

ELMER 2.0

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

Tiago Chedraoui Silva
RECOMB-CBB

University of São Paulo / Cedar-Sinai

2018-04-09





Overview

- 1 Introduction
- 2 Objectives
- 3 Methods: Algorithms and tools
- 4 Analysis
- 5 Conclusion

Enhancer Linking by Methylation/Expression Relationship

ELMER 2.0

1

Bioinformatics

doi:10.1093/bioinformatics/xxxxx

Advance Access Publication Date: Day Month Year

Systems Biology

OXFORD

Systems Biology

ELMER 2.0: An R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles

Tiago Chedraoui Silva^{1,2}, Simon G Coetzee², Lijing Yao³, Nicole Yeager², Dennis J Hazelett², Houtan Noushmehr⁵, De-Chen Lin^{4,*}, Benjamin P Berman^{2,4,*}

¹Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil,

²Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA, USA

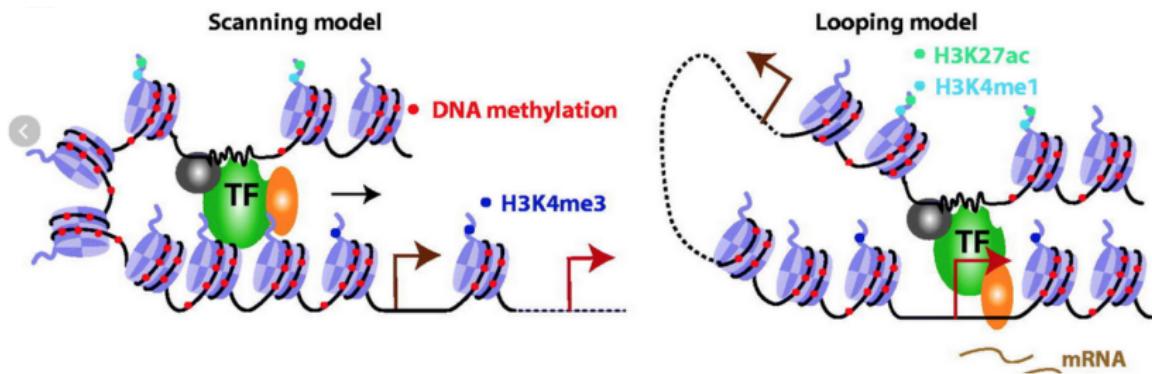
³Roche Sequencing Solutions, Belmont, CA, USA

⁴ Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA and

⁵ Department of Neurosurgery, Henry Ford Hospital, Detroit, MI, USA

* To whom correspondence should be addressed.

Enhancer-mediated gene regulation

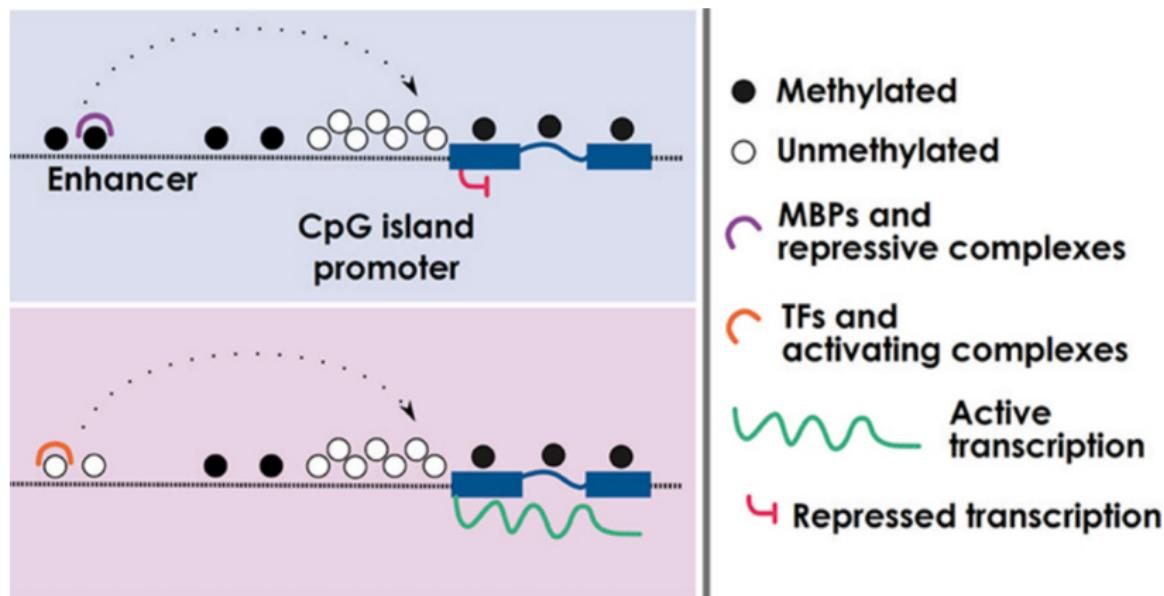


Source: Yao et al. Genome Biology (2015)

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).

Enhancer-mediated gene regulation



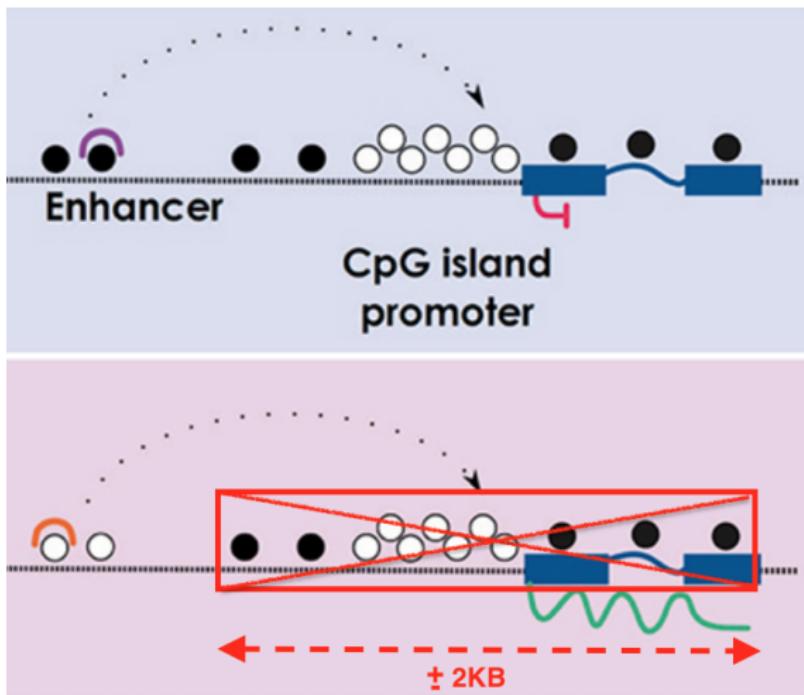
Source: Carrio et al. Frontiers in aging neuroscience (2015)

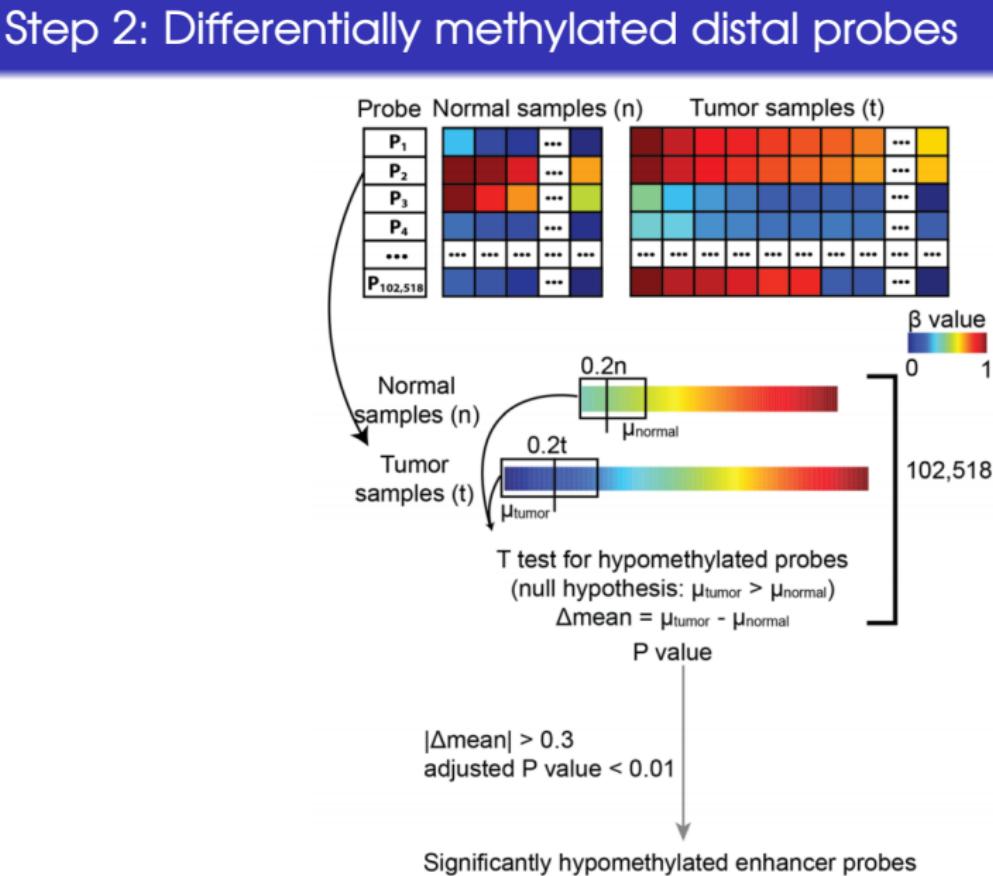
Algorithm

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

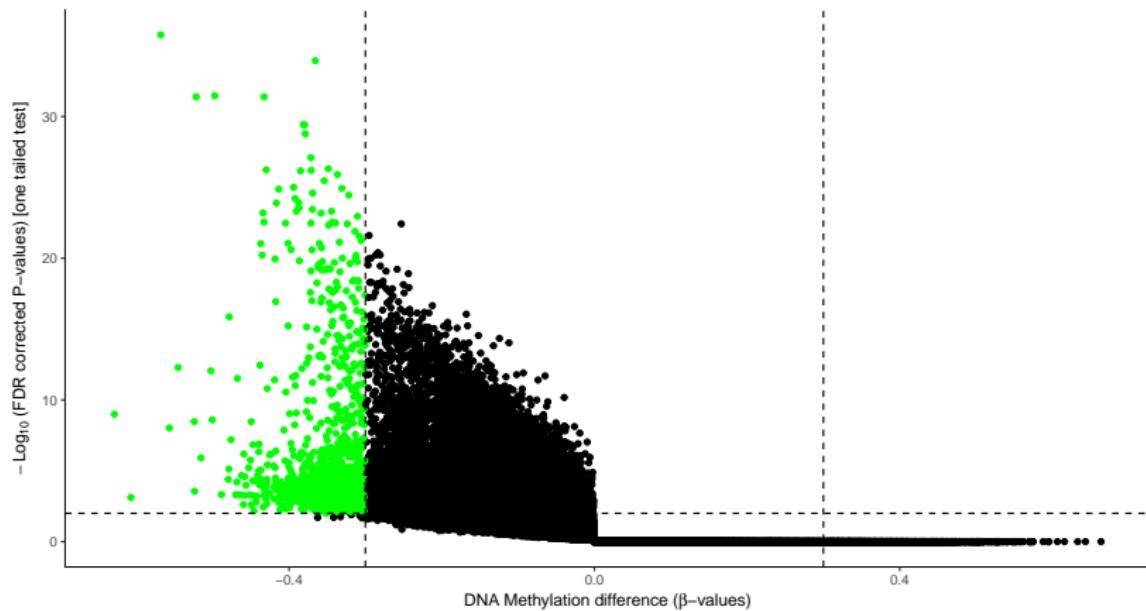




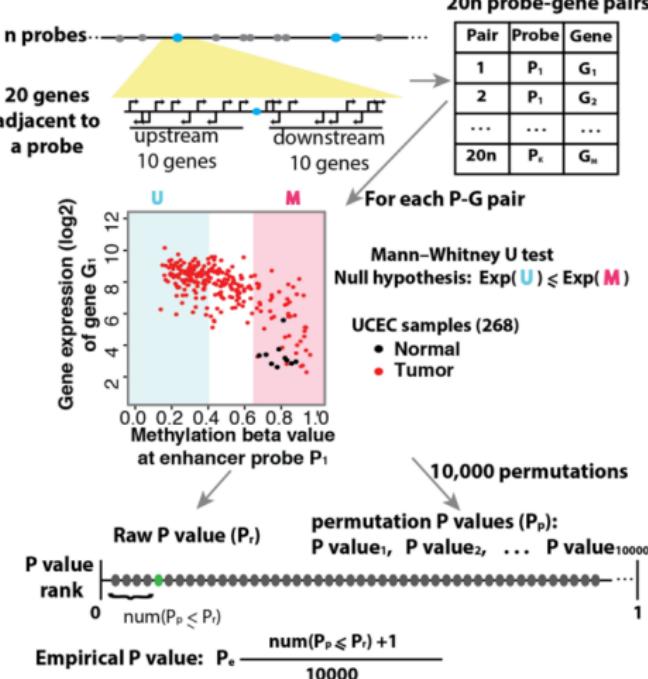
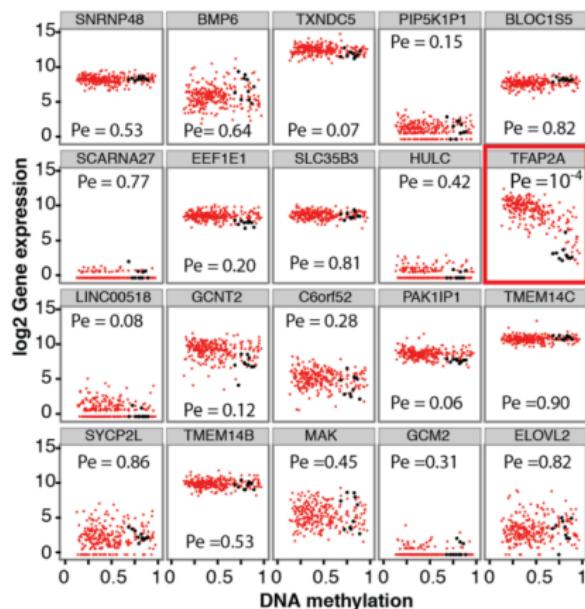
Step 2: Differentially methylated distal probes

Volcano plot – Probes hypomethylated in esad vs normal

- Not Significant
- Hypomethylated in esad



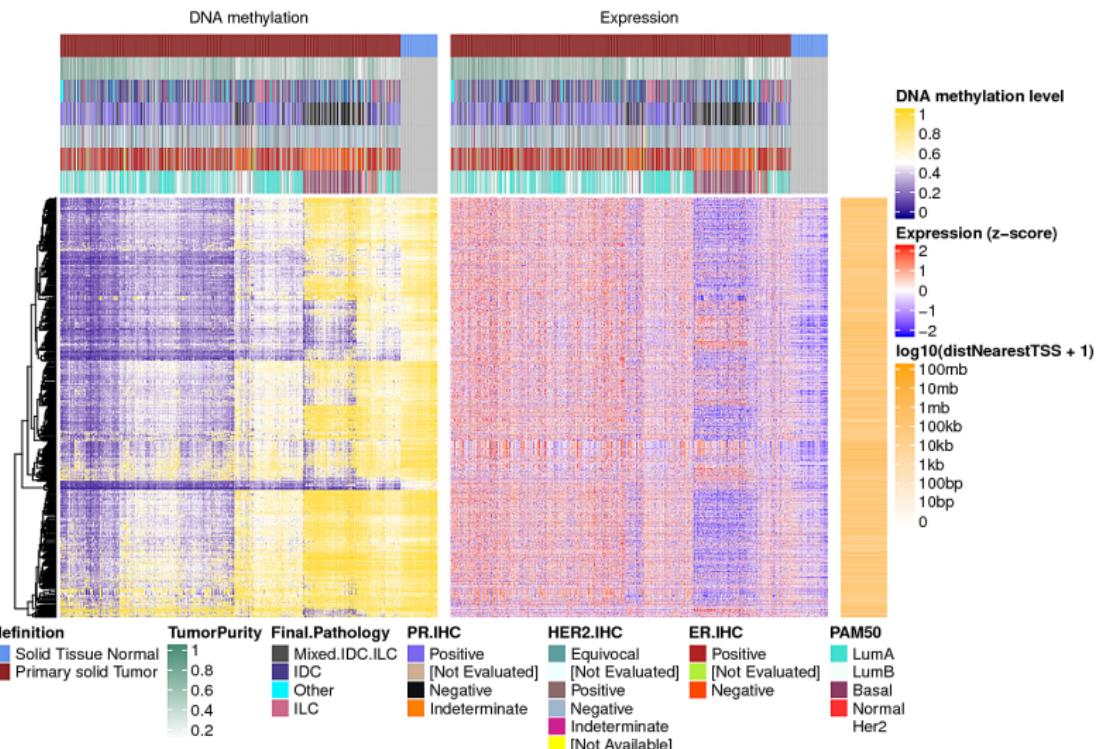
Step 3: Identification of putative target gene(s)

A**B**

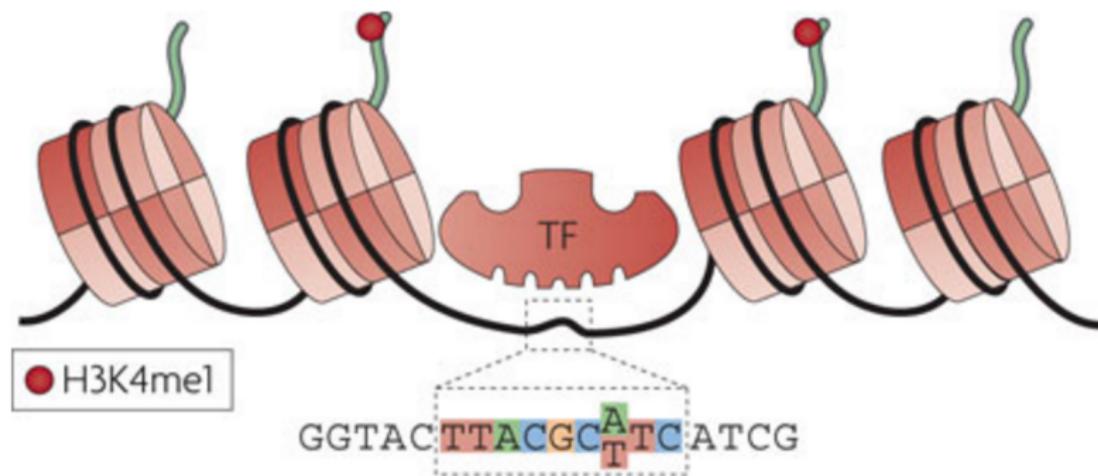
Source: Yao et al. Genome Biology (2015)

Step 3: Probe-target gene pairs inferred

Correspondence between probe DNA methylation and distal gene expression

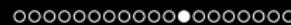


Step 4: Motif enrichment analysis



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

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

Model	LOGO	Transcription factor	Quality	TF family	TF subfamily
ASCL1_HUMAN.H11MO.0.A		ASCL1 (GeneCards)	A	MyoD / ASC-related factors[1.2.2]	Achaete-Scute-like factors[1.2.2.2]
ASCL2_HUMAN.H11MO.0.D		ASCL2 (GeneCards)	D	MyoD / ASC-related factors[1.2.2]	Achaete-Scute-like factors[1.2.2.2]
AHR_HUMAN.H11MO.0.B		AHR (GeneCards)	B	PAS domain factors[1.2.5]	Ahr-like factors[1.2.5.1]
EPAS1_HUMAN.H11MO.0.B		EPAS1 (GeneCards)	B	PAS domain factors[1.2.5]	Ahr-like factors[1.2.5.1]
HIF1A_HUMAN.H11MO.0.C		HIF1A (GeneCards)	C	PAS domain factors[1.2.5]	Ahr-like factors[1.2.5.1]
AIRE_HUMAN.H11MO.0.C		AIRE (GeneCards)	C	AIRE[5.3.1]	AIRE[5.3.1.0.1]
ALX1_HUMAN.H11MO.0.B		ALX1 (GeneCards)	B	Paired-related HD factors[3.1.3]	ALX[3.1.3.1]
ALX3_HUMAN.H11MO.0.D		ALX3 (GeneCards)	D	Paired-related HD factors[3.1.3]	ALX[3.1.3.1]
ALX4_HUMAN.H11MO.0.D		ALX4 (GeneCards)	D	Paired-related HD factors[3.1.3]	ALX[3.1.3.1]
AP2A_HUMAN.H11MO.0.A		TFAP2A (GeneCards)	A	AP-2[1.3.1]	AP-2alpha[1.3.1.0.1]
AP2B_HUMAN.H11MO.0.B		TFAP2B (GeneCards)	B	AP-2[1.3.1]	AP-2beta[1.3.1.0.2]
AP2D_HUMAN.H11MO.0.D		TFAP2D (GeneCards)	D	AP-2[1.3.1]	AP-2delta[1.3.1.0.4]
AP2C_HUMAN.H11MO.0.A		TFAP2C (GeneCards)	A	AP-2[1.3.1]	AP-2gamma[1.3.1.0.3]

HOCOMOCO v11 (<http://hocomoco11.autosome.ru/human/mono?full=true>), Accessed: 25-12-2017

Step 4: Motif enrichment analysis

Objective

Evaluate the enrichment of transcription factors in certain genomic regions.

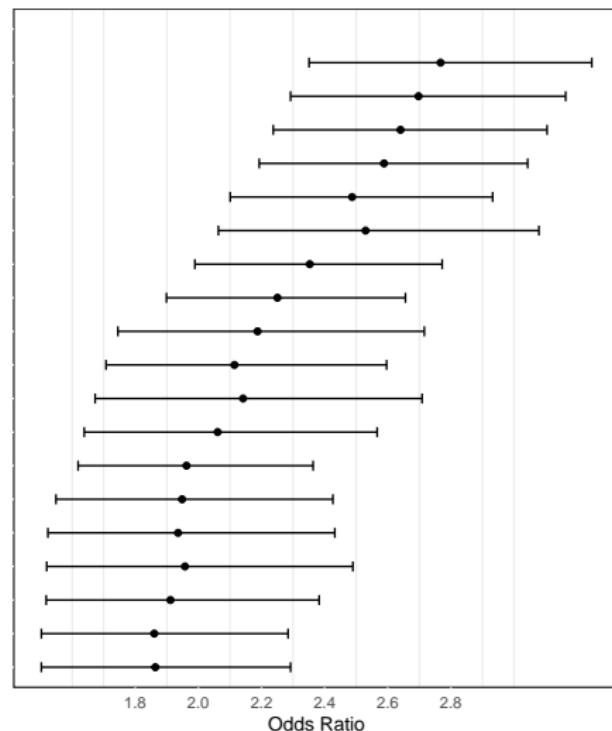
- ① Perform motif matching of transcription factors in probes regions (window $\pm 250\text{bp}$). Performed using HOMER (Hypergeometric Optimization of Motif EnRichment) with HOCOMOCO motifs.
- ② Evaluate which transcription factors are more likely to occur in those regions than in background regions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

Fisher's exact test

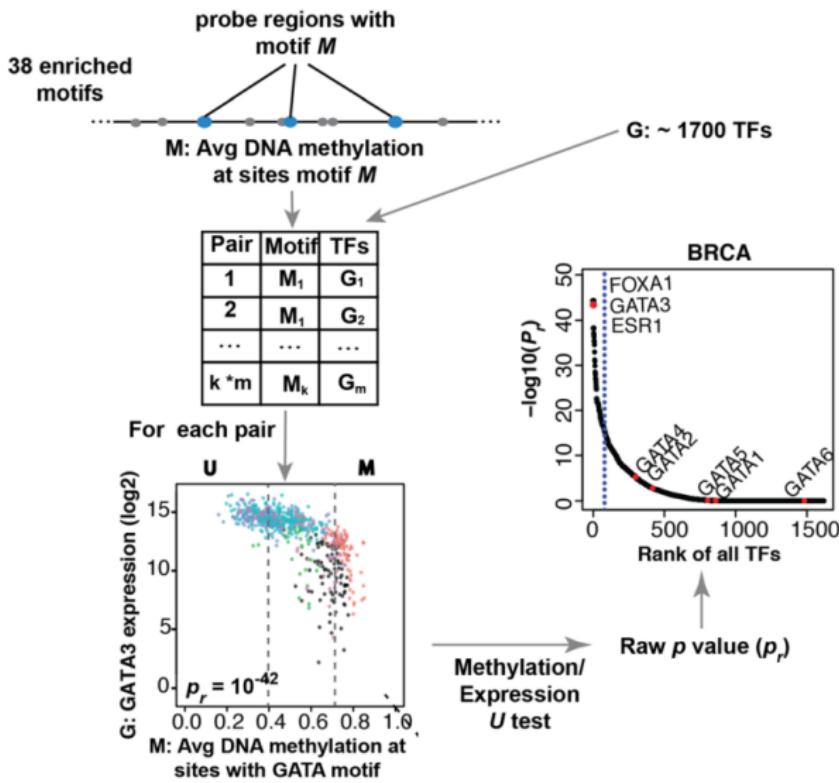
- a: nb of input regions with a match for TF motif.
- b: nb of input regions with no match for TF motif.
- c: nb of background regions with a match for TF motif.
- d: nb of background regions with no match for TF motif.

Step 4: Motif enrichment analysis

Motif	Odds ratio (95% CI)	# probes (% of paired)
FOSL2	2.77 (2.35–3.25)	201 (0.23%)
FOSL1	2.7 (2.29–3.16)	202 (0.23%)
FOSB	2.64 (2.24–3.1)	193 (0.22%)
FOS	2.59 (2.19–3.04)	193 (0.22%)
JUN	2.49 (2.1–2.93)	184 (0.21%)
BATF	2.53 (2.06–3.08)	118 (0.13%)
JUND	2.35 (1.99–2.77)	186 (0.21%)
JUNB	2.25 (1.9–2.66)	181 (0.2%)
HXB13	2.19 (1.74–2.72)	94 (0.11%)
PIT1	2.11 (1.71–2.6)	106 (0.12%)
PRRX1	2.14 (1.67–2.71)	78 (0.09%)
CDX1	2.06 (1.64–2.57)	91 (0.1%)
LMX1A	1.96 (1.62–2.36)	134 (0.15%)
BATF	1.95 (1.55–2.43)	91 (0.1%)
NKX32	1.94 (1.52–2.43)	83 (0.09%)
HME1	1.96 (1.52–2.49)	74 (0.08%)
IRX3	1.91 (1.52–2.38)	90 (0.1%)
PO4F3	1.86 (1.5–2.28)	106 (0.12%)
PO4F1	1.86 (1.5–2.29)	104 (0.12%)



Step 5: Identification of master regulator TF



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

Step 5: Master Regulator TF table

motif	OR	top.potential.TF.family	pvalue.TF.family	top.potential.TF.subfamily	pvalue.TF.subfamily
All		All	All	All	All
HXB13_HUMAN.H11MO.0.A	2.19	HOXB7	6.39e-7	HOXA13	0.00000105
CDX1_HUMAN.H11MO.0.C	2.06	HOXB7	6.39e-7	CDX2	8.20e-7
HXD9_HUMAN.H11MO.0.D	1.98	HOXB7	6.39e-7	HOXA13	0.00000105
PDX1_HUMAN.H11MO.1.A	1.89	HOXB7	6.39e-7	PDX1	0.0000355
HXC11_HUMAN.H11MO.0.D	1.84	HOXB7	6.39e-7	HOXA13	0.00000105
HXB6_HUMAN.H11MO.0.D	1.84	HOXB7	6.39e-7	HOXB7	6.39e-7
HXD8_HUMAN.H11MO.0.D	1.84	HOXB7	6.39e-7	HOXC8	0.00000134
CDX2_HUMAN.H11MO.0.A	1.83	HOXB7	6.39e-7	CDX2	8.20e-7
HXD12_HUMAN.H11MO.0.D	1.77	HOXB7	6.39e-7	HOXA13	0.00000105
HXC9_HUMAN.H11MO.0.C	1.74	HOXB7	6.39e-7	HOXA13	0.00000105

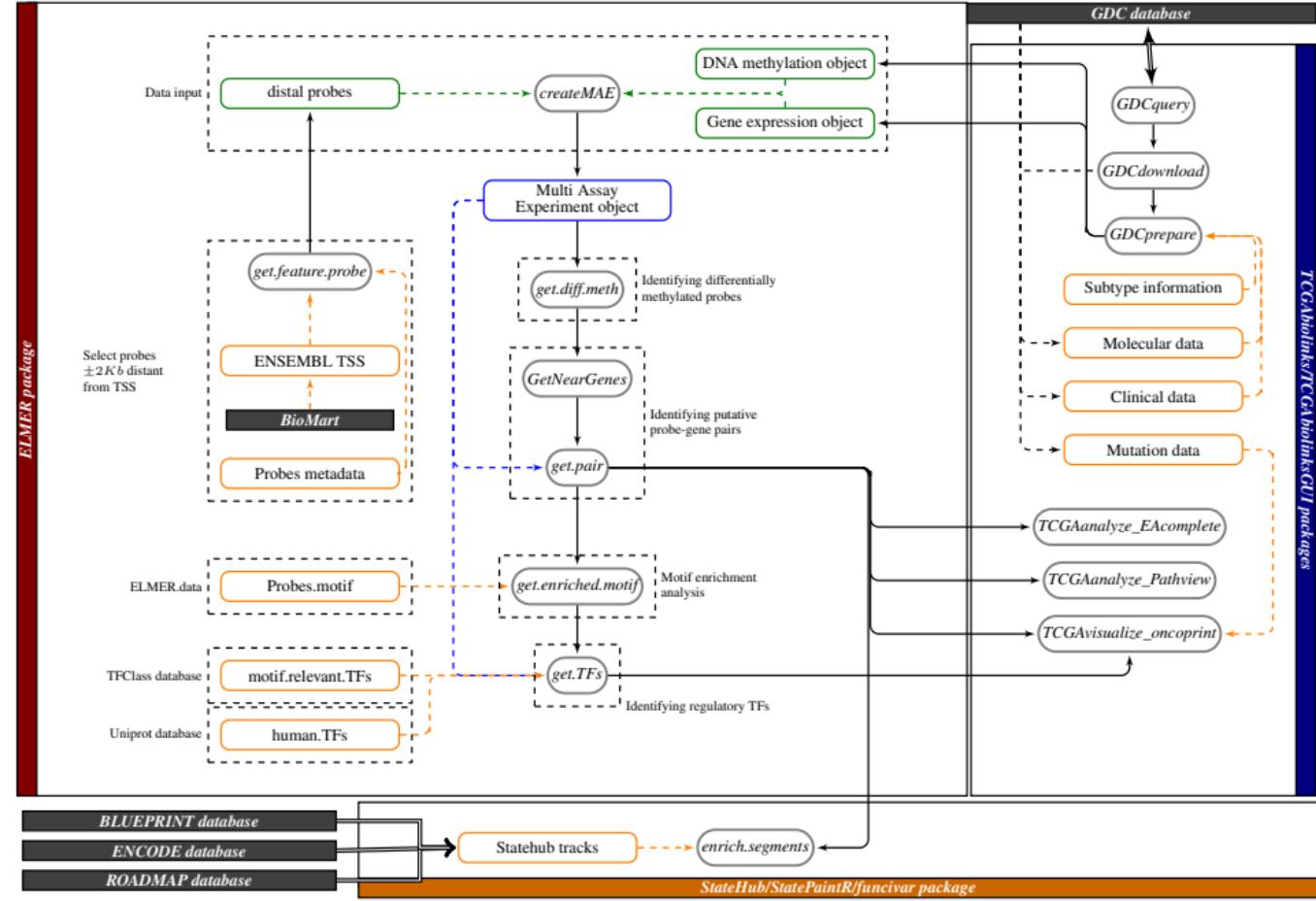
Showing 1 to 10 of 31 entries

Previous

1 2 3 4 Next

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	91	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



Difference of groups *U* and *M* definition in *supervised* and *unsupervised* mode

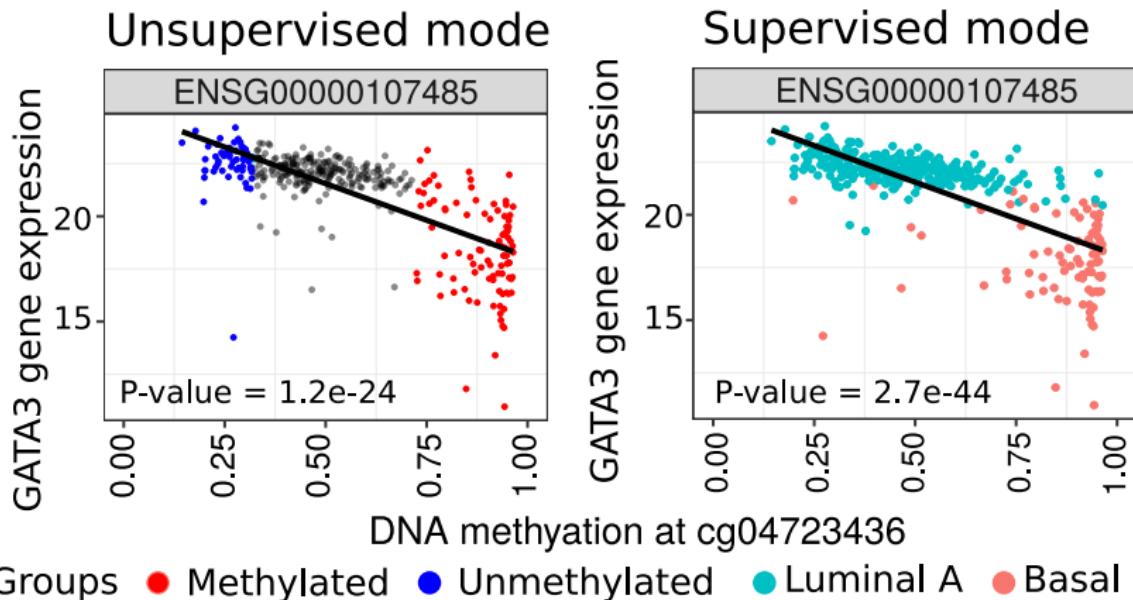


Figure: A: *unsupervised* mode; when `minSubgroupFrac` argument is set to 40%, the methylated group is defined as the highest quintile and the unmethylated group as the lowest quintile; B: *supervised* mode; methylated and unmethylated group are defined as one of the known molecular subtypes.



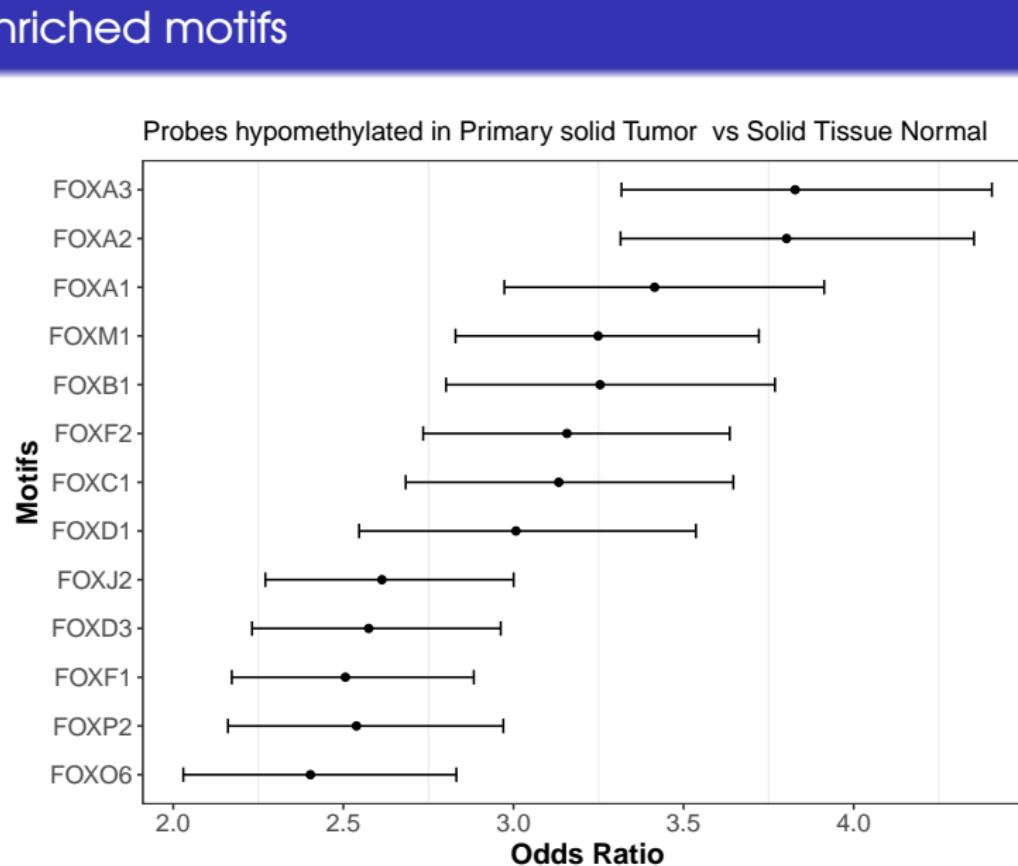
Case study: TCGA Breast Invasive Carcinoma (BRCA)

Table: Summary of the available samples in TCGA for BRCA

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

Table: Results supervised mode

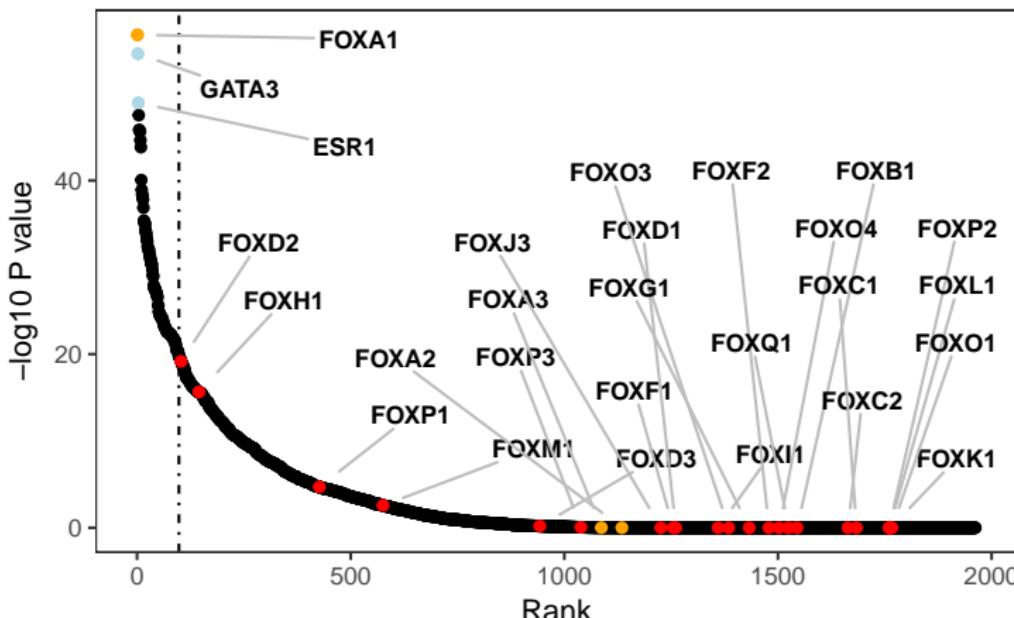
Inferred gene-probe pairs	2167
Enriched motifs	312
Master Regulator TF	17

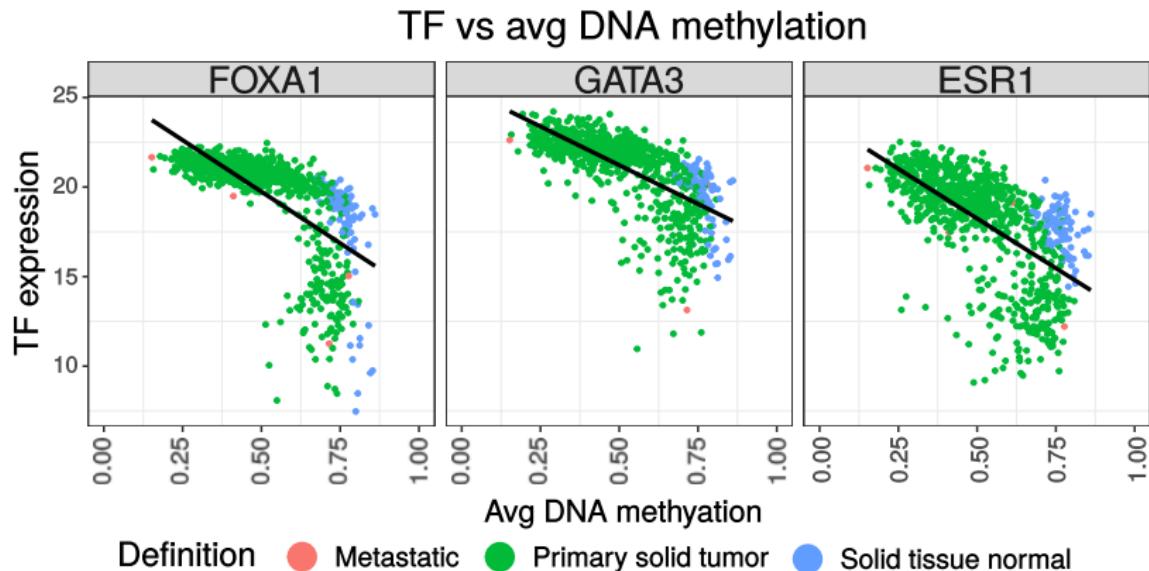




Motif: FOXA3

TF classification • None • Same family • Same subfamily • Top 3







Candidate master regulator TF

RESEARCH ARTICLE | OPEN ACCESS

Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours

André Albergaria, Joana Paredes, Bárbara Sousa, Fernanda Milanezi, Vítor Carneiro, Joana Bastos, Sandra Costa, Daniella Vieira, Nair Lopes, Eric W Lam, Nuno Lunet and Fernando Schmitt 

Breast Cancer Research 2009 11:R40 | DOI: 10.1186/bcr2327 | © Albergaria et al.; licensee BioMed Central Ltd. 2009

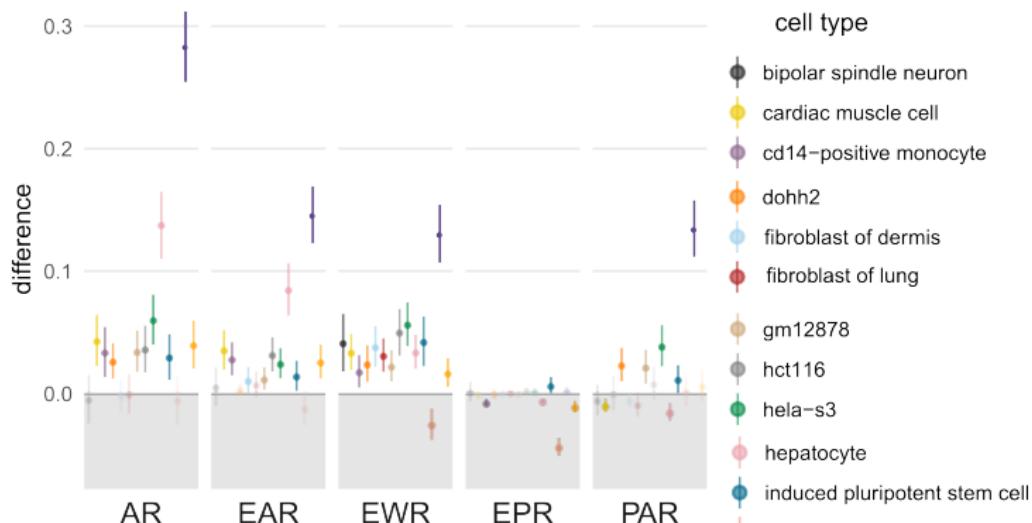
Received: 5 January 2009 | Accepted: 23 June 2009 | Published: 23 June 2009

Article | [OPEN](#)

Retinoic acid receptor alpha is associated with tamoxifen resistance in breast cancer

Henrik J. Johansson, Betzabe C. Sanchez, Filip Mundt, Jenny Forshed, Aniko Kovacs, Elena Panizza, Lina Hultin-Rosenberg, Bo Lundgren, Ulf Martens, Gyöngyvér Máthé, Zohar Yakhini, Khalil Helou, Kamilla Krawiec, Lena Kanter, Anders Hjerpe, Olle Stål, Barbro K. Linderholm & Janne Lehtio 

Characterization of chromatin state context of enriched probes using FunciVar



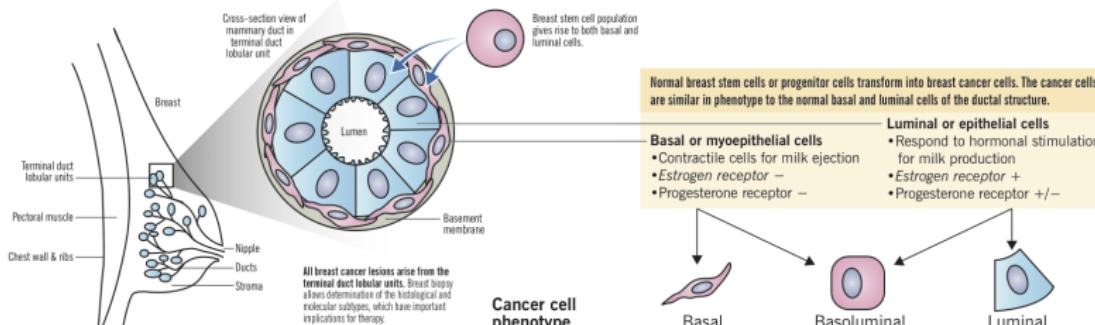
Acronyms

- AR: Active region
- EAR: active enhancer
- EWR: Weak Enhancer
- EPR: poised enhancer
- PAR: active promoter

Supervised analysis: BRCA molecular subtypes

Breast cancer pathogenesis and histologic vs. molecular subtypes

Eric Wong and Jemma Rebello



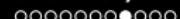
Histological subtypes	Ductal	Lobular	Cancer cell phenotype	Molecular subtypes
Preinvasive cancer 25% Cells limited to basement membrane	Ductal carcinoma in situ (DCIS) 80% May spread through ducts and distort duct architecture 1% progress to invasive cancer per year Usually unilateral	Lobular carcinoma in situ (LCIS) 20% Does not distort duct architecture Same genetic abnormality as ILC – E-cadherin loss 1% progress per year Can be bilateral		Triple negative ER-, PR-, HER2-
Invasive cancer 75% Extension beyond the basement membrane	Invasive ductal carcinoma (IDC) 79% Usually from DCIS precursor Cause fibrous response, producing a palpable mass on examination Metastasis through lymphatics and blood	Invasive lobular carcinoma (ILC) 10% Usually from LCIS precursor Minimal fibrous response, presents less often with palpable mass Metastasis through abdominal viscera to GI, ovaries, uterus Almost always ER+		HER2+ ER+/PR+
			% of breast cancers Receptor expression Histologic grade Prognosis Response to medical therapy	Luminal B Luminal A
			15-20% 10-15% 20% 40%	
			High (grade III) Level of cell differentiation	
			Poor	Good
			Chemotherapy Trastuzumab	Endocrine

Curr Treat Options Oncol. 2000 Aug;1(3):199-209.
Clin Transl Oncol. 2008 Dec;10(12):777-85.

Nat Clin Pract Oncol. 2007 Sep;4(9):516-25.
Rostami BE

Triple negative tumours respond best to chemotherapy, similar to other aggressive cancers.

Luminal A tumours respond best to endocrine therapy, e.g. antiestrogen or aromatase inhibitor.



Supervised analysis: Candidate regulatory TFs

TF	LumA (vs basal)	LumB (vs basal)	LumA (vs normal)	LumB (vs normal)	Basal (vs LumA)	Basal (vs LumB)	Basal (vs HER2)	HER2 (vs Basal)
AR	x	x	x					
BCL11A					x	x	x	
CEBPB					x	x	x	
E2F3						x	x	
EMX1	x	x	x					
ESR1	x	x	x	x				
ETV6					x	x	x	
FOXA1	x	x	x	x				x
FOXP1	x	x						x
GATA3	x	x	x	x				x
HOXB1	x	x						
HOXB2	x	x						x
HOXB3								x
HOXC10								x
KLF5						x	x	
LMX1B	x	x	x					
MNX1								x
MYB	x			x				
NFIL3					x	x		
PBX1		x		x		x		
RARA	x	x	x					
RUNX3					x	x		
SOX8						x		
SOX9						x		
SOX11					x	x		
ZNF467	x	x	x			x		
ZIC1					x	x	x	

TF master regulator: Basal

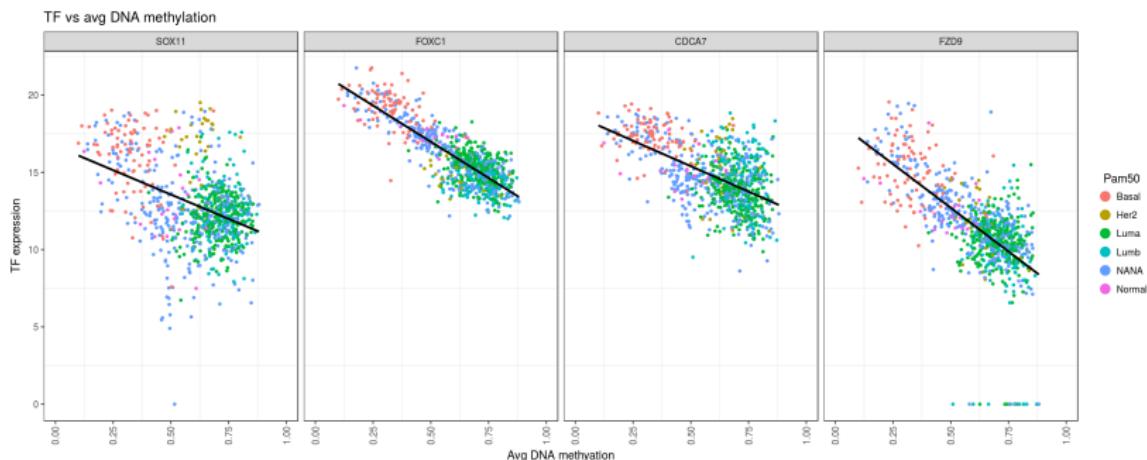


Figure: SOX11 and top3 TF expression vs avg DNA methylation of paired enriched probes for SOX10 - Probes hypermethylated in LumA vs Basal



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

Next steps: TF knockdown

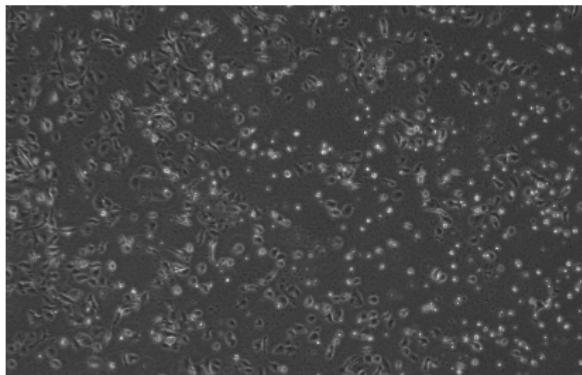
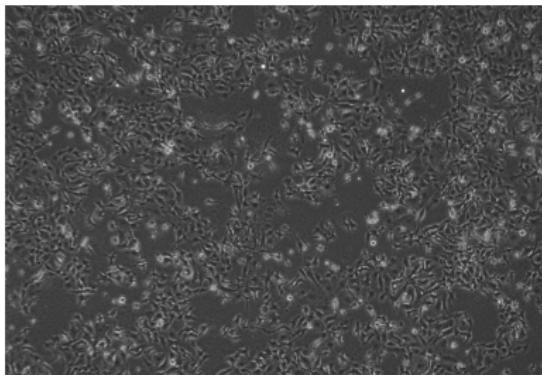


Figure: Candidate master regulator Transcription Factors (TF) knockdown in the SKGT4 human esophageal adenocarcinoma cell line. Figure produced by Dr. Dechen Lin.