

Bioinformatic tool to integrate and understand  
aberrant epigenomic and genomic changes  
associated with cancer

## Methods, development and analysis

# Tiago Chedraoui Silva

## PhD Dissertation Defense

University of São Paulo (USP)

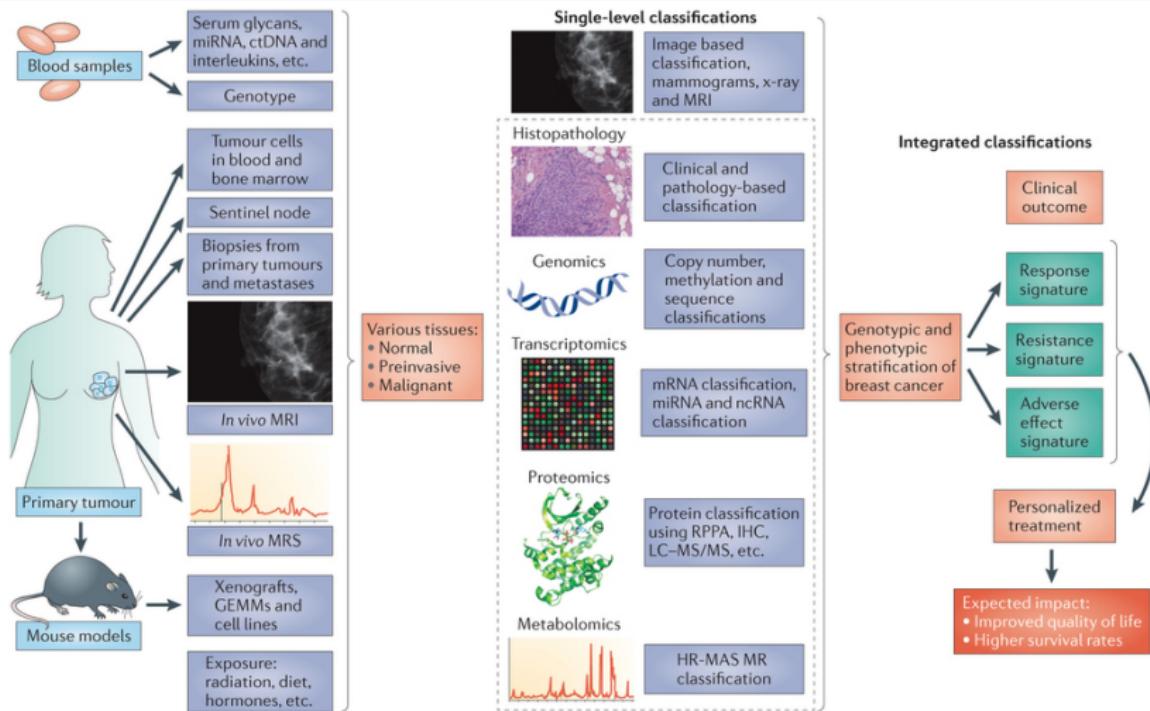
2018-01-30



## Overview

- 1 Introduction
  - 2 Objectives
  - 3 TCGAbiolinks
  - 4 ELMER
  - 5 Case of study
  - 6 Glioma analysis
  - 7 Conclusion

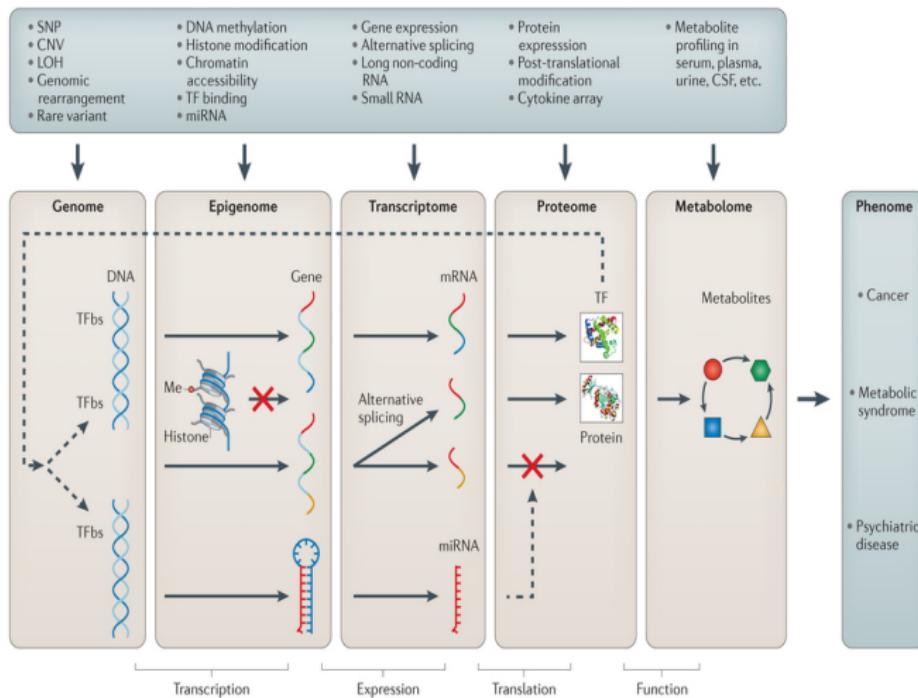
# Context



Nature Reviews | Cancer

Kristensen, Vessela N., et al. Principles and methods of integrative genomic analyses in cancer. Nature Reviews Cancer

# Understand molecular behaviours



Nature Reviews | Genetics

Ritchie, Marilyn D., et al. "Methods of integrating data to uncover genotype-phenotype interactions." *Nature Reviews Genetics*

# Integrative statistical analysis

## Integrative statistical analysis

Analysis of at least two different types of omics data

## Objectives

- ① Understand molecular behaviours, mechanisms and relationships between and within the different types of molecular structures
- ② Understand the taxonomy of diseases, thereby classifying individuals into latent classes of disease subtype
- ③ Predict an outcome or phenotype (survival, efficacy of therapy) for patients.

## Genetics

The study of heritable changes in gene activity or function due to the **direct alteration of the DNA sequence**.

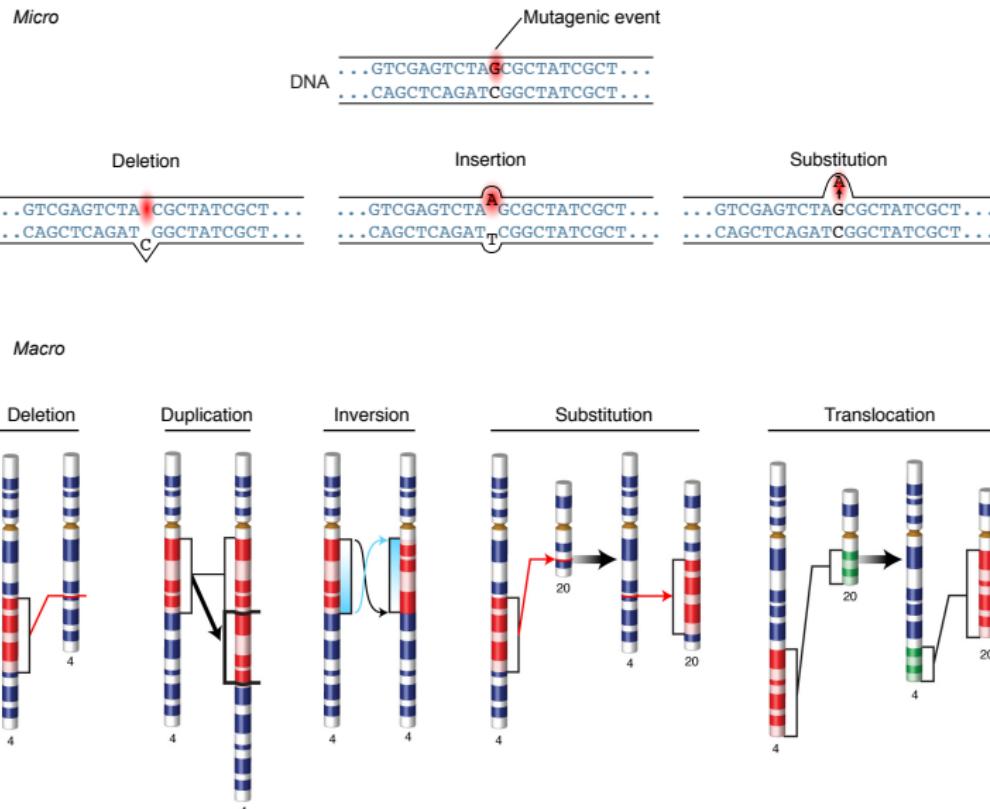
For example, point mutations, deletions, insertions, and translocation.

## Epigenetics

The study of changes in gene activity or function **not associated with any change of the DNA sequence**.

For example, DNA methylation and Histone modifications.

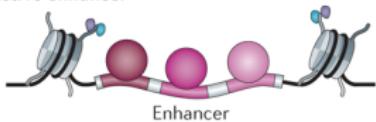
# Genetics alterations



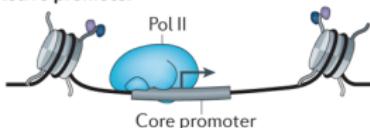
**a Chromatin as accessibility barrier**



**b Active enhancer**



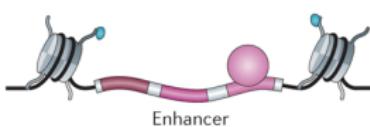
**c Active promoter**



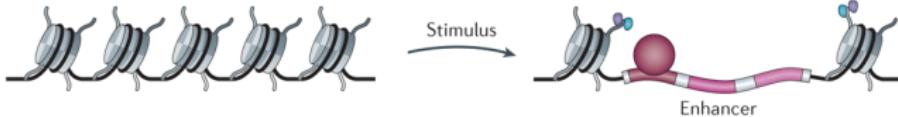
**d Closed or poised enhancer**



**e Primed enhancer**



**f Latent enhancer**



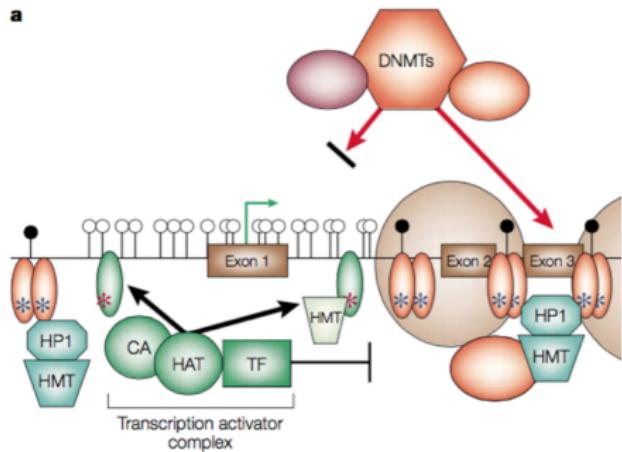
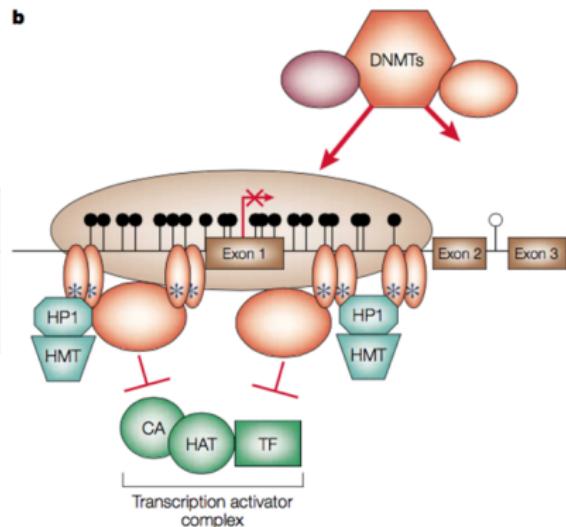
● ● ● TFs

— DNA binding motifs

● ● ● DNA-binding proteins:  
TFs, CTCF, repressors  
and polymerases

● H3K4me1  
● H3K4me3

● H3K27ac  
● H3K27me3

**a****b**

<ul style="list-style-type: none"> <li>● Methylated CpG</li> <li>○ Unmethylated CpG</li> <li>● HDAC</li> </ul>	<ul style="list-style-type: none"> <li>● Methylated Lys4 on H3</li> <li>● Methylated Lys9 on H3</li> <li>● Co-repressor</li> </ul>	<ul style="list-style-type: none"> <li>● MBP</li> <li>● Nucleosome with deacetylated histones</li> <li>● Nucleosome with acetylated histones</li> </ul>
--	--	---

Source: Peter A. Jones & Stephen B. Baylin. Nature Reviews Genetics 3, 415(428 (2002) doi:10.1038/nrg816

# Introduction - Databases

# Encyclopedia of DNA Elements (ENCODE)

Build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements.

# Roadmap Epigenomics Project

Epigenomic maps for stem cells and primary ex vivo tissues selected to represent the normal counterparts of tissues and organ systems frequently involved in human disease.

The Cancer Genome Atlas (TCGA)

Comprehensively map the important genomic changes involved in the major types and subtypes of cancer.

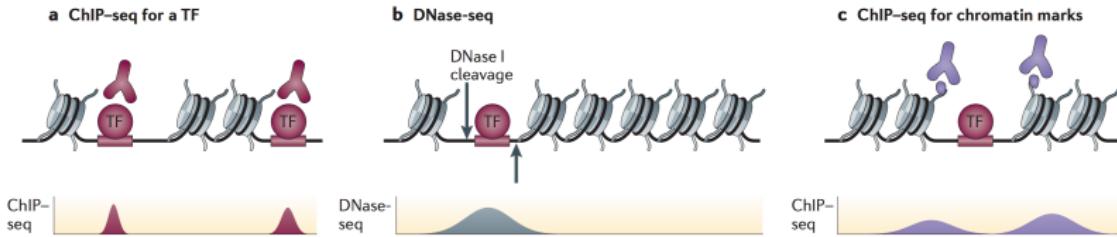
## Integrative analysis

Global transcriptional networks

If the targets of all transcription factors were known, then one could easily infer which transcription factors must be activated in a tumor to yield the observed cancer signature.

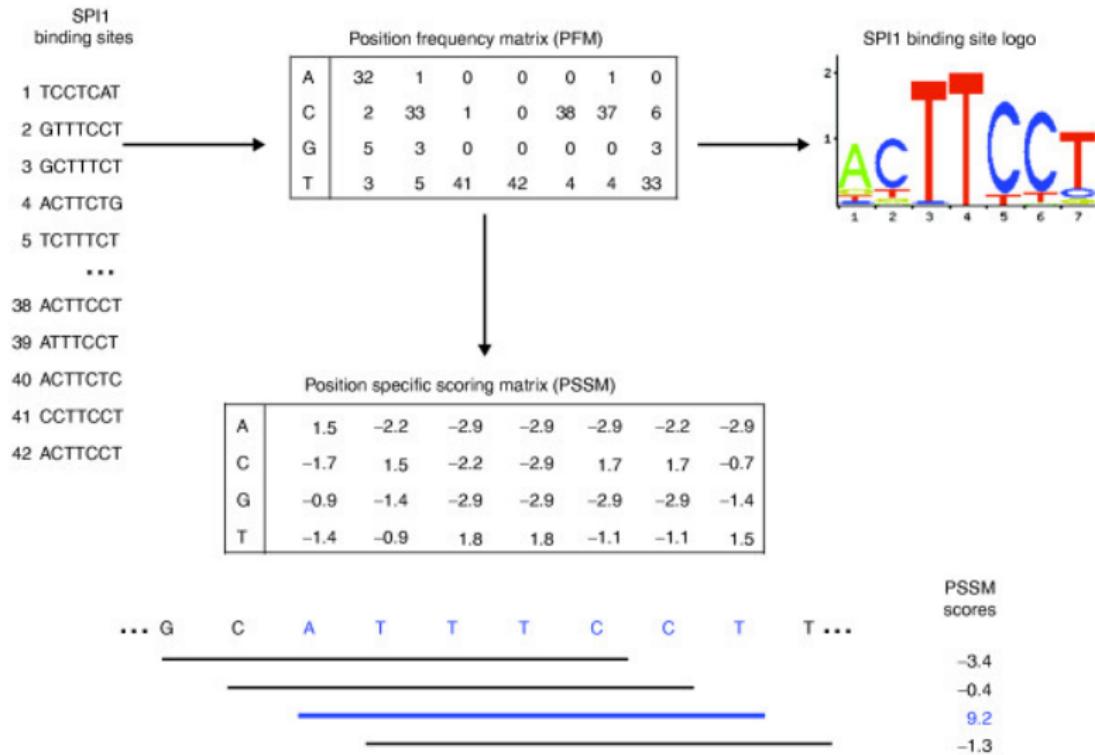
Identification of transcription factor-binding sites (TFBSs)

- Identification using High throughput experimental methods (i.e. ChIP-sequencing)
  - Prediction using in-silico sequence-based methods



Source: Daria Shlyueva, et al. *Nature Reviews Genetics* 15, 272(286 (2014) doi:10.1038/nrg3682

# TFBS prediction

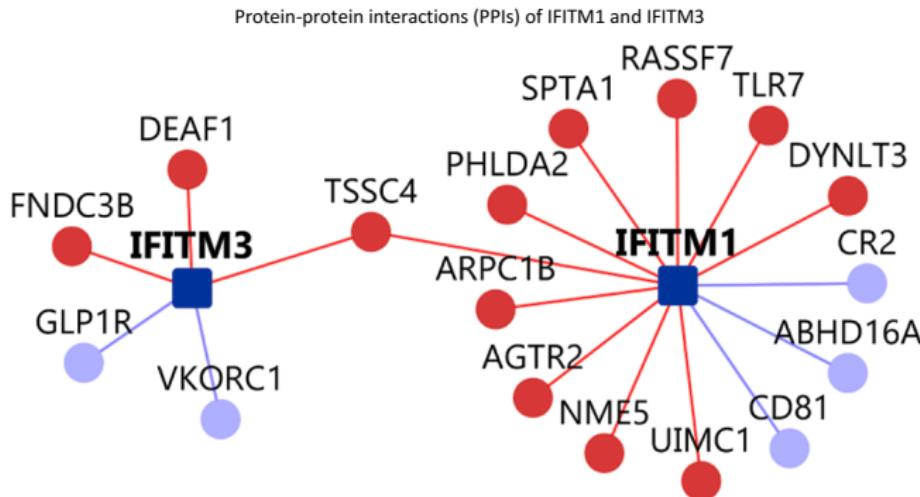


Source: Worsley-Hunt, R., et al. Genome medicine. 2011 Oct 10;3(10):65.

## Integrative analysis

Protein interaction networks and cancer signatures

Protein-protein interactions (PPIs) are essential to almost every process in a cell and are crucial for understanding cell physiology in normal and disease states.



Ganapathiraju MK. Predicted protein interactions of IFITMs which inhibit Zika virus infection [version 1]. F1000Research 2016, 5:1919 (doi: 10.12688/f1000research.9364.1)

F1000Research

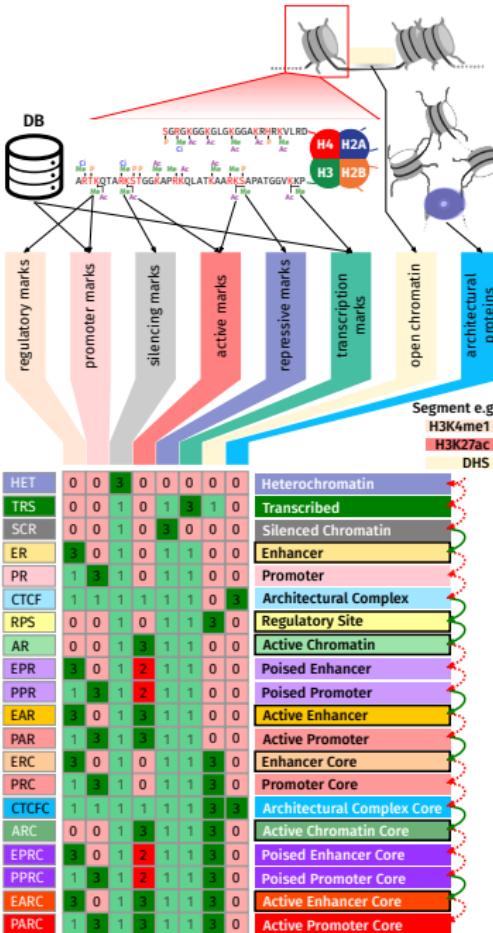
# Integrative analysis

## Epigenome

Regulates gene expression by organizing the nuclear architecture of the chromosomes, restricting or facilitating transcription factor access to DNA, and preserving a memory of past transcriptional activities

Table: Core set of histone modification marks and other epigenomic marks

Epigenomic marks	Role
Histone H3 lysine 4 trimethylation (H3K4me3)	Promoter regions
Histone H3 lysine 4 monomethylation (H3K4me1)	Enhancer regions
Histone H3 lysine 36 trimethylation (H3K36me3)	Transcribed regions
Histone H3 lysine 27 trimethylation (H3K27me3)	Polycomb repression
Histone H3 lysine 9 trimethylation (H3K9me3)	Heterochromatin regions
Histone H3 acetylated at lysine 27 (H3K27ac)	Increase activation of genomic elements
Histone H3 lysine 9 acetylation (H3K9ac)	Transcriptional activation
DNase hypersensitivity	Regions of accessible chromatin
DNA methylation	Repressed regulatory regions



Source: Coetze, Simon G., et al. "StateHub-StatePaintR: rapid and reproducible chromatin state evaluation for custom genome."

## Problem

- Data is spread across different databases and stored in different formats
  - Lack of computational tools and methods that can integrate and interpret such information

Steps performed manually by the user

- ① Access all databases
  - ② Select and process the data necessary to the project
  - ③ Integrate that data using multiple downstream analysis tools to extract
  - ④ Interpret the relevant biological information

## Main goals

- Develop tools for searching, retrieving and analyzing pan-cancer genomic data from several databases (GDC, ENCODE, ROADMAP)
  - Investigate the intergenic epigenomic changes associated with distinct biological and clinical subgroups of gliomas first discovered by our laboratory

## Secondary goals

- Use standard data structure to organize the data and the metadata
  - Publish tools the open-source Bioconductor environment.
  - Use learning machine algorithms for classifying an independent set of gliomas based on newly identified regulatory networks as related to pan-cancer deregulation;

# Tools

## TCGAbiolinks

platforms all downloads top 5% posts 17 / 2 / 2 / 3 In Bioc 2 years  
build ok

DOI: [10.18129/B9.bioc.TCGAbiolinks](https://doi.org/10.18129/B9.bioc.TCGAbiolinks)  

TCGAbiolinks: An R/Bioconductor package for integrative analysis with GDC data

## TCGAbiolinksGUI

platforms all downloads top 20% posts 0 In Bioc 0.5 years  
build ok

DOI: [10.18129/B9.bioc.TCGAbiolinksGUI](https://doi.org/10.18129/B9.bioc.TCGAbiolinksGUI)  

"TCGAbiolinksGUI: A Graphical User Interface to analyze cancer molecular and clinical data"

## ELMER

platforms all downloads top 20% posts 0 In Bioc 2 years  
build ok

DOI: [10.18129/B9.bioc.ELMER](https://doi.org/10.18129/B9.bioc.ELMER)  

Inferring Regulatory Element Landscapes and Transcription Factor Networks Using Cancer Methylomes

# TCGAbiolinks functions

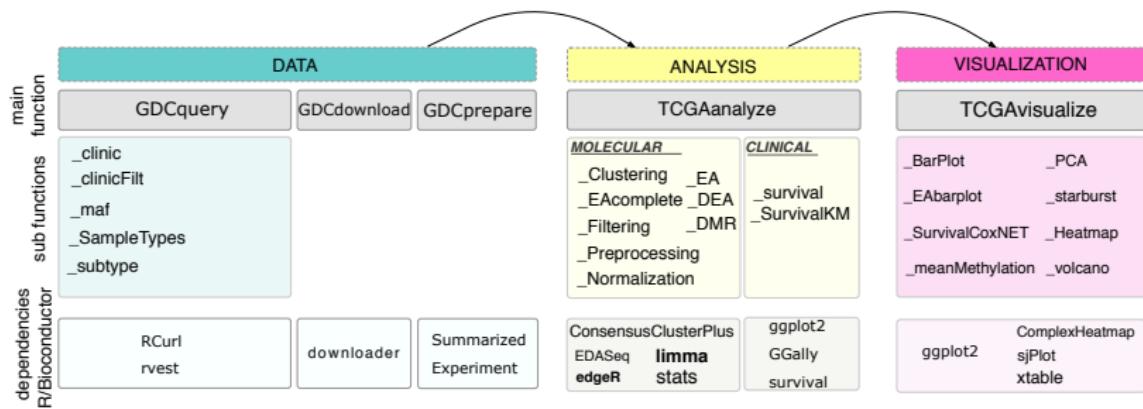


Figure: Overview of TCGAbiolinks functions

## Comparing TCGAbiolinks to competing software

		Packages						
Features	Sub-features	TCGAAbioliinks	TCGAAAssembler	canEnvolve	TCGA2stat	Firehose-FirebrowseR	RTCGAToolbox	cBio Portal CGDS-R
Availability	Platform	B	R	W	C	CW	B	CW
Genome of reference	Access to data aligned against the GRCh38/hg38	X	X					
	Access to data aligned against the GRCh37/hg19	X	X	X	X	X	X	X
Query TCGA Cases	Individual TCGA samples (e.g. TCGA-01-0001)	X	X			X		
Download	All TCGA platforms	X						
Data type analysis	mRNA	X		X	X	X	X	X
	miRNA	X		X	X	X	X	X
	copy number	X		X	X	X	X	X
	DNA methylation	X			X	X	X	X
	Clinical	X		X	X	X	X	X
	Protein			X		X		X
	Mutation	X		X	X	X	X	X
Integrative analysis	DNA methylation and gene expression	X				X		
Other	Extensible to other BioC packages	X						

# Data structure: SummarizedExperiment

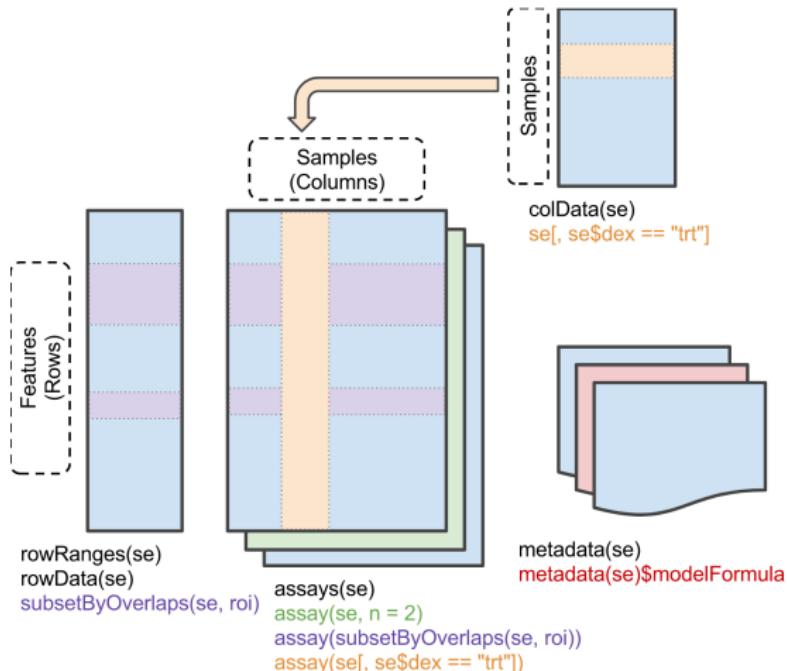


Figure: Example of Sumarized Experiment object. Figure reproduced from SummarizedExperiment manual (MORGAN et al., 2017).

# Graphical user interface (GUI)

**GDC Data**

- Get GDC data
- Molecular data
- Mutation data
- Clinical data
- Subtype data

**Analysis**

- Clinical analysis
- Epigenetic analysis
- Transcriptomic analysis
- Genomic analysis

**Integrative analysis**

- Starburst plot
- ELMER

**Configuration**

- Configuration

**Help Documents**

- Tutorial/Vignettes
- Need help ?
- References

**Prepare completed**

Saved in:

```
/home/shiny/TCGAbiolinksGUI/TCGA-LUSC_DNA_Methylation__hg38.rda
/home/shiny/TCGAbiolinksGUI/TCGA-LUSC_DNA_Methylation__hg38_samples_information.csv
```

**GDC search results: Summary**

**Experimental strategy**

Category	Value
Project	TCGA-LUSC
File size (MB)	~200
Sample	Methylation Array 100%

**Tissue definition**

Category	Value
Project	Primary solid Tumor 100%

**Data type**

Category	Value
Project	Methylation Beta Value 100%

**Vital status**

Category	Value
Project	dead 100%

**Tumor stage**

Category	Value
Project	stage Ib 50% stage Ia 50%

**Race**

Category	Value
Project	white 50% not reported 50%

**Gender**

Category	Value
Project	male 100%

**Molecular data search**

**Database:**

- Harmonized database (hg38)
- Legacy database (hg19)

**Project filter:** Lung Squamous Cell Carcinoma (TCGA-LUSC)

**Data Category filter:** DNA Methylation

**Platform filter:** Illumina Human Methylation 450

**Sample type filter:**

**Barcode:** TCGA-34-S231-01  
TCGA-77-7138-01

**Clinical filters:**

**Visualize Data**

**Download & Prepare**

**Data types:**

- SummarizedExperiment
- Dataframe
- Add gistic2 and mutation information ?

**File name:** TCGA-LUSC\_DNA\_Methylation\_\_hg38.rda

**Download and prepare data**

# Enhancer Linking by Methylation/Expression Relationship

## ELMER



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

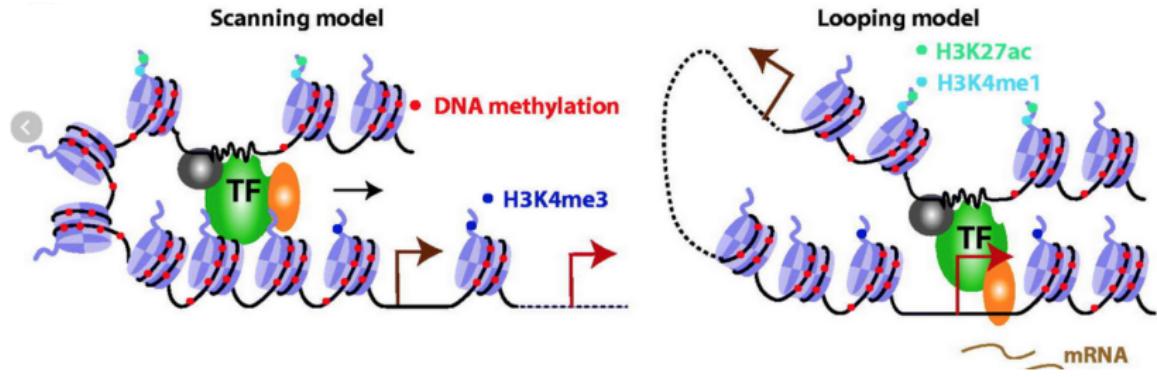


Lijing Yao<sup>1</sup>, Hui Shen<sup>2</sup>, Peter W Laird<sup>2</sup>, Peggy J Farnham<sup>1\*</sup> and Benjamin P Berman<sup>1,3\*</sup>

### 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.

# Enhancer-mediated gene regulation



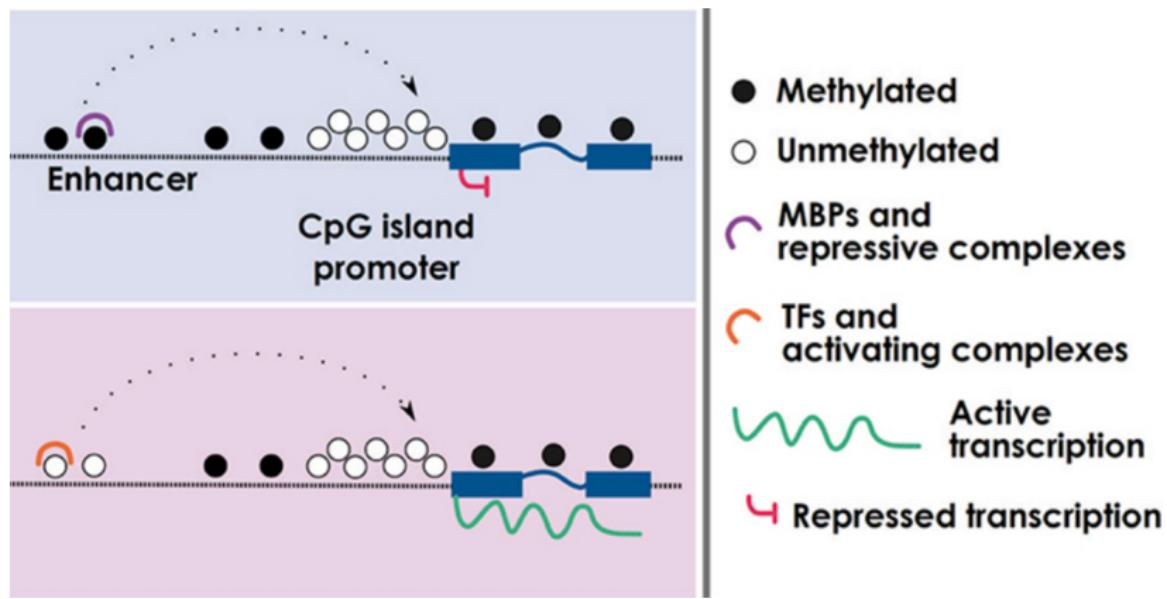
Source: Yao et al. *Genome Biology* (2015) 16:105.



# 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>.

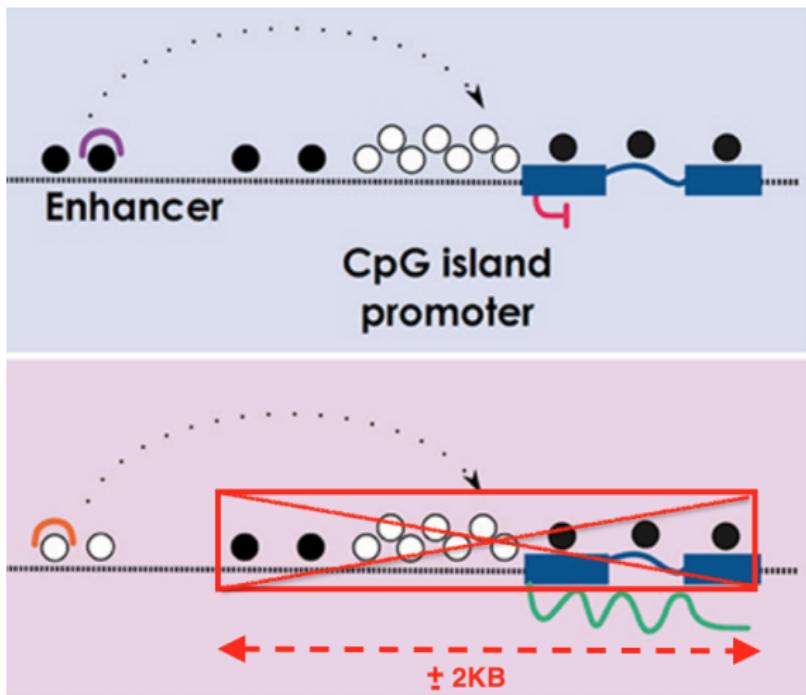


# Algorithm

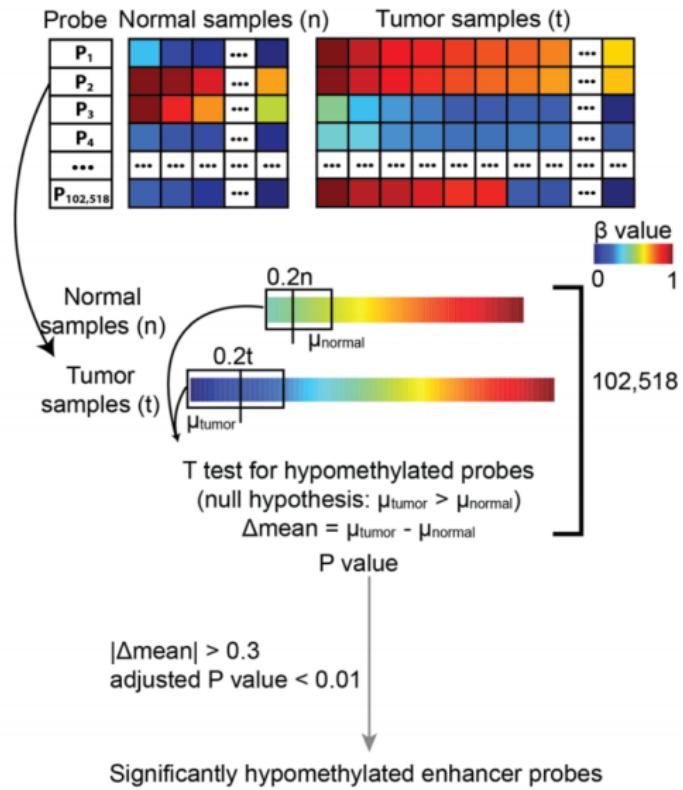
## Steps

- 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 putative target gene.
- ⑤ Identify regulatory TFs whose expression associate with DNA methylation at motifs.

## Step 1: Identify distal probes



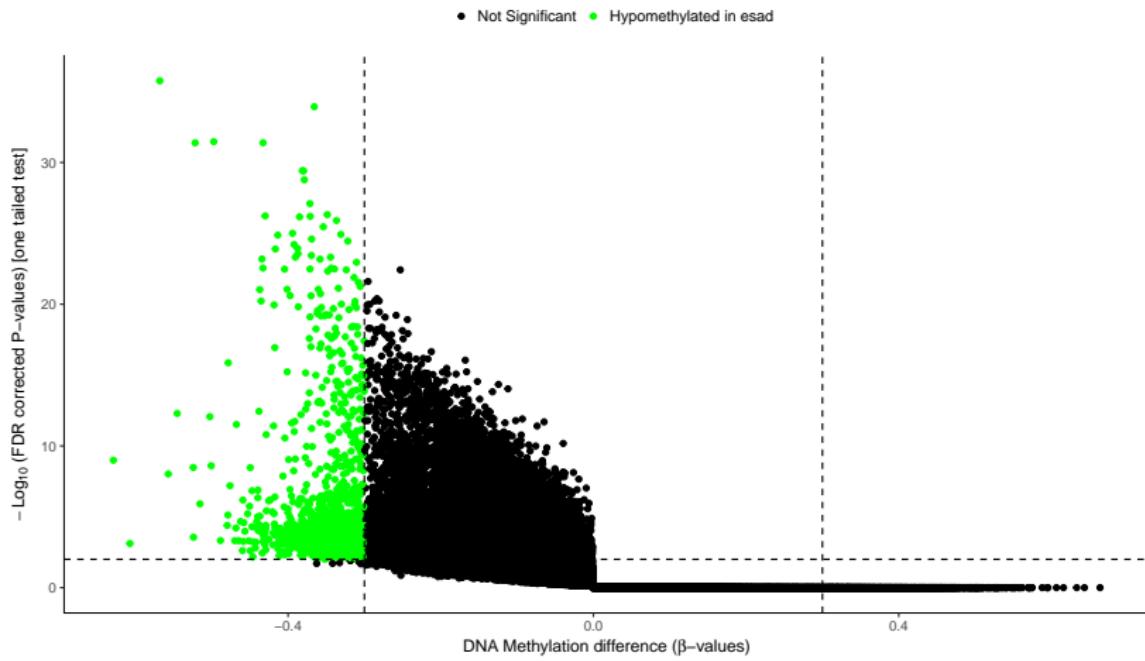
## Step 2: Differentially methylated distal probes



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

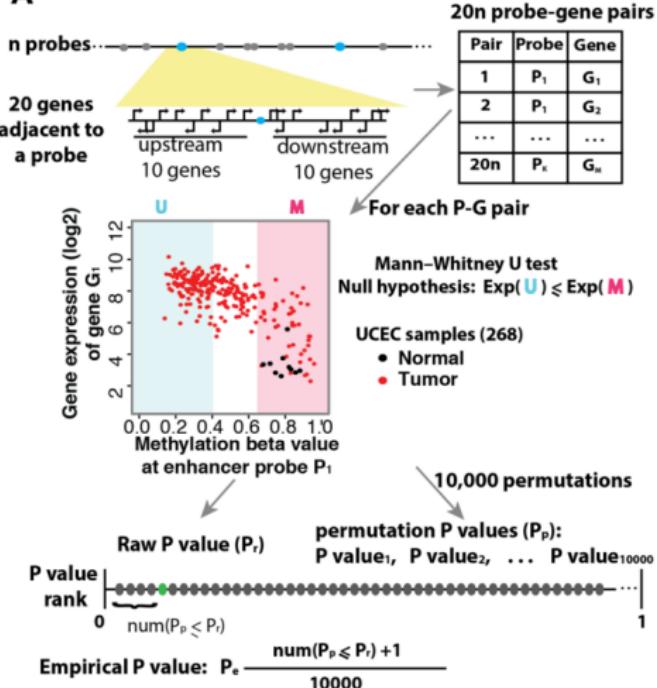
## Step 2: Differentially methylated distal probes

### Volcano plot – Probes hypomethylated in esad vs normal

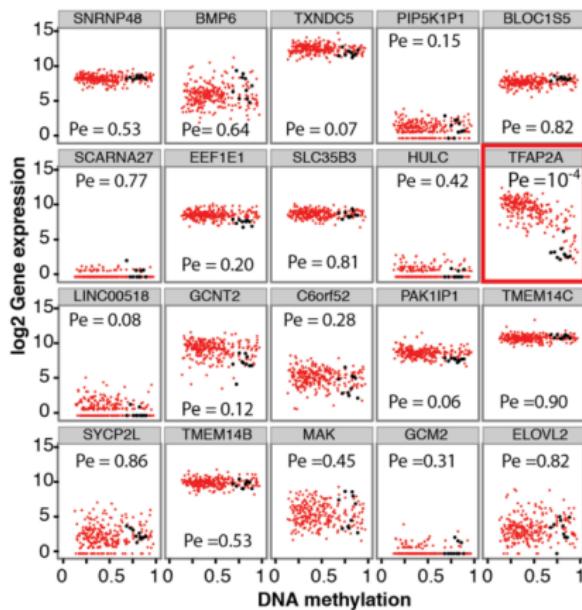


## Step 3: Identification of putative target gene(s)

**A**

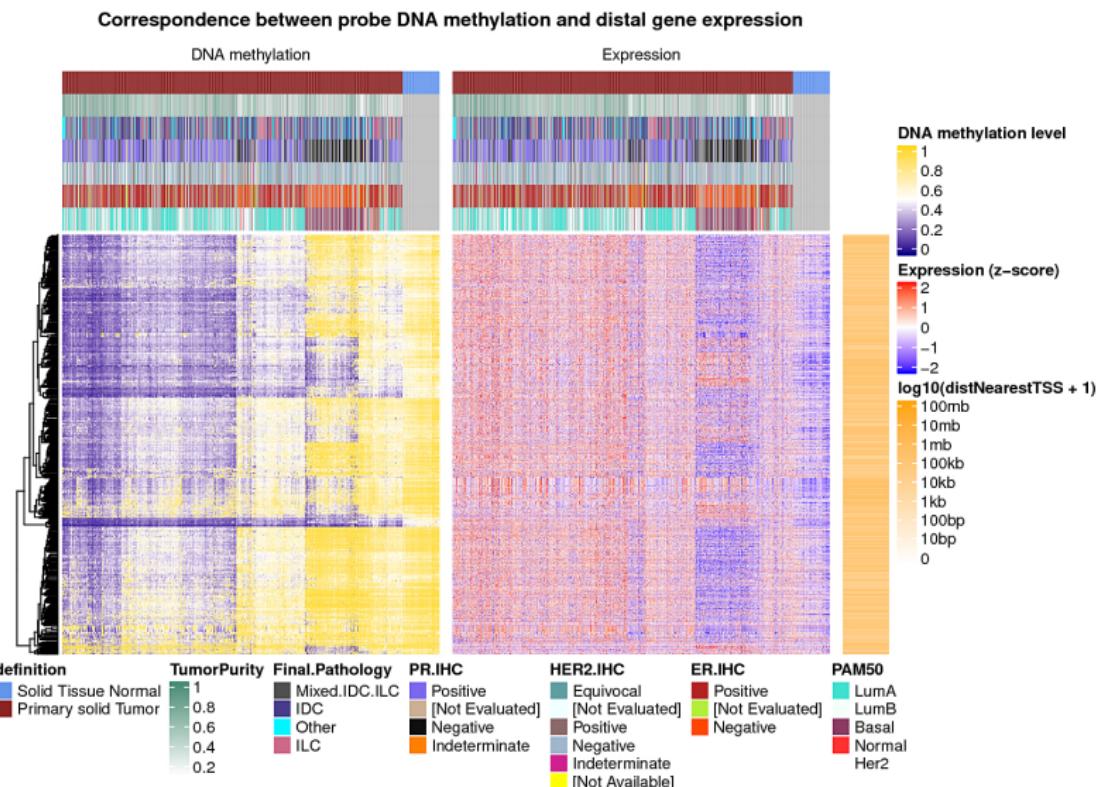


**B**

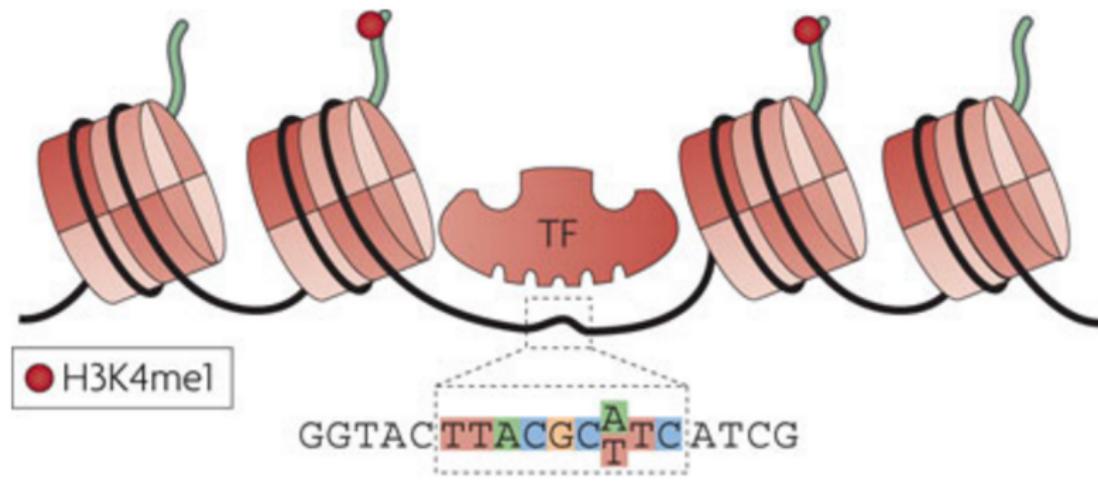


Source: Yao et al. Genome Biology (2015) 16:105.

# Step 3: Probe-target gene pairs inferred



## Step 4: Motif enrichment analysis



Nature Reviews | Genetics

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

## 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.

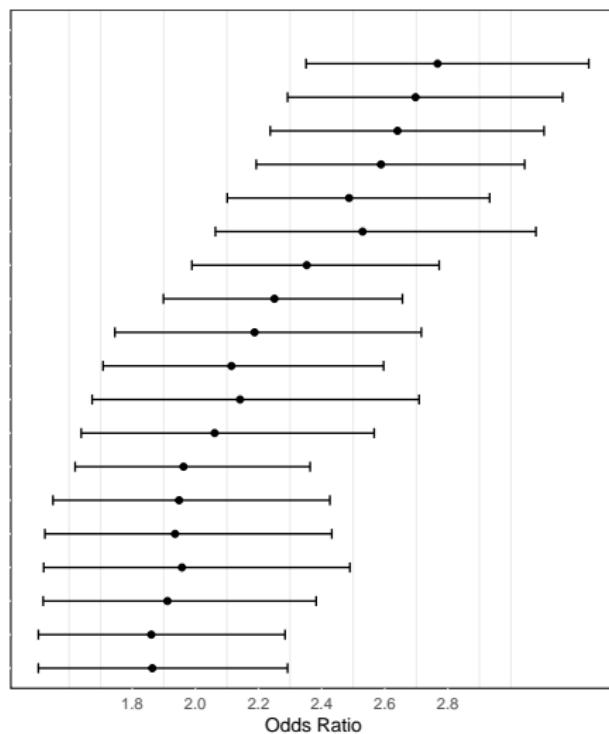
- ① Perform motif matching of transcription factors in probes regions (window  $\pm 250\text{bp}$ ). Performed using HOMER (Hypergeometric Optimization of Motif EnRichment) with HOCOMOCO motifs.
- ② Evaluate which transcription factors are more likely to occur in those regions than in background regions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

### Fisher's exact test

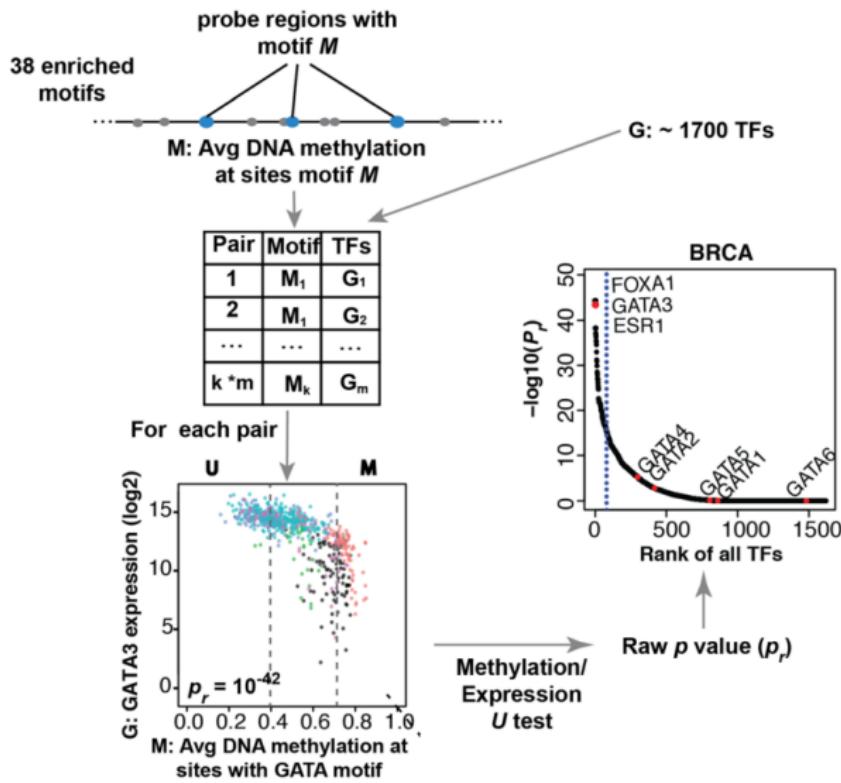
- a: nb of input regions with match for TF motif.
- b: nb of input regions with no match for TF motif.
- c: nb of background regions with match for TF motif.
- d: nb of background regions with no match for TF motif.

## 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) 16:105.

## 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

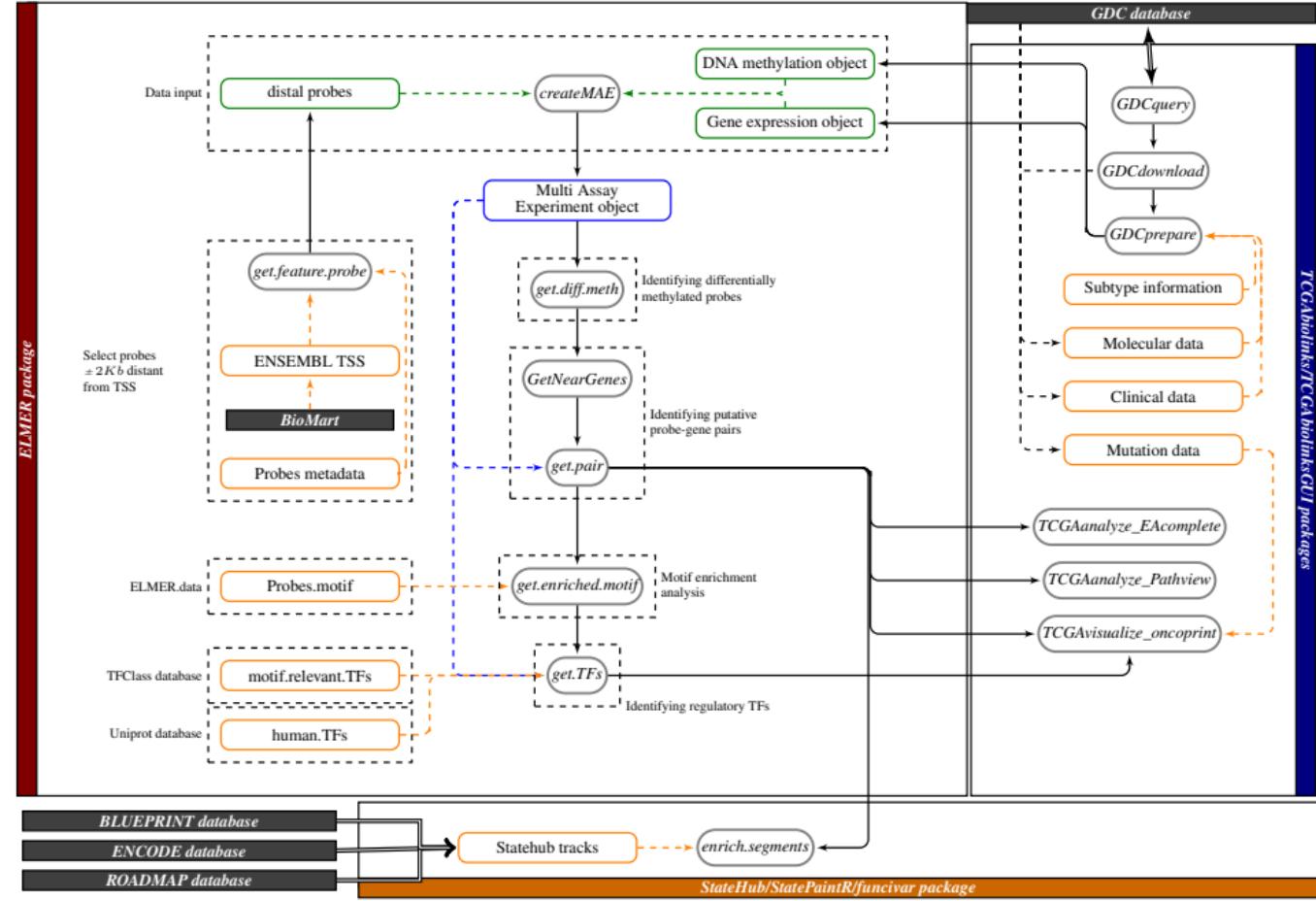
Showing 1 to 10 of 31 entries

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

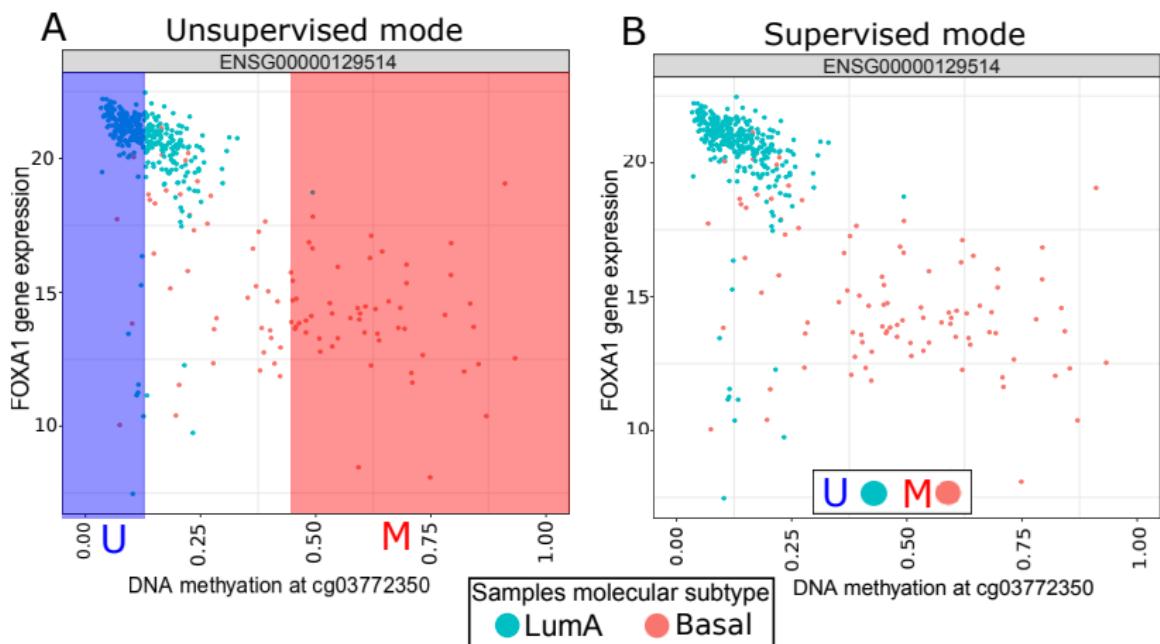


Figure: A: *unsupervised mode*; when `minSubgroupFrac` argument is set to 40%, the methylated group is defined as the highest quintile and the unmethylated group as the lowest quintile; B: *supervised mode*; methylated and unmethylated group are defined as one of the known molecular subtypes.

## Case study: TCGA Breast Invasive Carcinoma (BRCA)

Table: Summary of the available samples in TCGA for BRCA

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

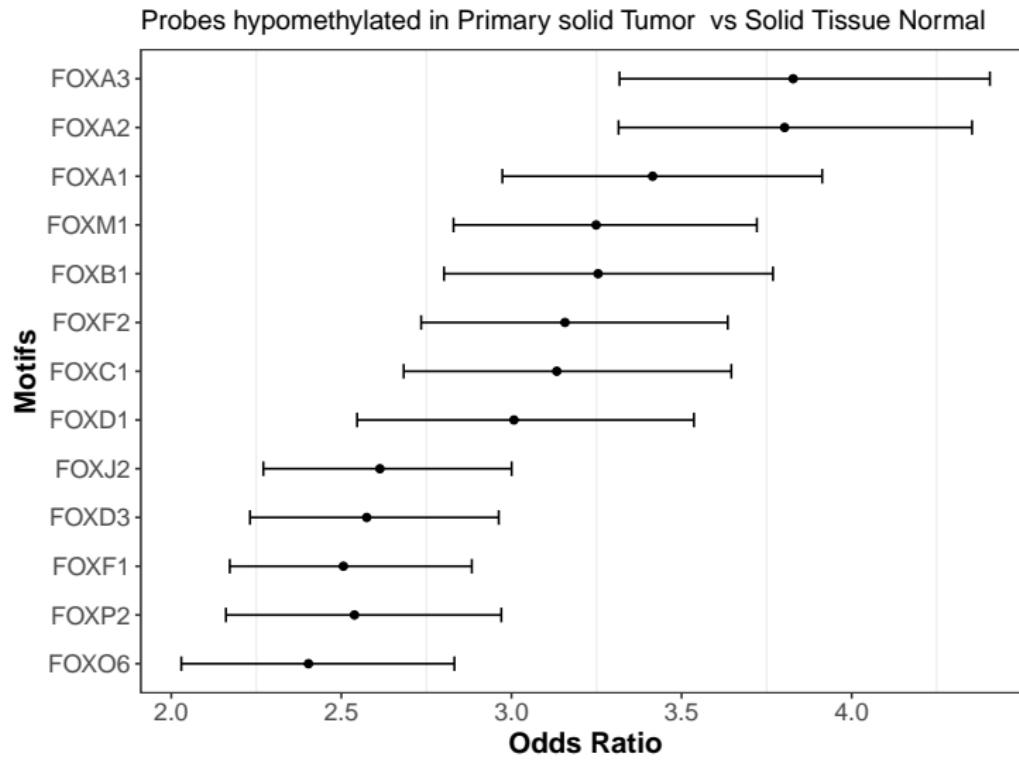
Table: Results supervised mode

---

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

---

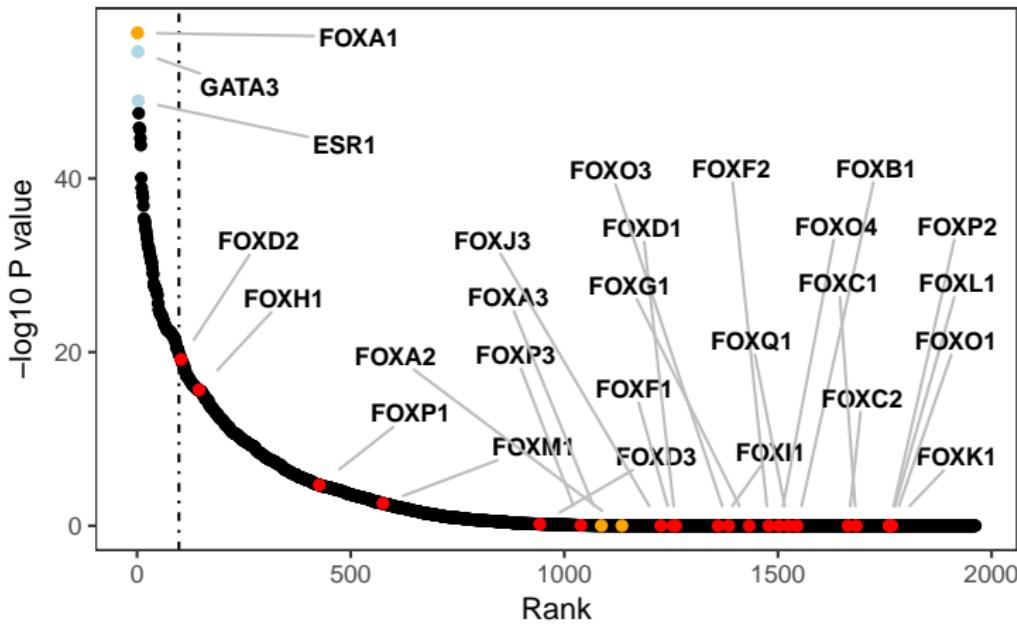
# Top enriched motifs



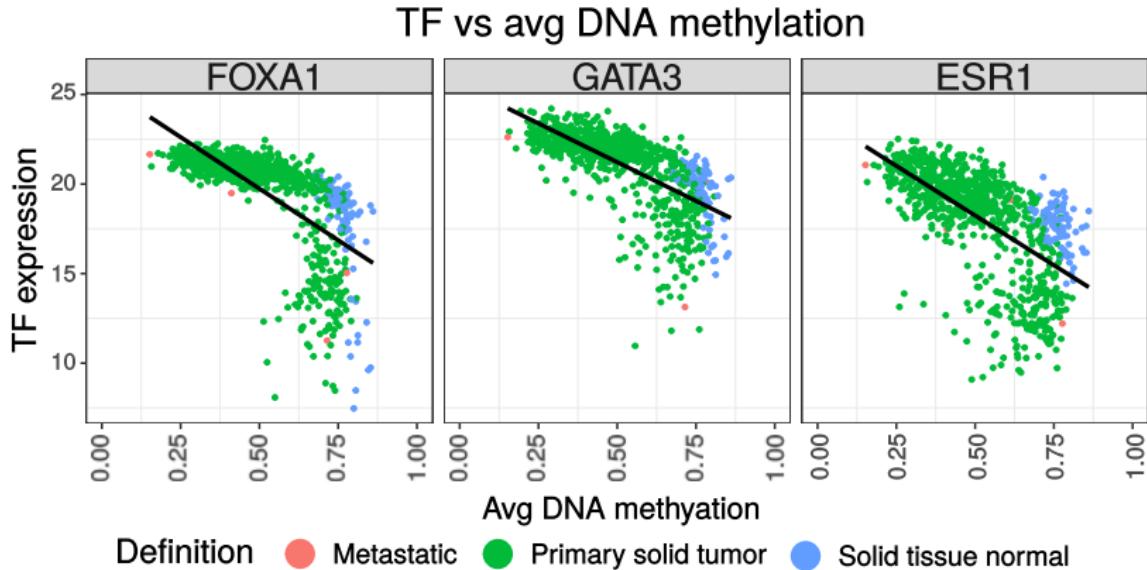
## TF ranking plot - FOXA3 motif

## Motif: FOXA3

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



## DNA methylation at motifs vs TF expression



# Candidate 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 ERα-positive breast cancer.

Magnani L<sup>1</sup>, Patten DK<sup>1</sup>, Nguyen VT<sup>1</sup>, Hong SP<sup>1</sup>, Steel JH<sup>1</sup>, Patel N<sup>1</sup>, Lombardo Y<sup>1</sup>, Faronato M<sup>1</sup>, Gomes AR<sup>1</sup>, Woodley L<sup>1</sup>, Page K<sup>2</sup>, Guttry D<sup>2</sup>, Primrose L<sup>2</sup>, Fernandez Garcia D<sup>2</sup>, Shaw J<sup>2</sup>, Viola P<sup>3</sup>, Green A<sup>4</sup>, Nolan C<sup>4</sup>, Ellis IO<sup>4</sup>, Rakha EA<sup>4</sup>, Shousha S<sup>1</sup>, Lam EW<sup>1</sup>, Győrffy B<sup>5</sup>, Lupien M<sup>6,7</sup>, Coombes RC<sup>1</sup>.

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 Hjerpe, Olle Stål, Barbro K. Linderholm & Janne Lehtiö

# Characterization of chromatin state context of enriched probes using FunciVar



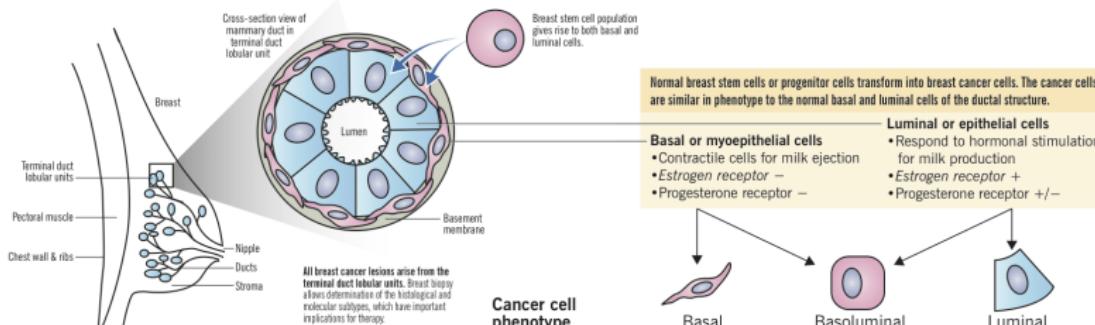
**Acronyms**

- 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	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	<b>Triple negative</b> ER-, PR-, HER2-	<b>HER2+</b> HER2	<b>Luminal B</b> High (grade III) Level of cell differentiation	<b>Luminal A</b> Low (grade I)
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+	<b>Poor</b> Correlates to histologic grade	<b>Good</b>	<b>Chemotherapy</b> Trastuzumab	<b>Endocrine</b> Luminal A tumours respond best to endocrine therapy, e.g. antiestrogen or aromatase inhibitor.

Curr Opin Oncol. 2000 Aug;13(3):199-209.  
Clin Transl Oncol. 2008 Dec;10(12):777-85.

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

# BRCA supervised analysis

Groups annotation were taken from: <https://doi.org/10.1016/j.cell.2015.09.033>

x	freq
<fctr>	<int>
Basal	85
Her2	34
LumA	288
LumB	117
Normal	22
NA	314
6 rows	

Argument	Value
<fctr>	<fctr>
Mode	Supervised
All: minSubgroupFrac	100%
DNA methylation differences: min mean difference	0.3
DNA methylation differences: p-value adj cut-off	0.01
Pairs correlation: # permutations	10000
Pairs correlation: raw p-value cut-off	0.001
Pairs correlation: empirical p-values cut-off	0.001
Enrichement motif: minimum # probes (enriched motif)	10
Enrichement motif:lower.OR	1.1
9 rows	

## Supervised analysis: Candidate regulatory TFs

## TF master regulator: LumA

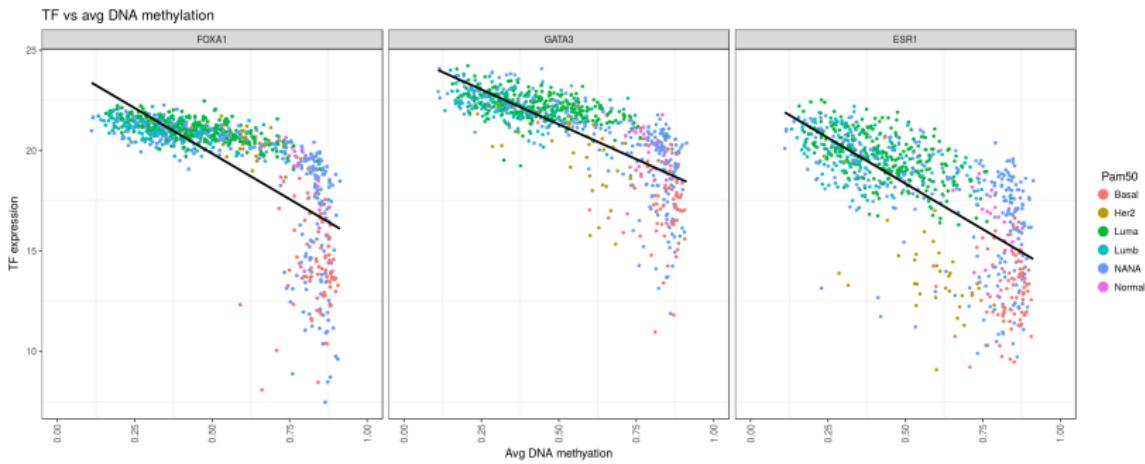


Figure: FOXA1 and top3 TF expression vs avg DNA methylation of paired enriched probes for FOXA3 motif - Probes hypomethylated in LumA vs Normal-like

# TF master regulator: Basal

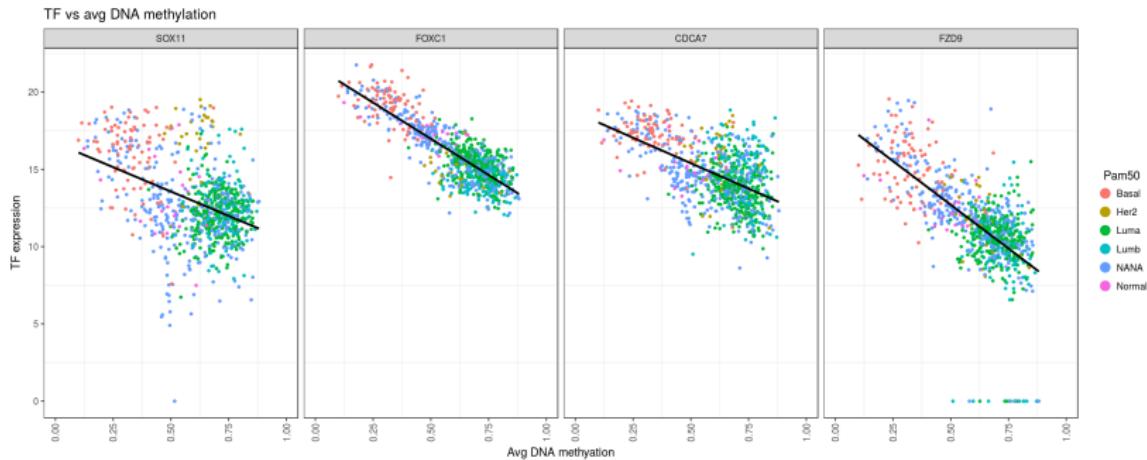


Figure: SOX11 and top3 TF expression vs avg DNA methylation of paired enriched probes for SOX10 - Probes hypermethylated in LumA vs Basal

## TF master regulator: Basal

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

**Jonathan H. Shepherd<sup>1,3</sup>, Ivan P. Uray<sup>3</sup>, Abhijit Mazumdar<sup>3</sup>, Anna Tsimelzon<sup>2</sup>, Michelle Savage<sup>3</sup>, Susan G. Hilsenbeck<sup>2</sup>, Powell H. Brown<sup>1,3</sup>**

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

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

negative breast cancer cell lines to regain hormonal sensitivity.<sup>41</sup> 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.<sup>13</sup> These findings together highlight a critical role of FOXA1 in maintaining the luminal and

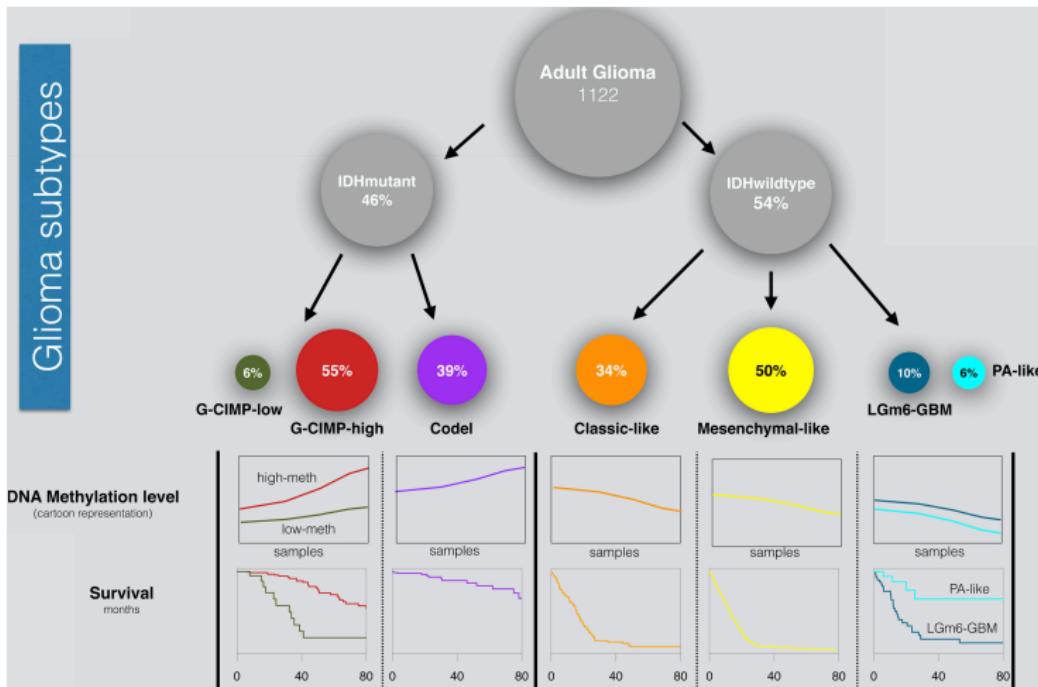
### GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland.

Kouros-Mehr H<sup>1</sup>, Slorach EM, Sternlicht MD, Werb Z.

## Author information

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

# Glioma analysis: glioma subgroups



Source: Ceccarelli, Michele, et al. Cell 164.3 (2016): 550-563. DOI: <http://dx.doi.org/10.1016/j.cell.2015.12.028>

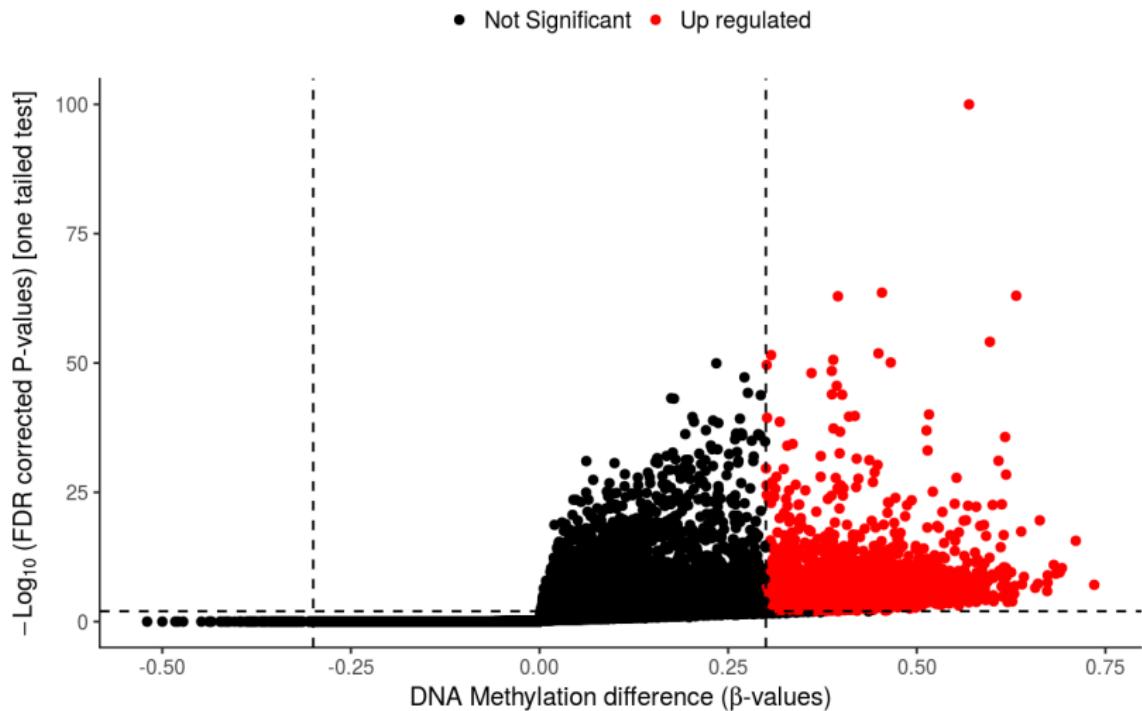
# G-CIMP supervised analysis

x	freq <int>
G-CIMP-high	233
G-CIMP-low	11
2 rows	

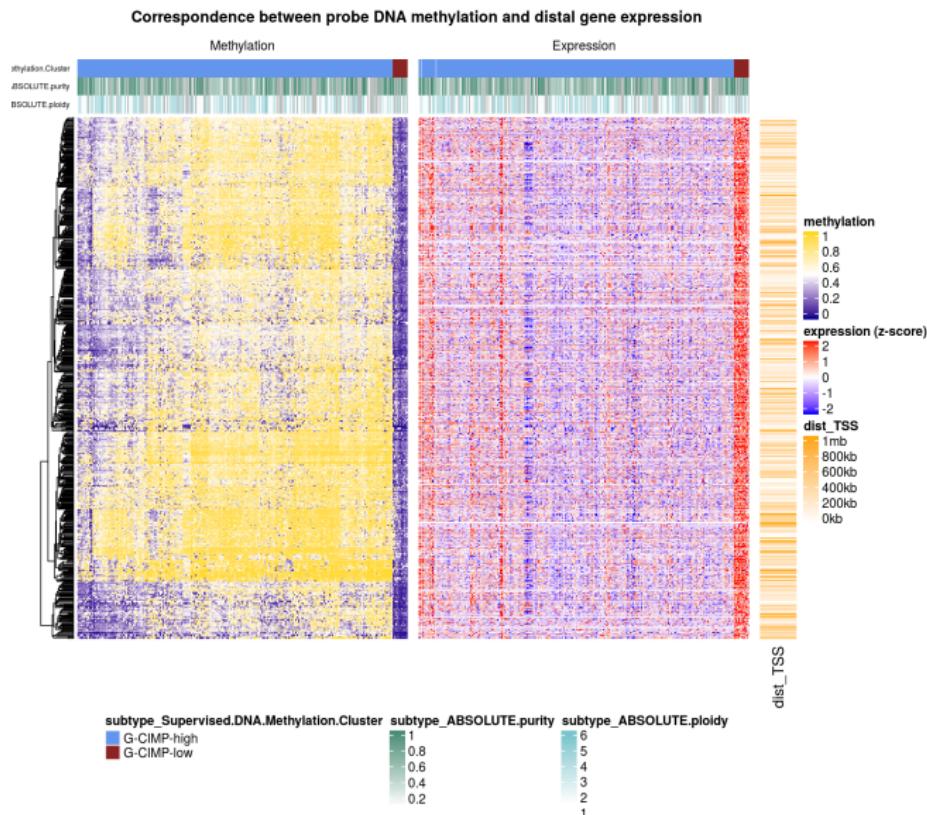
Argument	Value
<fctr>	<fctr>
createMAE: genome	hg38
Mode	supervised
All: minSubgroupFrac	100%
DNA methylation differences: min mean difference	0.3
DNA methylation differences: p-value adj cut-off	0.01
Pairs correlation: # permutations	10000
Pairs correlation: raw p-value cut-off	0.001
Pairs correlation: empirical p-values cut-off	0.001
Enrichement motif: minimun # probes (enriched motif)	10
Enrichement motif:lower.OR	1.1
1-10 of 10 rows	

## Differentially methylated distal probes

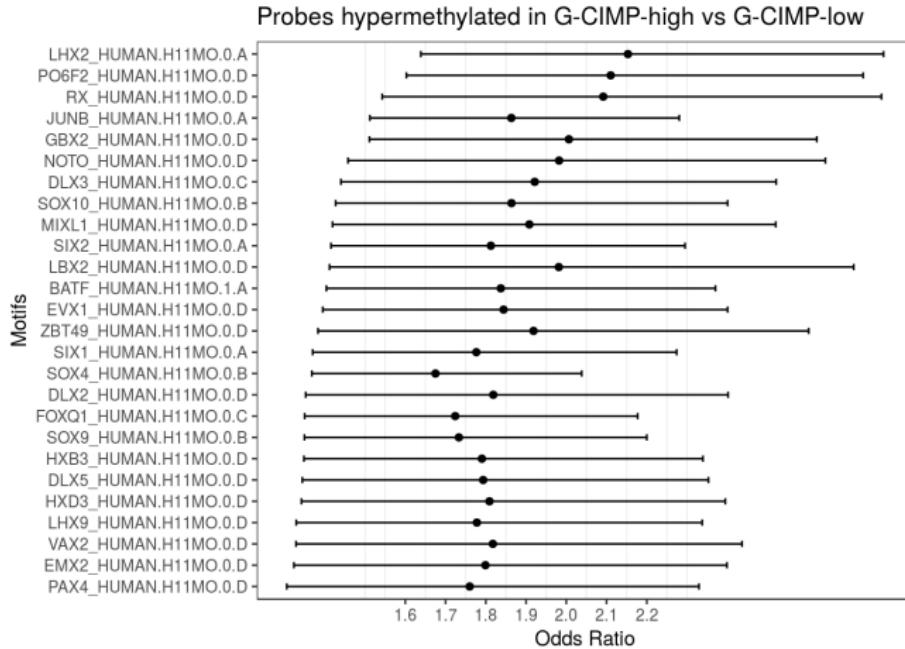
## Volcano plot - Probes hypermethylated in gcimp-high vs g-cimp-low



# Paired gene-probes



# Enriched motifs



# Candidate master regulator Transcription Factors (TF)

## Top TFs

- HOXD13
- HMX3
- ZSCAN16
- SOX11
- PAX3
- FOXM1

HOXD13

# DNA methylation profiles of long- and short-term glioblastoma survivors

Thoria Shinawi,<sup>1</sup> Victoria K. Hill,<sup>1</sup> Dietmar Krex,<sup>2</sup> Gabriele Schackert,<sup>2</sup> Dean Gentle,<sup>1</sup> Mark R. Morris,<sup>3</sup> Wenbin Wei,<sup>4</sup> Garth Cruickshank,<sup>5</sup> Fionnán R. Maher<sup>1</sup> and Farida Latifi<sup>1,\*</sup>

<sup>1</sup>Centre for Rare Diseases and Personalized Medicine; Department of Medical & Molecular Genetics; School of Clinical and Experimental Medicine; University of Birmingham College of Medical and Dental Sciences; Birmingham, UK; <sup>2</sup>Department of Neurosurgery; University Hospital Dresden; Dresden, Germany; <sup>3</sup>School of Applied Sciences; University of Wolverhampton; Wolverhampton, UK; <sup>4</sup>School of Cancer Sciences; University of Birmingham; Birmingham, UK; <sup>5</sup>Department of Neurosurgery; University of Birmingham; Queen Elizabeth Hospital; Birmingham, UK

**Keywords:** DNA methylation, gliomas, IDH1, long-term survivors, short-term survivors

Glioblastoma (GBM) is the most common and malignant type of primary brain tumor in adults and prognosis of most GBM patients is poor. However, a small percentage of patients show a long term survival of 36 mo or longer after diagnosis. Epigenetic profiles can provide molecular markers for patient prognosis: recently, a G-CIMP positive phenotype associated with *IDH1* mutations has been described for GBMs with good prognosis. In the present analysis we performed genome-wide DNA methylation profiling of short-term survivors (STS; overall survival < 1 y) and long-term survivors (LTS; overall survival > 3 y) by utilizing the HumanMethylation450K BeadChips to assess quantitative methylation at > 480,000 CpG sites. Cluster analysis has shown that a subset of LTS showed a G-CIMP positive phenotype that was tightly associated with *IDH1* mutation status and was confirmed by analysis of the G-CIMP signature genes. Using high stringency criteria for differential hypermethylation between non-cancer brain and tumor samples, we identified 2,638 hypermethylated CpG loci (890 genes) in STS GBMs, 3,101 hypermethylated CpG loci (1,062 genes) in LTS (wild type *IDH1*) and 11,293 hypermethylated CpG loci in LTS (mutated for *IDH1*), reflecting the CIMP positive phenotype. The location of differentially hypermethylated CpG loci with respect to CpG content, neighborhood context and functional genomic distribution was similar in our sample set, with the majority of CpG loci residing in CpG islands and in gene promoters. Our preliminary study also identified a set of CpG loci differentially hypermethylated between STS and LTS cases, including members of the homeobox gene family (*HOXD8*, *HOXD13* and *HOXC4*), the transcription factors *NR2F2* and *TFAP2A*, and *Dickkopf 2*, a negative regulator of the wnt/β-catenin signaling pathway.

# FOXM1 and PAX3

## FoxM1: a potential drug target for glioma

Yu Li<sup>1</sup>, Sicong Zhang<sup>1</sup>, and Suyun Huang<sup>1,2</sup>

<sup>1</sup>Department of Neurosurgery, the University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

### Abstract

Malignant glioma is an aggressive disease and there is no effective therapy. Recently, with the elucidation of mechanisms for glioma formation and progression, the critical molecules involved in the process are considered as therapeutic targets and numerous of drugs against these targets are ongoing for evaluation in clinic trial. FoxM1 has been recognized as one of the common pathways in cancer cells including glioma cells. FoxM1 signal network is reported to be critical in glioma by promoting cell proliferation, invasion, angiogenesis and cancer stem cell self-renewal. FoxM1 may represent a novel therapeutic target and FoxM1 inhibitors may provide a new therapeutic strategy against glioma.

### Keywords

FoxM1; glioma; proliferation; invasion; angiogenesis; self-renewal; drug target

### PAX3 is overexpressed in human glioblastomas and critically regulates the tumorigenicity of glioma cells.

Xia L<sup>1</sup>, Huang Q, Nie D, Shi J, Gong M, Wu B, Gong P, Zhao L, Zuo H, Ju S, Chen J, Shi W.

#### ✉ Author information

#### Abstract

Paired box 3 (PAX3) is overexpressed in glioma tissues compared to normal brain tissues, however, the pathogenic role of PAX3 in human glioma cells remains to be elucidated. In this study, we selected the human glioma cell lines U251, U87, SHG-44, and the normal human astrocytes, 1800, which have differential PAX3 expression depending upon the person. siRNA targeting PAX3 and PAX3 overexpression vectors were transfected into U87 and SHG-44 glioma cell lines, and cell proliferation, invasion, apoptosis, and differentiation were examined by CCK-8 assays, transwell chamber assays, tunnel staining, Annexin V/PI analysis, and Western blotting, respectively. In addition, we used subcutaneous tumor models to study the effect of PAX3 on the growth of glioma cells in vivo. We found that PAX3 was upregulated in the three glioma cell lines. PAX3 knockdown inhibited cell proliferation and invasion, and induced apoptosis in the U87MG glioblastoma cell line, whereas PAX3 upregulation promoted proliferation, inhibited apoptosis, and increased invasion in the SHG-44 glioma cell line. Moreover, we found that targeting PAX3 expression in glioma cell lines together with chemotherapeutic treatment could increase glioma cell susceptibility to

SOX11

## **Highly specific overexpression of the transcription factor SOX11 in human malignant gliomas.**

Weigle B<sup>1</sup>, Ebner B, Temme A, Schwind S, Schmitz M, Kiessling A, Rieger MA, Schackert G, Schackert HK, Rieber EP

## Author information

### Abstract

Malignant glioma comprises the majority of primary human brain tumors with 16,800 new cases reported each year in the USA. Its prognosis remains dismal despite numerous attempts to improve conventional therapeutic modalities. Therefore, much effort is devoted to the exploration of alternative forms of treatment such as immunotherapy. The identification of potential target structures highly overexpressed in brain tumors is a crucial prerequisite for the activation of the immune defense against malignant glioma cells. By screening an expression database for genes highly expressed in glioblastoma multiforme (GBM), we identified the Pit-Oct-Unc (POU) cooperating transcription factor SOX11 that is known to be crucially involved in brain development. Analysis of the expression pattern of SOX11 in different normal adult and fetal tissues by multiple tissue dot blot and by a highly sensitive quantitative PCR assay confirmed the selective overexpression of SOX11 in fetal brain tissue. Examination of tissue specimens obtained from malignant gliomas and from normal brain by quantitative real-time PCR (Q-RT-PCR) revealed upregulation of SOX11 in almost all tumor samples (15/16) as compared to the pooled normal brain. Seventy-five percent of the tumor samples (12/16) showed a 5- to more than 600-fold overexpression. We conclude that, after downregulation of SOX11 in the adult brain, its expression is reactivated during tumorigenesis and that SOX11 therefore represents a promising novel molecular target for adjuvant therapy of malignant gliomas.

PMID: 15583815

## Next steps: TF knockdown

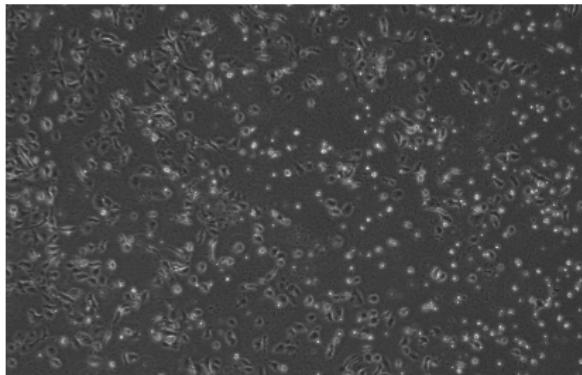
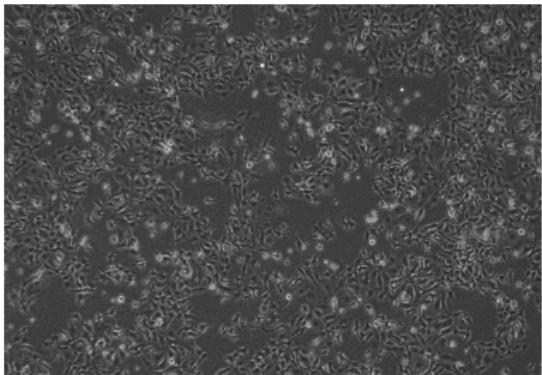


Figure: Candidate master regulator Transcription Factors (TF) knockdown in the SKGT4 human esophageal adenocarcinoma cell line. Figure produced by Dr. Dechen Lin.

# Conclusion

## Conclusion

- Tools were able to provide methods for (epi) genomics integrative analysis using data from several databases.
- Several genes identified using those tools are consistent with literature.

## Future works

- Extend ELMER algorithm to work with regions (WGBS).
- Provided a processing pipeline for WGBS data.
- Use mutation information instead of DNA methylation to identify regulatory regions mutated that might have affected the regulation of upstream/downstream genes.

#### Results: First authorship - articles, software and material online

-  **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
  -  **Silva TC, Colaprico A, Olsen C et al.**  
TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages F1000Research 2016, 5:1542
  -  **Silva, Tiago C., et al.**  
"TCGAbiolinksGUI: A graphical user interface to analyze GDC cancer molecular and clinical data." bioRxiv (2017): 147496.
  -  **Silva, Tiago Chedraoui, et al.**  
"Enhancer Linking by Methylation/Expression Relationships with the R package ELMER version 2." bioRxiv (2017): 148726.
  -  **Workflow:**  
TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages  
<https://www.bioconductor.org/help/workflows/TCGAWorkflow/>
  -  **Workshop:**  
Integrative analysis workshop with TCGAbiolinks and ELMER  
<https://bioinformaticsfmrp.github.io/Bloc2017.TCGAbiolinks.ELMER/index.html>
  -  **Softwares:**  
<http://bioconductor.org/packages/TCGAbiolinksGUI/>  
<http://bioconductor.org/packages/TCGAbiolinks/>  
<http://bioconductor.org/packages/ELMER/>

## Results: Co-authorship - articles and software

-  M Ceccarelli, et al.  
Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma  
*Cell*, 2016
  -  Malta, Tathiane M., et al.  
Glioma CpG island methylator phenotype (G-CIMP): biological and clinical implications  
*Neuro-oncology* (2017).
  -  Lin, De-Chen, et al.  
Identification of distinct mutational patterns and new driver genes in oesophageal squamous cell carcinomas and adenocarcinomas  
*Gut*, pp.gutjnl-2017.
  -  Cava, Claudia, et al.  
SpidermiR: An R/Bioconductor Package for Integrative Analysis with miRNA Data  
*International journal of molecular sciences*. 2017 Jan 27;18(2):274.
  -  Softwares:  
<http://bioconductor.org/packages/SpidermiR/>

## Acknowledgments

University of São Paulo

- Houtan Noushmerh
  - Camila Ferreira de Souza
  - Tathiane Malta



Cedars-Sinai

- Ben Berman
  - Simon Coetzee
  - Dennis Hazelett
  - Nicole Yeager
  - Dechen Lin



# 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?