

Enhancer Linking by Methylation/Expression Relationship: a case study using Breast Cancer

Tiago Chedraoui Silva

University of São Paulo (USP)
Cedars-Sinai Medical Center

Tuesday 6th June, 2017

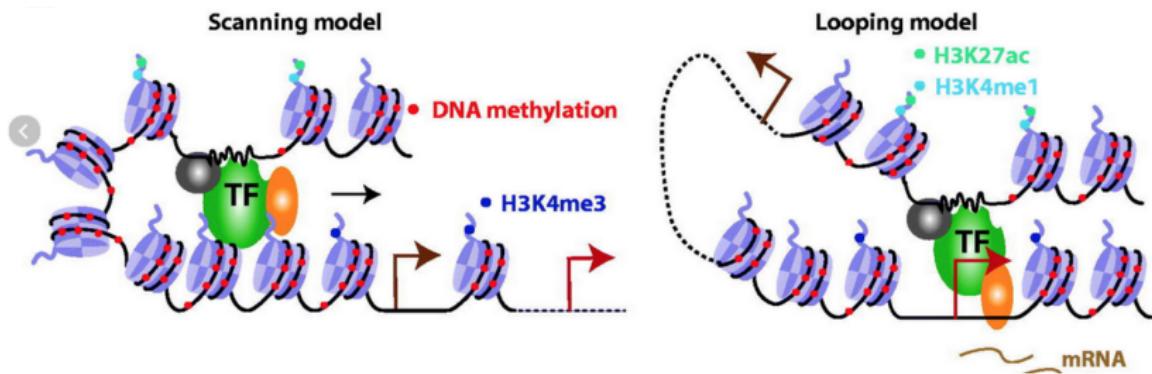




Overview

- ① Introduction
- ② ELMER
- ③ Comparison between versions
- ④ Case of study - TCGA BRCA

Enhancer-mediated gene regulation



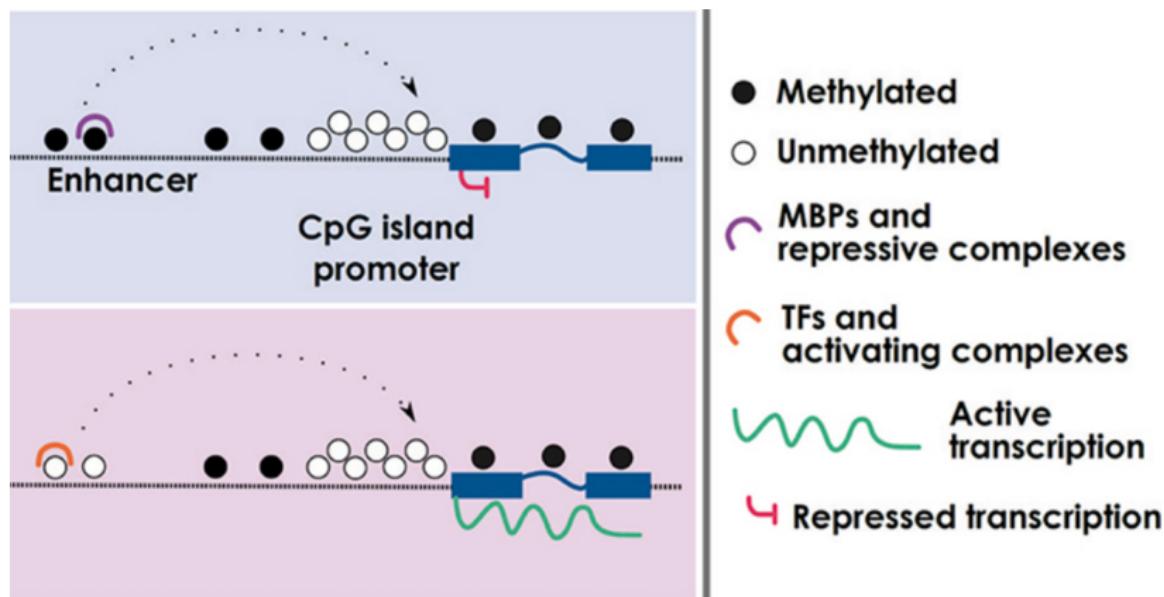
Yao et al. Genome Biology (2015) 16:105.



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: Front. Aging Neurosci., 05 March 2015 <http://dx.doi.org/10.3389/fnagi.2015.00019>

Enhancer Linking by Methylation/Expression Relationship

ELMER

platforms	all	downloads	top 20%	posts	0	in Bloc	1.5 years
build	ok	commits	0.67	test coverage	unknown		

Yao et al. *Genome Biology* (2015) 16:105
DOI 10.1186/s13059-015-0668-3



METHOD

Open Access

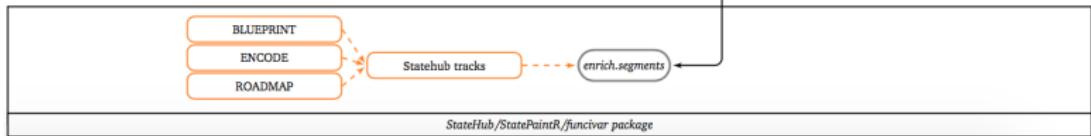
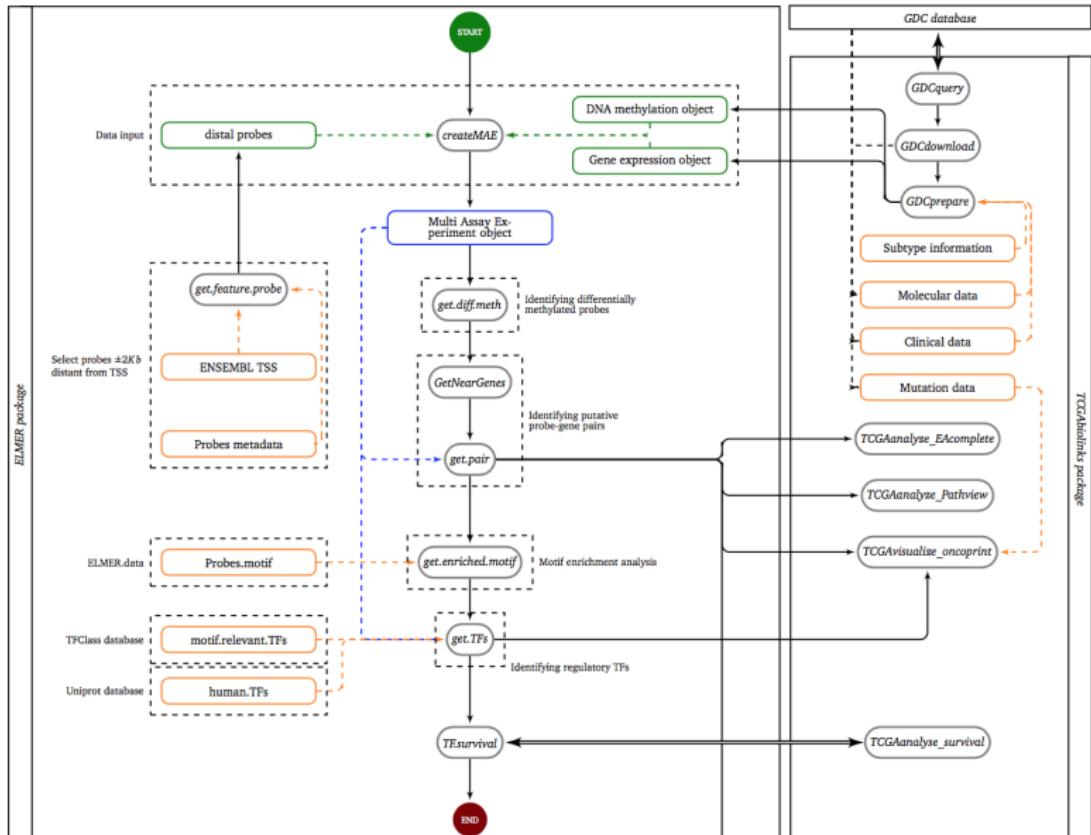


Inferring regulatory element landscapes and transcription factor networks from cancer methylomes

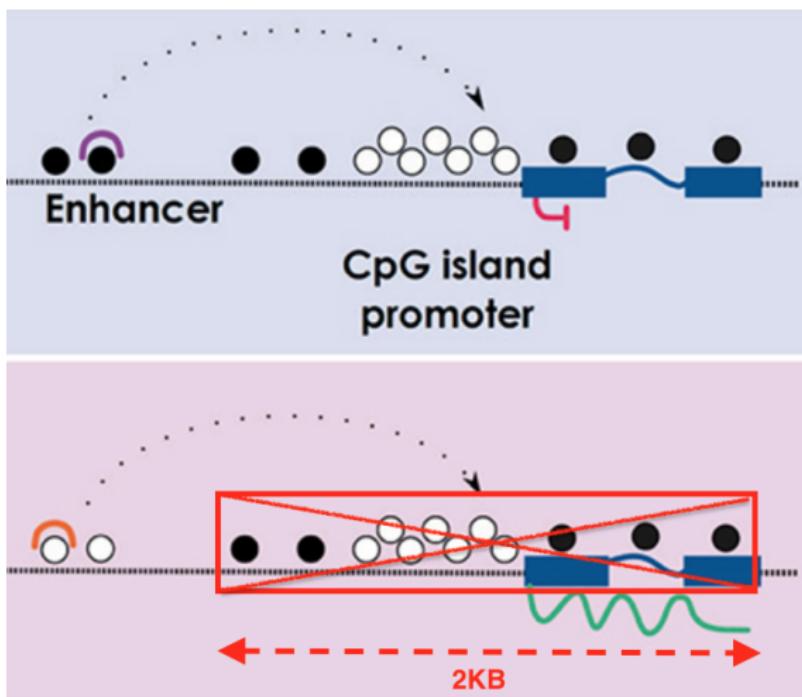
Lijing Yao¹, Hui Shen², Peter W Laird², Peggy J Farnham^{1*} and Benjamin P Berman^{1,3*}

Abstract

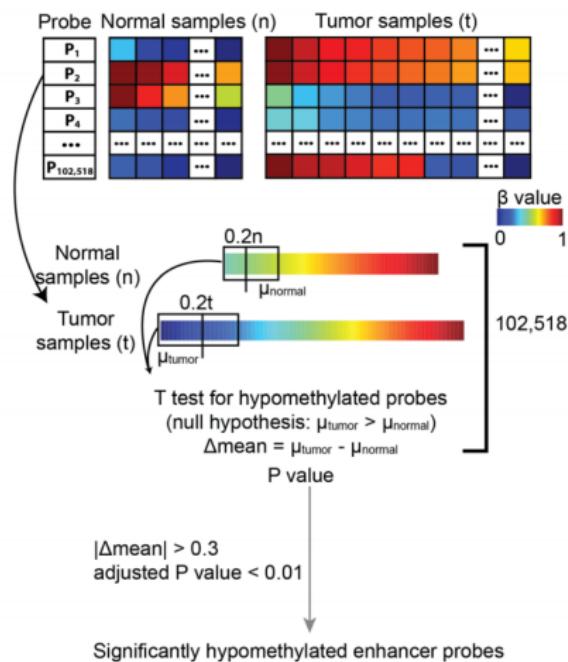
Recent studies indicate that DNA methylation can be used to identify transcriptional enhancers, but no systematic approach has been developed for genome-wide identification and analysis of enhancers based on DNA methylation. We describe ELMER (Enhancer Linking by Methylation/Expression Relationships), an R-based tool that uses DNA methylation to identify enhancers and correlates enhancer state with expression of nearby genes to identify transcriptional targets. Transcription factor motif analysis of enhancers is coupled with expression analysis of transcription factors to infer upstream regulators. Using ELMER, we investigated more than 2,000 tumor samples from The Cancer Genome Atlas. We identified networks regulated by known cancer drivers such as GATA3 and FOXA1 (breast cancer), SOX17 and FOXA2 (endometrial cancer), and NFE2L2, SOX2, and TP63 (squamous cell lung cancer). We also identified novel networks with prognostic associations, including RUNX1 in kidney cancer. We propose ELMER as a powerful new paradigm for understanding the *cis*-regulatory interface between cancer-associated transcription factors and their functional target genes.



Step 1: Identify distal probes



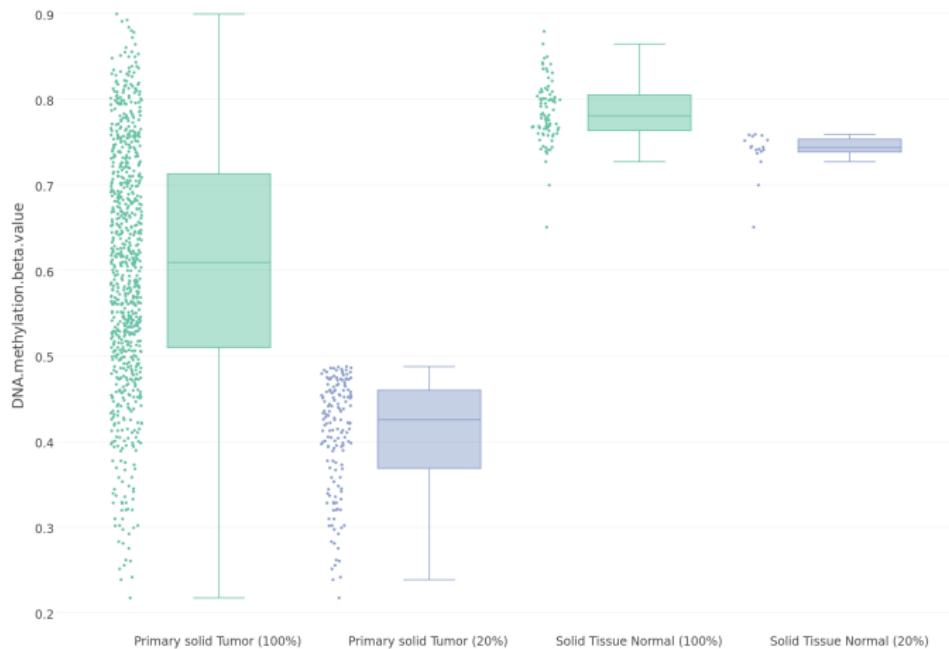
Step 2: Differentially methylated distal probes



Yao et al. Genome Biology (2015) 16:105.

Step 2: Differentially methylated distal probes

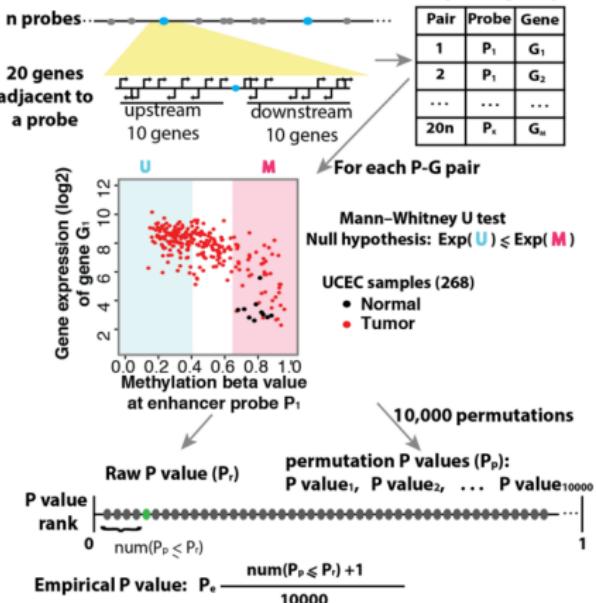
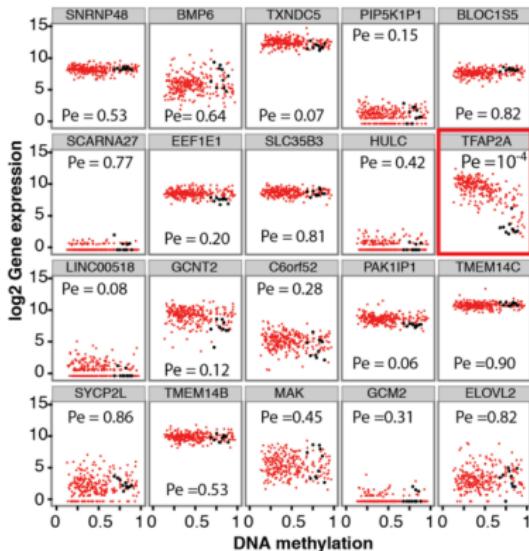
Boxplot DNA methylation (probe cg00001809)



Step 2: Differentially methylated distal probes

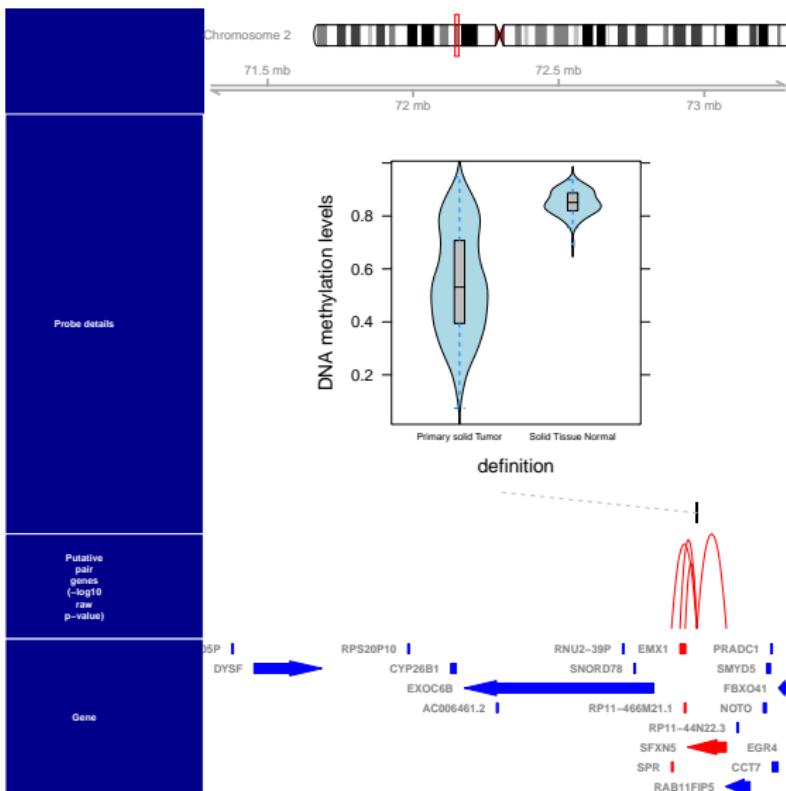
probe	pvalue	Primary.solid.Tumor_Minus_Solid.Tissue.Normal	adjust.p
cg00001809	1.97e-35	-0.32	1.26e-34
cg00008695	1.62e-67	-0.44	3.72e-66
cg00009553	6.84e-31	-0.52	3.61e-30
cg00009922	1.83e-69	-0.38	4.60e-68
cg00012529	2.22e-18	-0.38	6.54e-18
cg00014333	1.11e-31	-0.30	6.10e-31
cg00014484	1.01e-44	-0.34	9.36e-44

Step 3: Get pairs

A**B**

Yao et al. Genome Biology (2015) 16:105.

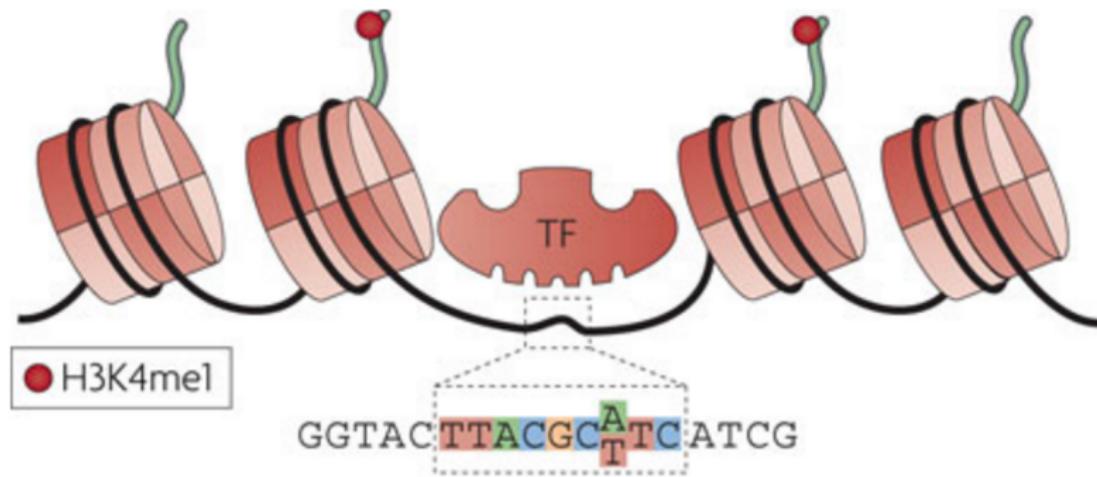
Step 3: Pairs inferred



Step 3: Pairs inferred

Probe	GenelD	Symbol	Distance	Sides	Raw.p	Pe
cg05120309	ENSG00000196405	EVL	0	L3	3,04E-56	1,00E-04
cg25343204	ENSG00000106541	AGR2	0	R1	4,00E-56	1,00E-04
cg14058239	ENSG00000141424	SLC39A6	0	R1	5,17E-56	1,00E-04
cg14986386	ENSG00000183323	CCDC125	0	L2	3,26E-55	1,00E-04
cg04723436	ENSG00000107485	GATA3	110455	R3	6,01E-55	2,00E-04
cg08815672	ENSG00000075275	CELSR1	0	L2	1,93E-54	1,00E-04
cg12633764	ENSG00000279685	MAPT-IT1	2591	L1	4,44E-54	1,00E-04
cg02711212	ENSG00000245750	DRAIC	0	R1	2,39E-53	2,00E-04
cg08070476	ENSG00000106537	TSPAN13	63251	L2	2,56E-53	1,00E-04
cg12633764	ENSG00000186868	MAPT	0	L2	4,22E-53	1,00E-04
cg00601836	ENSG00000091831	ESR1	0	R2	5,41E-53	3,00E-04
cg02187712	ENSG00000124664	SPDEF	11491	L1	1,22E-52	2,00E-04
cg25417042	ENSG00000138166	DUSP5	0	R2	1,35E-52	1,00E-04
cg25999891	ENSG00000072310	SREBF1	0	L3	4,43E-52	1,00E-04
cg16013309	ENSG00000171604	CXXC5	10507	L1	5,97E-52	1,00E-04
cg23005227	ENSG00000114737	CISH	0	R1	6,80E-52	1,00E-04
cg14058239	ENSG00000134759	ELP2	3177	R2	9,77E-52	1,00E-04
cg02546690	ENSG00000187486	KCNJ11	423601	L8	1,68E-51	4,00E-04
cg00836101	ENSG00000276107	CTD-2033D15.2	2512	R2	2,29E-51	1,00E-04
cg13720048	ENSG00000081923	ATP8B1	0	L2	2,33E-51	1,00E-04
cg26703758	ENSG00000187486	KCNJ11	380882	L8	3,81E-51	4,00E-04

Step 4: Enriched motifs



Nature Reviews | Genetics

Next-generation genomics: an integrative approach R. David Hawkins et al. *Nature Reviews Genetics* 11, 476-486

Step 4: Enriched motifs

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

Reset Select Columns Get CSV HUMAN_mono_motifs.tsv PWMS for HUMAN transcription factors

Model	LOGO	Transcription factor	TF family	TF subfamily
AP2C_HUMAN.H10MO.A		TFAP2C (GeneCards)	AP-2[1.3.1]	AP-2gamma[1.3.1.0.3]
BATF_HUMAN.H10MO.A		BATF (GeneCards)	B-ATF-related factors[1.1.4]	B-ATF[1.1.4.0.1]
MAX_HUMAN.H10MO.A		MAX (GeneCards)	bHLH-ZIP factors[1.2.6]	Myc / Max factors[1.2.6.5]
MYC_HUMAN.H10MO.A		MYC (GeneCards)	bHLH-ZIP factors[1.2.6]	Myc / Max factors[1.2.6.5]
USF1_HUMAN.H10MO.A		USF1 (GeneCards)	bHLH-ZIP factors[1.2.6]	USF factors[1.2.6.2]
USF2_HUMAN.H10MO.A		USF2 (GeneCards)	bHLH-ZIP factors[1.2.6]	USF factors[1.2.6.2]
CEBPA_HUMAN.H10MO.A		CEBPA (GeneCards)	C/EBP-related[1.1.8]	C/EBP[1.1.8.1]
CEBPB_HUMAN.H10MO.A		CEBPB (GeneCards)	C/EBP-related[1.1.8]	C/EBP[1.1.8.1]
CEBPE_HUMAN.H10MO.A		CEBPE (GeneCards)	C/EBP-related[1.1.8]	C/EBP[1.1.8.1]
CREB1_HUMAN.H10MO.A		CREB1 (GeneCards)	CREB-related factors[1.1.7]	CREB-like factors[1.1.7.1]
E2F1_HUMAN.H10MO.A		E2F1 (GeneCards)	E2F-related factors[3.3.2]	E2F[3.3.2.1]

Step 4: Enriched motifs

Definition

- a: probes with motif in the set (from probes-gene pair)
- b: probes without motif in the set (from probes-gene pair)
- c: probes with motif for all distal probes
- d: probes without motif for all distal probes

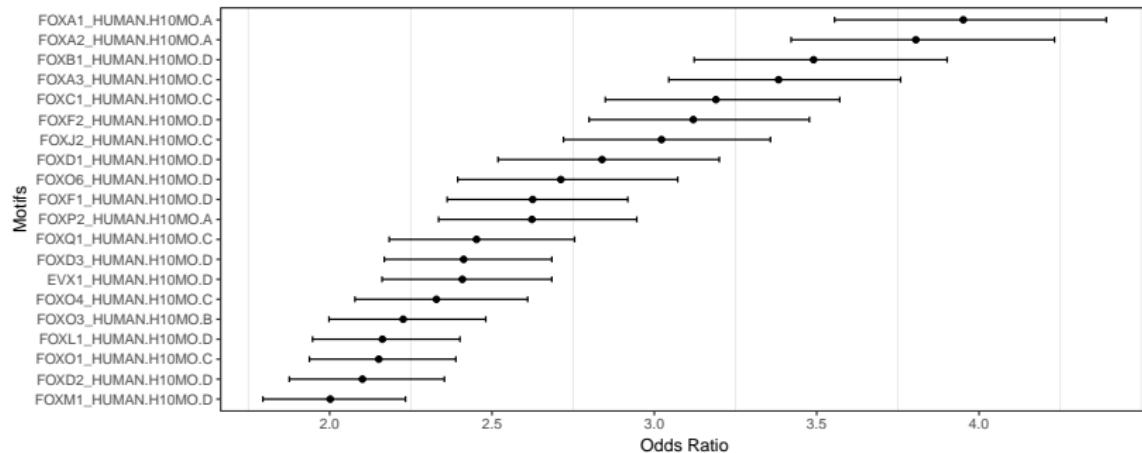
Odds ratio

$$p = \frac{a}{a+b}, P = \frac{c}{c+d} \quad (1)$$

$$OR = \frac{\frac{p}{1-p}}{\frac{P}{1-P}} = \frac{\frac{a}{b}}{\frac{c}{d}} \quad (2)$$

$$SD = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \quad (3)$$

Step 4: Motif enrichment



Step 4: Motif enrichment

motif	NumOfProbes	OR	lowerOR	upperOR
FOXA1_HUMAN.H10MO.A	603	3.95	3.56	4.39
FOXA2_HUMAN.H10MO.A	578	3.81	3.42	4.23
FOXB1_HUMAN.H10MO.D	467	3.49	3.12	3.90
FOXA3_HUMAN.H10MO.C	615	3.38	3.04	3.76
FOXC1_HUMAN.H10MO.C	443	3.19	2.85	3.57
FOXF2_HUMAN.H10MO.D	519	3.12	2.80	3.48
FOXJ2_HUMAN.H10MO.C	617	3.02	2.72	3.36
FOXD1_HUMAN.H10MO.D	365	2.84	2.52	3.20
FOXO6_HUMAN.H10MO.D	325	2.71	2.39	3.07
FOXF1_HUMAN.H10MO.D	595	2.63	2.36	2.92
FOXP2_HUMAN.H10MO.A	402	2.62	2.34	2.95
FOXQ1_HUMAN.H10MO.C	401	2.45	2.18	2.75
FOXD3_HUMAN.H10MO.D	561	2.41	2.17	2.68
EVX1_HUMAN.H10MO.D	521	2.41	2.16	2.68
FOXO4_HUMAN.H10MO.C	427	2.33	2.08	2.61
FOXO3_HUMAN.H10MO.B	523	2.23	2.00	2.48
FOXL1_HUMAN.H10MO.D	638	2.16	1.95	2.40
FOXO1_HUMAN.H10MO.C	647	2.15	1.94	2.39
FOXD2_HUMAN.H10MO.D	433	2.10	1.88	2.35
FOXM1_HUMAN.H10MO.D	496	2.00	1.79	2.23

Step 5: Human TFs

UniProtKB - "transcription factor" AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]" Advanced Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

UniProtKB results

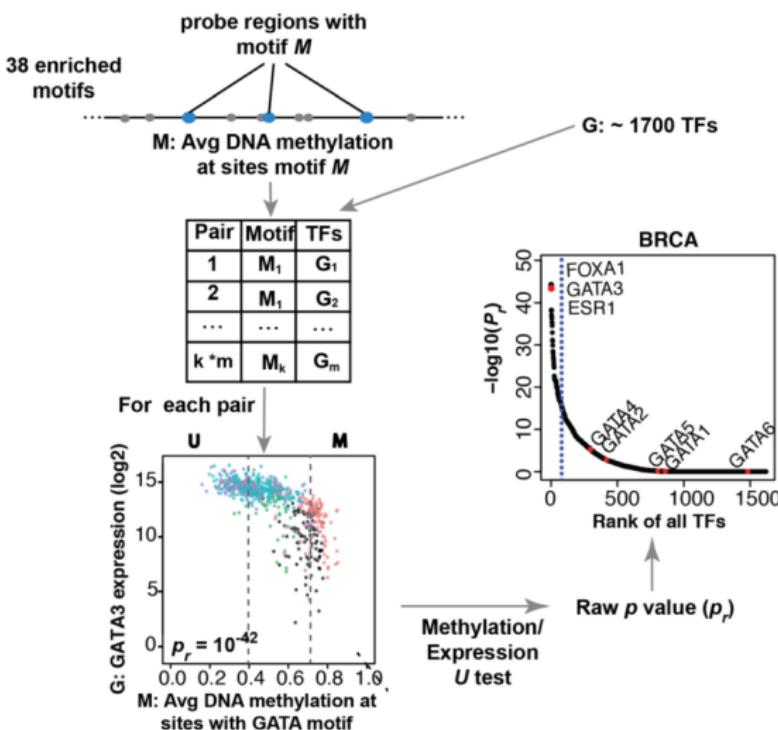
Filter by:

- Reviewed (2,017) ×
- Swiss-Prot
- Popular organisms (2,017) ×
- Human (2,017) ×
- Proteomes
- UP000005640 (2,017)
- Search terms
- Filter "transcription factor" as:
- gene ontology (2,060)
- protein family (25)
- protein name (409)
- View by:
- Results table
- Taxonomy
- Keywords
- Gene Ontology
- Enzyme class

1 to 25 of 2,017 Show 25 ↗

Entry	Entry name	Protein names	Gene names	Organism
P04637	P53_HUMAN	Cellular tumor antigen p53	TP53 P53	Homo sapiens (Human)
P15884	ITF2_HUMAN	Transcription factor 4	TCF4 BHLHB19, ITF2, SEF2	Homo sapiens (Human)
Q04206	TF65_HUMAN	Transcription factor p65	RELA NFKB3	Homo sapiens (Human)
Q00613	HSF1_HUMAN	Heat shock factor protein 1	HSF1 HSTF1	Homo sapiens (Human)
Q9NQB0	TF7L2_HUMAN	Transcription factor 7-like 2	TCF7L2 TCF4	Homo sapiens (Human)
P18074	ERCC2_HUMAN	TFIIE basal transcription factor co... TFIIE basal transcription factor co... TFIIE basal transcription factor co...	ERCC2 XPD, XPDC	Homo sapiens (Human)
P10275	ANDR_HUMAN	Androgen receptor	AR DHTR, NR3C4	Homo sapiens (Human)
P08047	SP1_HUMAN	Transcription factor Sp1	SP1 TSFP1	Homo sapiens (Human)
P05412	JUN_HUMAN	Transcription factor AP-1	JUN	Homo sapiens (Human)
P35222	CTNB1_HUMAN	Catenin beta-1	CTNNB1 CTNNB, OK/SW-cl.35, PRO2286	Homo sapiens (Human)
P35555	FBN1_HUMAN	Fibrillin-1	FBN1 FBN	Homo sapiens (Human)

Step 5: TF ranking plot



Yao et al. Genome Biology (2015) 16:105.

Step 5: Results

motif	top.pot.TF.family	top.pot.TF.subfamily	pot.TF.family	pot.TF.subfamily	top_5percent_TFs
AIRE.C	NA	NA	NA	NA	FOXA1;GATA3;ESR1;SPD...
ALX1.B	NA	NA	NA	NA	FOXA1;GATA3;ESR1;SPD...
ALX3.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;SPD...
ALX4.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;RAR...
ANDR.A	ESR1	AR	ESR1;AR	AR	FOXA1;GATA3;ESR1;SPD...
ARI3A.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;RAR...
ARI3A.S	NA	NA	NA	NA	FOXA1;GATA3;ESR1;RAR...
ARI5B.C	NA	NA	NA	NA	FOXA1;GATA3;ESR1;SPD...
ARX.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;RAR...
BARH1.D	EMX1	NA	EMX1;LBX2	NA	FOXA1;GATA3;ESR1;SPD...
BARH2.D	EMX1	NA	EMX1;LBX2	NA	FOXA1;GATA3;ESR1;CXX...
BARX1.D	EMX1	NA	EMX1;LBX2	NA	FOXA1;GATA3;ESR1;RAR...
BARX2.D	EMX1	NA	EMX1;LBX2	NA	FOXA1;GATA3;ESR1;CXX...
BATF3.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;SPD...
BATF.A	NA	NA	NA	NA	FOXA1;GATA3;ESR1;CXX...
BATF.S	NA	NA	NA	NA	FOXA1;GATA3;ESR1;CXX...
BCL6.C	ZSCAN16	NA	ZSCAN16;OVOL1	NA	FOXA1;GATA3;ESR1;RAR...
BHE22.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;SPD...
BPTF.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;RAR...
BRAC.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;RAR...
BRCA1.D	NA	NA	NA	NA	FOXA1;GATA3;ESR1;CXX...
BSH.D	EMX1	NA	EMX1;LBX2	NA	FOXA1;GATA3;ESR1;SPD...
CDC5L.D	MYB	MYB	MYB	MYB	FOXA1;GATA3;ESR1;RAR...
CDX1.C	NA	NA	NA	NA	FOXA1;GATA3;ESR1;CXX...
CDX2.C	NA	NA	NA	NA	FOXA1;GATA3;ESR1;CXX...
CEBPA.A	NA	NA	NA	NA	FOXA1;GATA3;ESR1;CXX...

Differences between versions

Table: Main differences between ELMER version 1 and 2)

Features	ELMER Version 1	ELMER Version 2
Main data structure	mee object (own data structure)	MAE object (Bioconductor data structure)
Auxiliary data	Manually created	Programmatically created
Number of human TF	1,982	1,987 (Uniprot database)
Number of TF motifs	91	640 (HOCOMOCO v10 database)
TF classification	78 families	80 families and 308 subfamilies (TFCClass database)
Analysis performed	Normal tumor samples vs experiment	Group 1 vs group 2
TCGA samples 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	HumanMethylation450	HumanMethylationEPIC / HumanMethylation450

TCGA Breast Invasive Carcinoma (BRCA) Samples

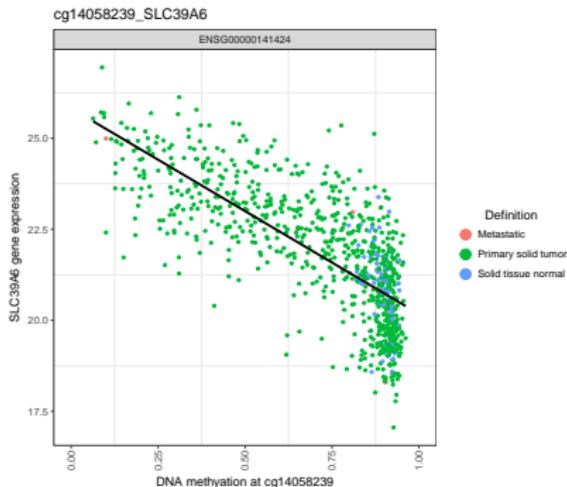
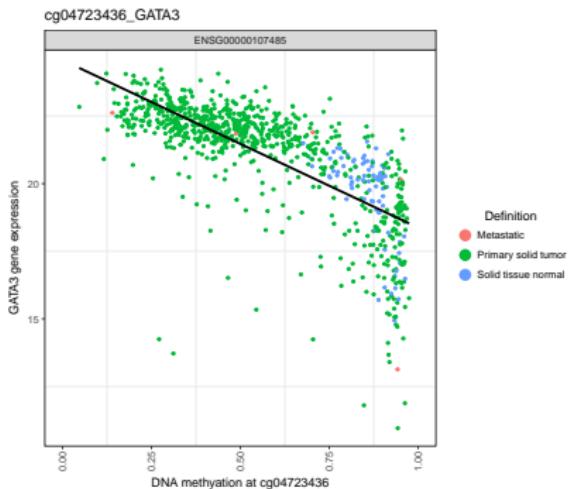
Table: Summary of the available samples in TCGA for BRCA

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

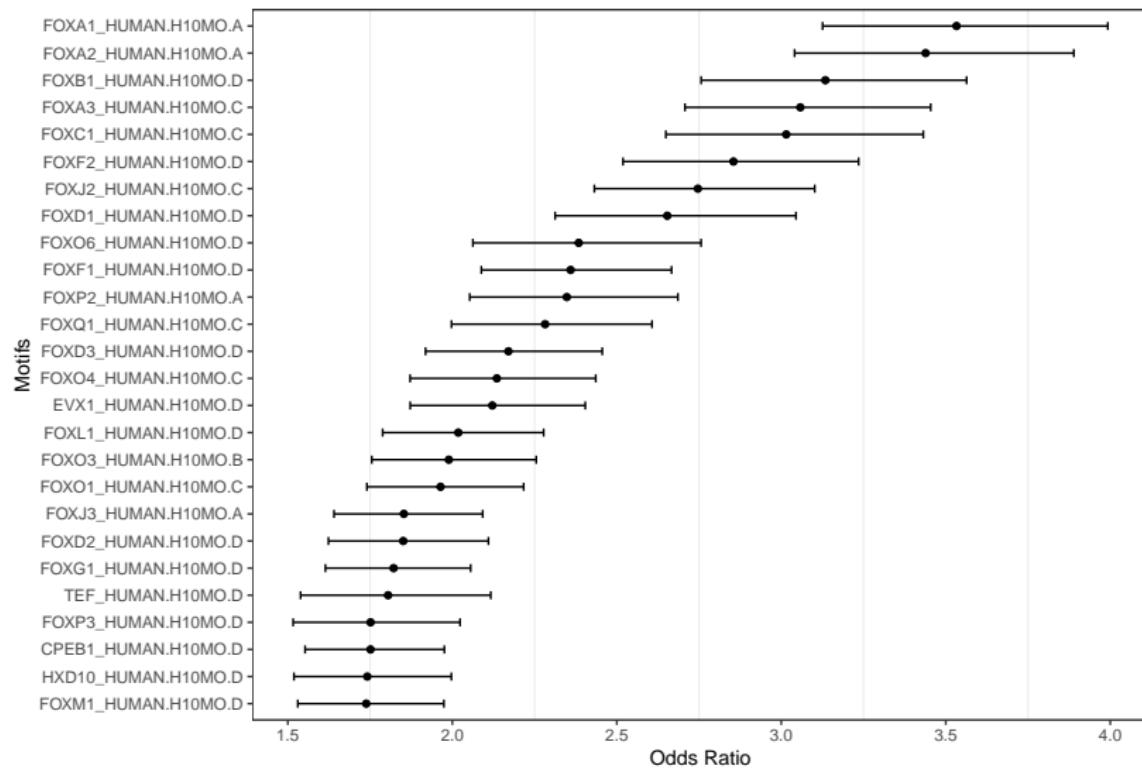
Table: Results

Inferred gene-probe pairs	2167
Enriched motifs	312
Regulatory TF factors	17

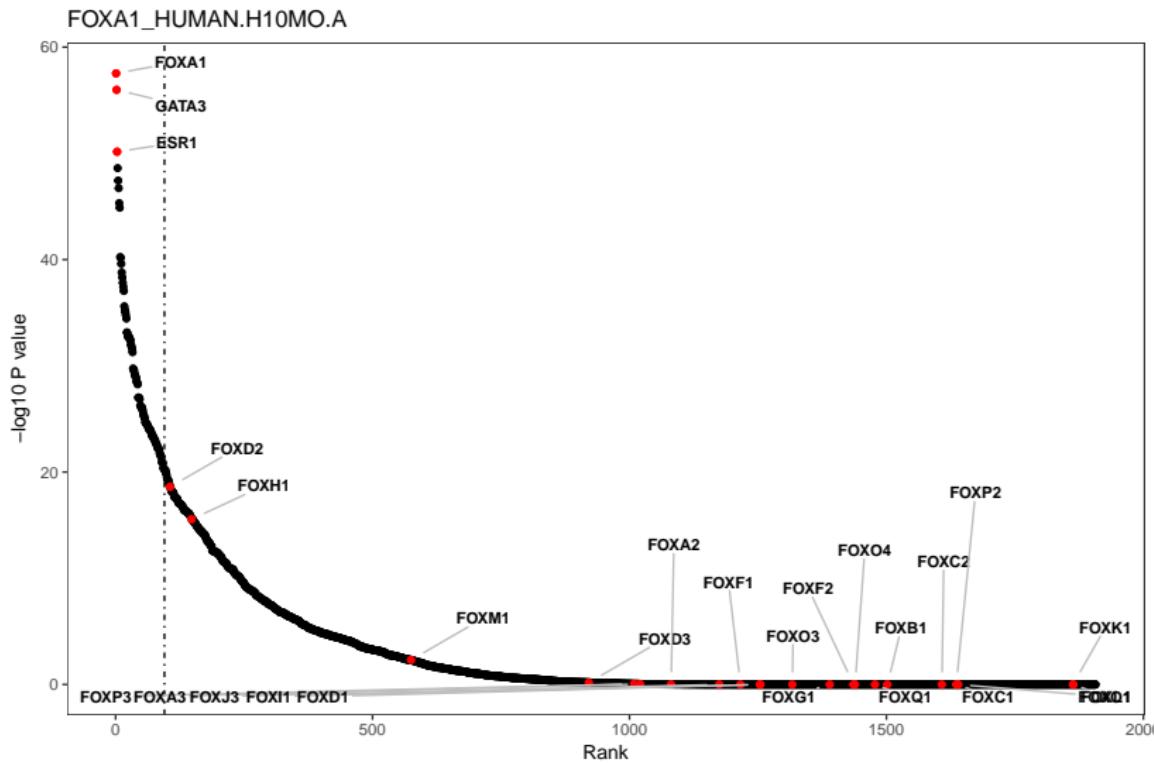
Step 3: Pairs inferred



Top enriched motifs



TF ranking



Regulatory TF

	top.potential.TF.family	freq
All		
1	EMX1	25
2	FOXA1	24
3	GATA3	8
4	LMX1B	5
5	RARA	5
6	ZSCAN16	4
7	PBX1	3
8	HOMEZ	2
9	MYB	2
10	MZF1	2
11	NR2F6	2
12	SPDEF	2
13	ZKSCAN1	2
14	ESR1	1
15	OVOL1	1

Regulatory 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

Oncotarget, 2015 Sep 8;6(26):21878-91.

The pioneer factor PBX1 is a novel driver of metastatic progression in ER α -positive breast cancer.

Magnani L¹, Patten DK¹, Nguyen VT¹, Hong SP¹, Steel JH¹, Patel N¹, Lombardo Y¹, Faronato M¹, Gomes AR¹, Woodley L¹, Page K², Gutierrez D², Primrose L², Fernandez Garcia D², Shaw J², Viola P³, Green A⁴, Nolan C⁴, Ellis IO⁴, Rakha EA⁴, Shousha S¹, Lam EW¹, Györffy B⁵, Lupien M^{6,7}, Coombes RC¹.

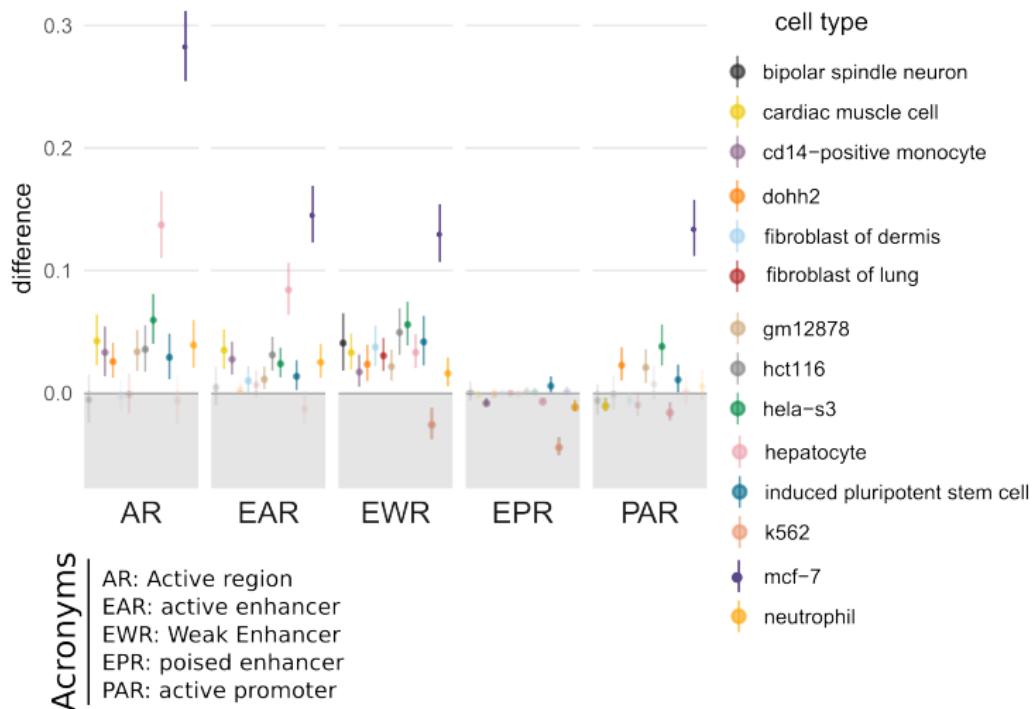
Author information

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 Hierpe, Olle Stål, Barbro K. Linderholm & Janne Lehtio

Annotating chromatin state and verifying enrichment



Comparing inferred results with MCF-7 chIA-PET

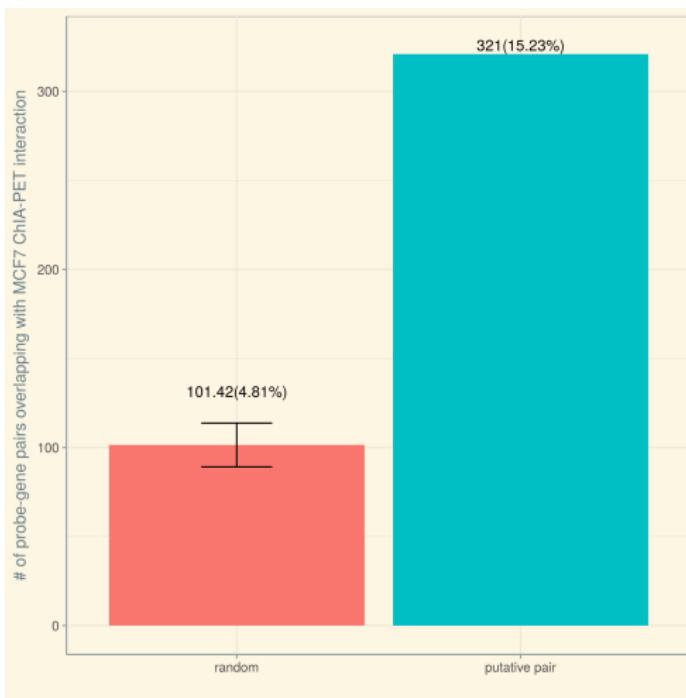


Figure: The graph shows the comparison of the number of probe-gene pairs

Acknowledgments

University of São Paulo

- Houtan Noushmerh



Cedars-Sinai

- Ben Berman
- Simon Coetzee
- Dennis Hazelett
- Nicole Yeager
- Huy Dinh
- Michelle Jones
- Alberto Reyes
- Iveth Corona



CEDARS-SINAI®

References



Ginsburg, Geoffrey S., and Huntington F. Willard.
Genomic and personalized medicine.
Vol. 1. Academic Press, 2008.



Silva TC, Colaprico A, Olsen C et al.
TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages
F1000Research 2016, 5:1542



H. Noushmehr, D. J. Weisenberger, et al.
Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma.
Cancer Cell, 17(5):510(522, May 2010.



S. Sharma, T. K. Kelly, and P. A. Jones.
Epigenetics in cancer.
Carcinogenesis, 31(1):27(36, Jan 2010.



Tiago Chedraoui Silva*, Antonio Colaprico*, et al.
TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res.,
Nucleic acids research, p.gkv1507, 2015



Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, et al. . Cell,
Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma
Cell, 164(3):550-563, 2016



Thank you for your attention!

Any questions?