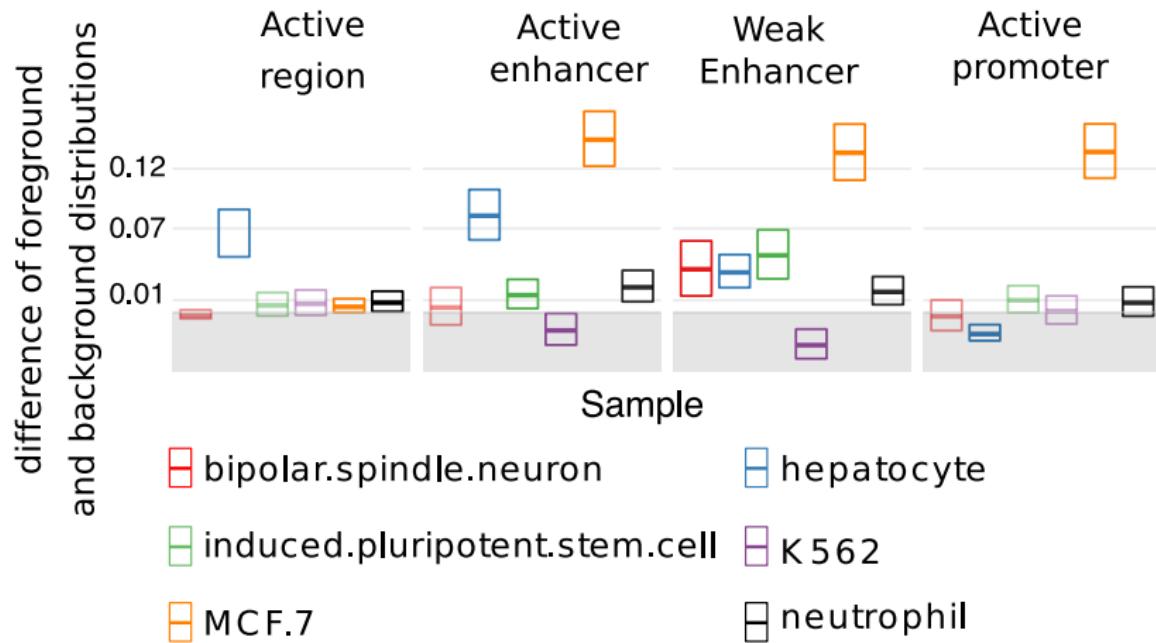
Source: Yao et al. Genome Biology (2015)

Step 3: Probe-target gene pairs inferred

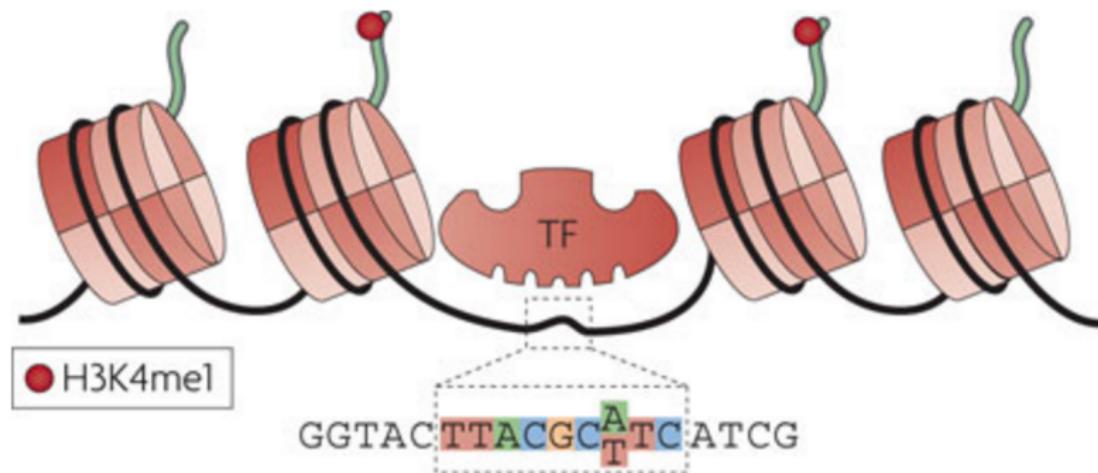
Correspondence between probe DNA methylation and distal gene expression



Chromatin state enrichment analysis



Step 4: Motif enrichment analysis



Nature Reviews | Genetics

Hawkins RD, et al. Next-generation genomics: an integrative approach. Nature Reviews Genetics (2010)

Step 4: TF motifs source

HOCOMOCO Home ▾ Human TFs ▾ Mouse TFs ▾ Tools ▾ Downloads ▾ Help Search: Human

Switch to CORE collection Reset Select Columns Get CSV HUMAN_mono_motifs.tsv PWMs for HUMAN transcription factors (full)

Model	LOGO	Transcription factor	TF family	TF subfamily
		FOX		
FOXA1_HUMAN.H11MO.0.A	T _g T _T T _T A _C _T _A _T _A	FOXA1 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXA[3.3.1.1]
FOXA2_HUMAN.H11MO.0.A	T _g T _T T _T A _C _T _A _T _A	FOXA2 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXA[3.3.1.1]
FOXA3_HUMAN.H11MO.0.B	T _g T _T T _T A _C _T _A _T _A	FOXA3 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXA[3.3.1.1]
FOXB1_HUMAN.H11MO.0.D	~A _A T _T T _T A _C A _T A _T	FOXB1 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXB[3.3.1.2]
FOXC1_HUMAN.H11MO.0.C	T _G T _T T _T A _C T _A _T _A	FOXC1 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXC[3.3.1.3]
FOXC2_HUMAN.H11MO.0.D	~T _T G _T _T _T _G _C _A _G	FOXC2 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXC[3.3.1.3]
FOXD1_HUMAN.H11MO.0.D	~T _T G _T T _T A _C _T _A	FOXD1 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXD[3.3.1.4]
FOXD2_HUMAN.H11MO.0.D	~G _T A _A T _A T _T _T _A _T _A	FOXD2 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXD[3.3.1.4]
FOXD3_HUMAN.H11MO.0.D	T _G T _T T _T _C _T _A	FOXD3 (GeneCards)	Forkhead box (FOX) factors[3.3.1]	FOXD[3.3.1.4]

HOCOMOCO v11 (<http://hocomoco11.autosome.ru/human/mono?full=true>), Accessed: 26-12-2017

Step 4: Motif enrichment analysis

Objective

Evaluate the enrichment of transcription factors in certain genomic regions.

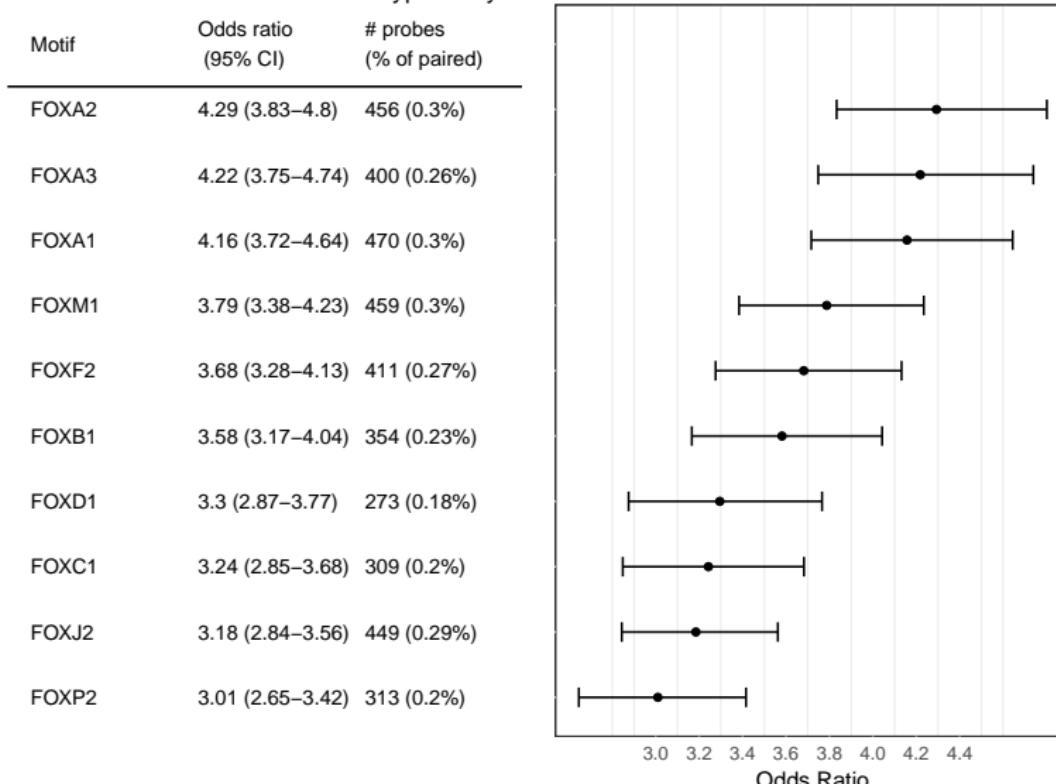
- ① Perform motif matching of transcription factors in probes regions (window $\pm 250\text{bp}$). Performed using HOMER (Hypergeometric Optimization of Motif EnRichment) with HOCOMOCO motifs.
- ② Evaluate which transcription factors are more likely to occur in those regions than in background regions using Fisher's exact test with FDR correction.

Fisher's exact test

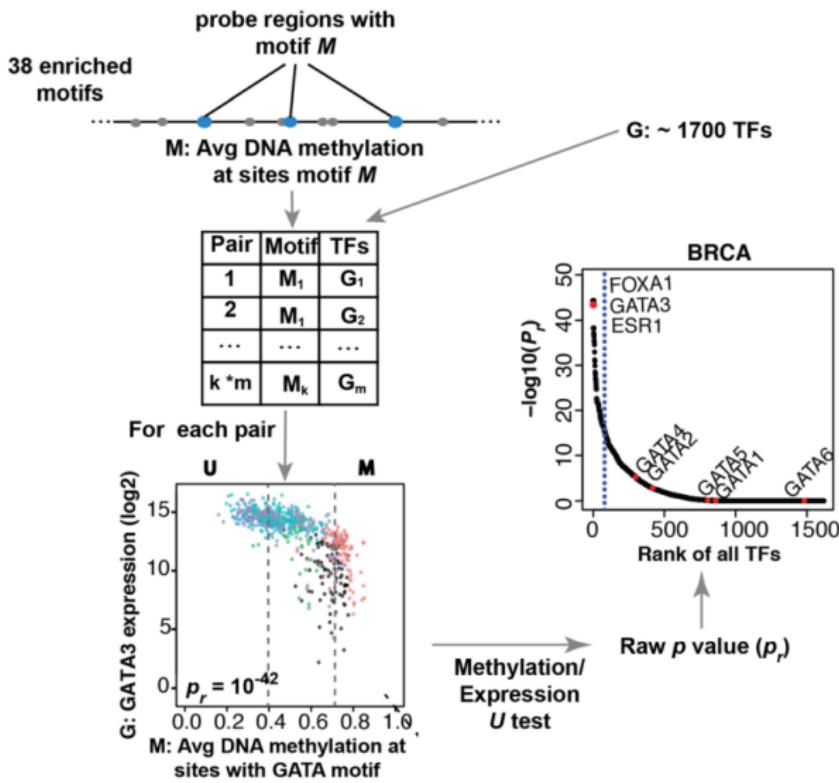
- a: nb of input regions with a match for TF motif.
- b: nb of input regions with no match for TF motif.
- c: nb of background regions with a match for TF motif.
- d: nb of background regions with no match for TF motif.

Step 4: Motif enrichment analysis

Probes hypomethylated in LumB vs Basal



Step 5: Identification of master regulator TF



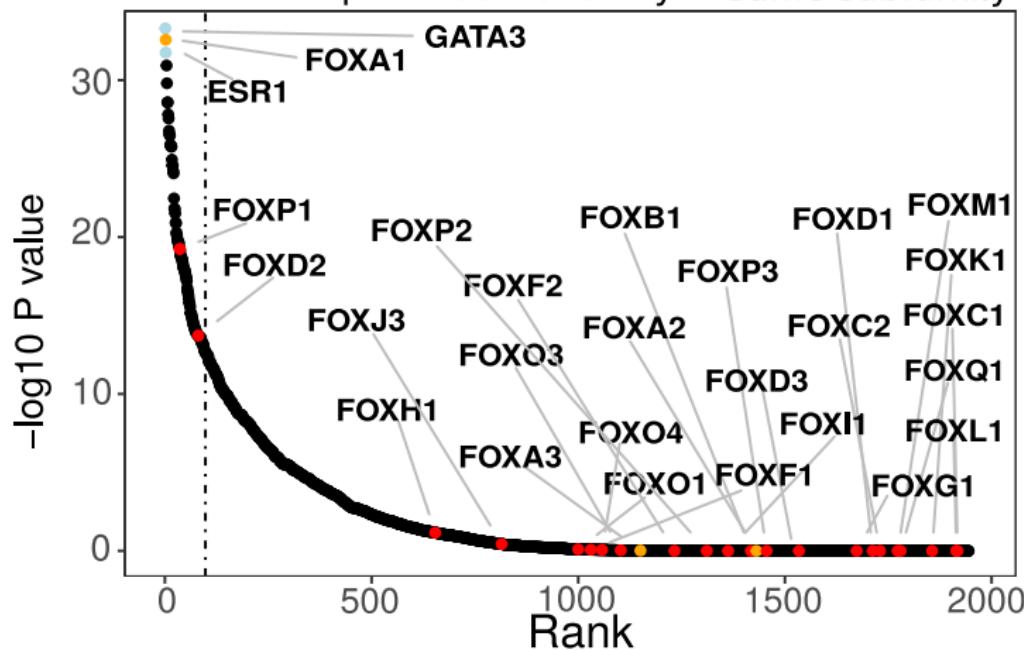
Source: Yao et al. Genome Biology (2015).

TFs ranking for the FOXA3 motif

TF classification (Motif: FOXA2)

Probes hypomethylated in LumB vs Basal

- None
- Top 3
- Same family
- Same subfamily



Step 5: Master Regulator TF table

motif	OR	potential.TF.family	pvalue
All	All	All	All
FOXA2_HUMAN.H11MO.0.A	4.29	FOXA1	3.29e-33
FOXA2_HUMAN.H11MO.0.A	4.29	FOXD2	1.94e-14
FOXA2_HUMAN.H11MO.0.A	4.29	FOXP1	5.85e-20
FOXA3_HUMAN.H11MO.0.B	4.22	FOXA1	3.29e-33
FOXA3_HUMAN.H11MO.0.B	4.22	FOXD2	1.94e-14
FOXA3_HUMAN.H11MO.0.B	4.22	FOXP1	5.85e-20
FOXA1_HUMAN.H11MO.0.A	4.16	FOXA1	3.29e-33
FOXA1_HUMAN.H11MO.0.A	4.16	FOXD2	1.94e-14
FOXA1_HUMAN.H11MO.0.A	4.16	FOXP1	5.85e-20
FOXM1_HUMAN.H11MO.0.A	3.79	FOXA1	3.29e-33

Showing 1 to 10 of 256 entries

Previous 1 2 3 4 5 ... 26 Next

Case study: TCGA Breast Invasive Carcinoma (BRCA)

Table: Summary of the available samples in TCGA for BRCA

Group	Samples w/ methylation (450K)	Samples w/ gene ex- pression (FPKM-UQ)	Samples w/ both
Primary solid Tumor	791	1102	778
Solid Tissue Normal	96	113	83

Table: Number of samples of the molecular subtypes of breast cancer

Molecular subtype	Number of samples
Basal	85
Her2	34
LumA	288
LumB	117
Normal-like	22

	Unsupervised	Supervised			
TF	Tumor (vs Normal)	LumB (vs basal)	Basal (vs LumB)	Basal (vs HER2)	HER2 (vs Basal)
<i>AR</i>	X	X			
<i>EMX1</i>	X	X			
<i>ESR1</i>	X	X			
<i>GATA3</i>	X	X			
<i>LMX1B</i>	X	X			
<i>NR2E3</i>	X	X			
<i>PBX1</i>	X	X			
<i>RARA</i>	X	X			
<i>ZNF467</i>	X	X			
<i>FOXA1</i>	X	X			X
<i>FOXP1</i>		X			X
<i>HOXB3</i>					X
<i>HOXC10</i>					X
<i>HOXC11</i>					X
<i>BCL11A</i>			X	X	
<i>E2F3</i>			X	X	
<i>ETV6</i>			X	X	
<i>KLF5</i>			X	X	
<i>SOX8</i>			X	X	
<i>SOX9</i>			X	X	
<i>SOX11</i>			X		
<i>ZIC1</i>			X	X	

Master regulator TF: molecular known subtypes

The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression

Jonathan H. Shepherd^{1,3}, Ivan P. Uray³, Abhijit Mazumdar³, Anna Tsimelzon², Michelle Savage³, Susan G. Hilsenbeck², Powel H. Brown^{1,3}

FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer

C Gong,^{1,2,6} K Fujino,^{1,3,6} L J Monteiro,¹ A R Gomes,¹ R Drost,⁴ H Davidson-Smith,⁵ S Takeda,³ U S Khoo,² J Jonkers,⁴ D Sproul,⁵ and E W-F Lam^{1,*}

negative breast cancer cell lines to regain hormonal sensitivity.⁴¹ In addition to promoting mammary luminal phenotype, FOXA1 might also have a more direct role in repressing the basal breast cancer phenotype. It has been shown that FOXA1 also inhibits the transcription of basal-type associated genes such as CD58, ANXA1, JAG1 and SOX9, whereas the loss of FOXA1 leads to the derepression of these basal genes.¹³ These findings together highlight a critical role of FOXA1 in maintaining the luminal and

GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland.

Kouros-Mehr H¹, Slorach EM, Sternlicht MD, Werb Z.

⊕ Author information

GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility

Main differences between ELMER old version (v.1) and the new version (v.2)

Features	ELMER Version 1	ELMER Version 2
Primary data structure	mee object (custom data structure)	MAE object (Bioconductor data structure)
Auxiliary data	Manually created	Programmatically created
Number of human TFs	1,982	2,014 (UniProt database)
Number of TF motifs	145	771 (HOCOMOCO v11 database)
TF classification	78 families	82 families and 331 subfamilies (TFClass database, HOCOMOCO)
Analysis performed	Normal vs tumor samples	Group 1 vs group 2
Statistical grouping	Unsupervised only	Unsupervised or supervised using labeled groups
TCGA data source	The Cancer Genome Atlas (TCGA)	The NCI's Genomic Data Commons (GDC)
Genome of reference	GRCh37 (hg19)	GRCh37 (hg19)/GRCh38 (hg38)
DNA methylation platforms	HM450	EPIC and HM450
Graphical User Interface (GUI)	None	TCGAbiolinksGUI
Automatic report	None	HTML summarizing results
Annotations	None	StateHub

Acknowledgments

Cedars-Sinai Hospital Center

- Benjamin P. Berman
 - Simon Coetzee
 - Dennis Hazelett
 - Nicole Gull
 - Dechen Lin



CEDARS-SINAI®

University of São Paulo

- Houtan Noushmehr



Thank You!