

R/BioConductor tools for functional interpretation of methylation changes in cancer DNA

Ben Berman, PhD

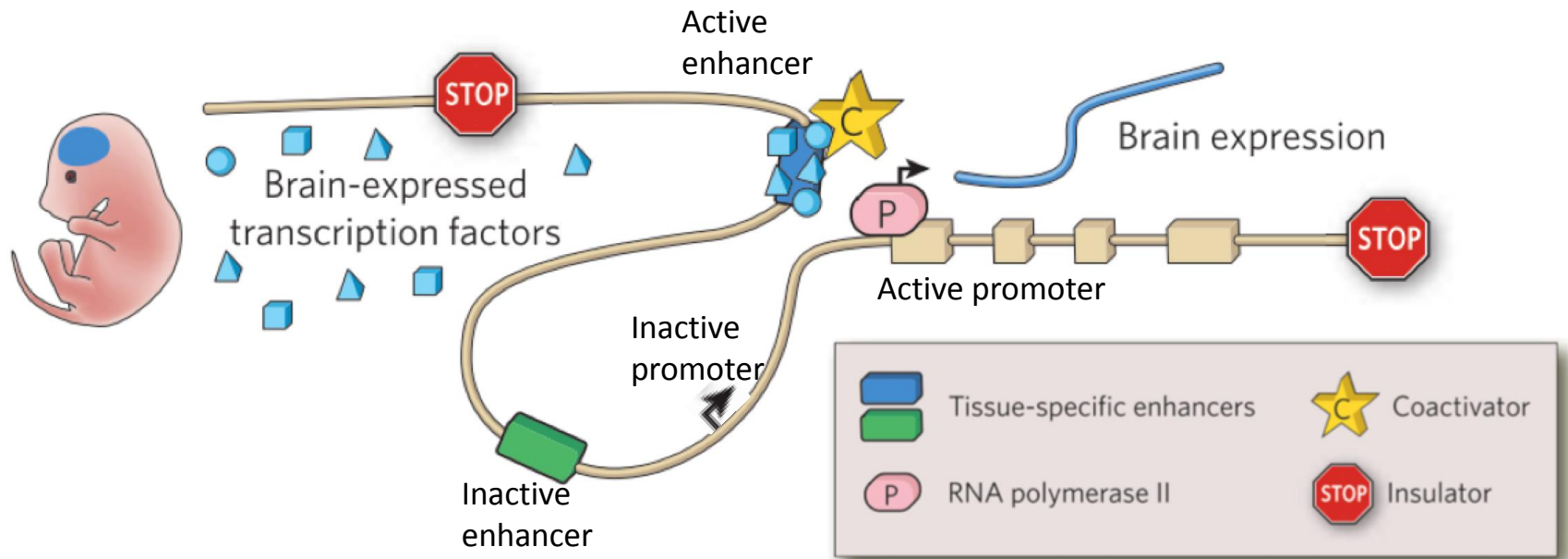
Center for Bioinformatics & Functional Genomics



CEDARS-SINAI MEDICAL CENTER®

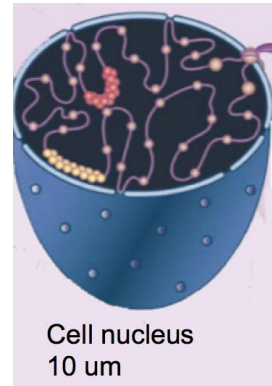
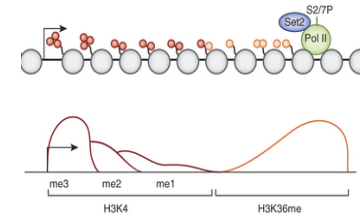
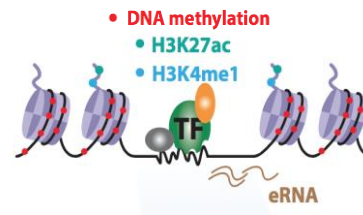
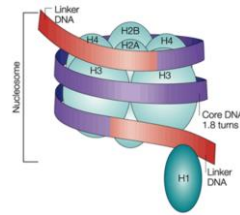
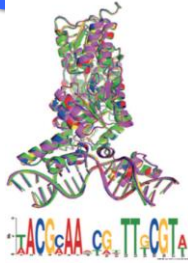
Los Angeles, CA

Key features of transcriptional regulation



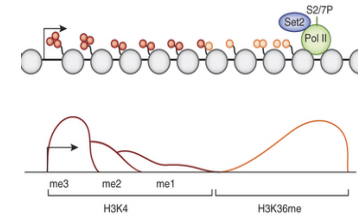
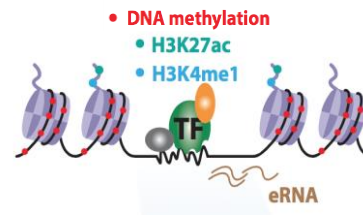
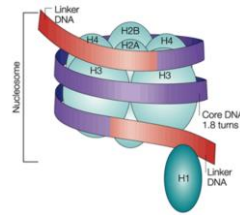
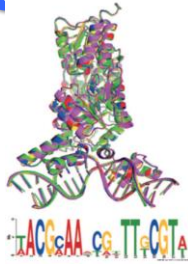
Visel et al. *Nature* 2009, 461(7261), 199–205

Mapping the chromatin structures of gene regulation (i.e. epigenomics)



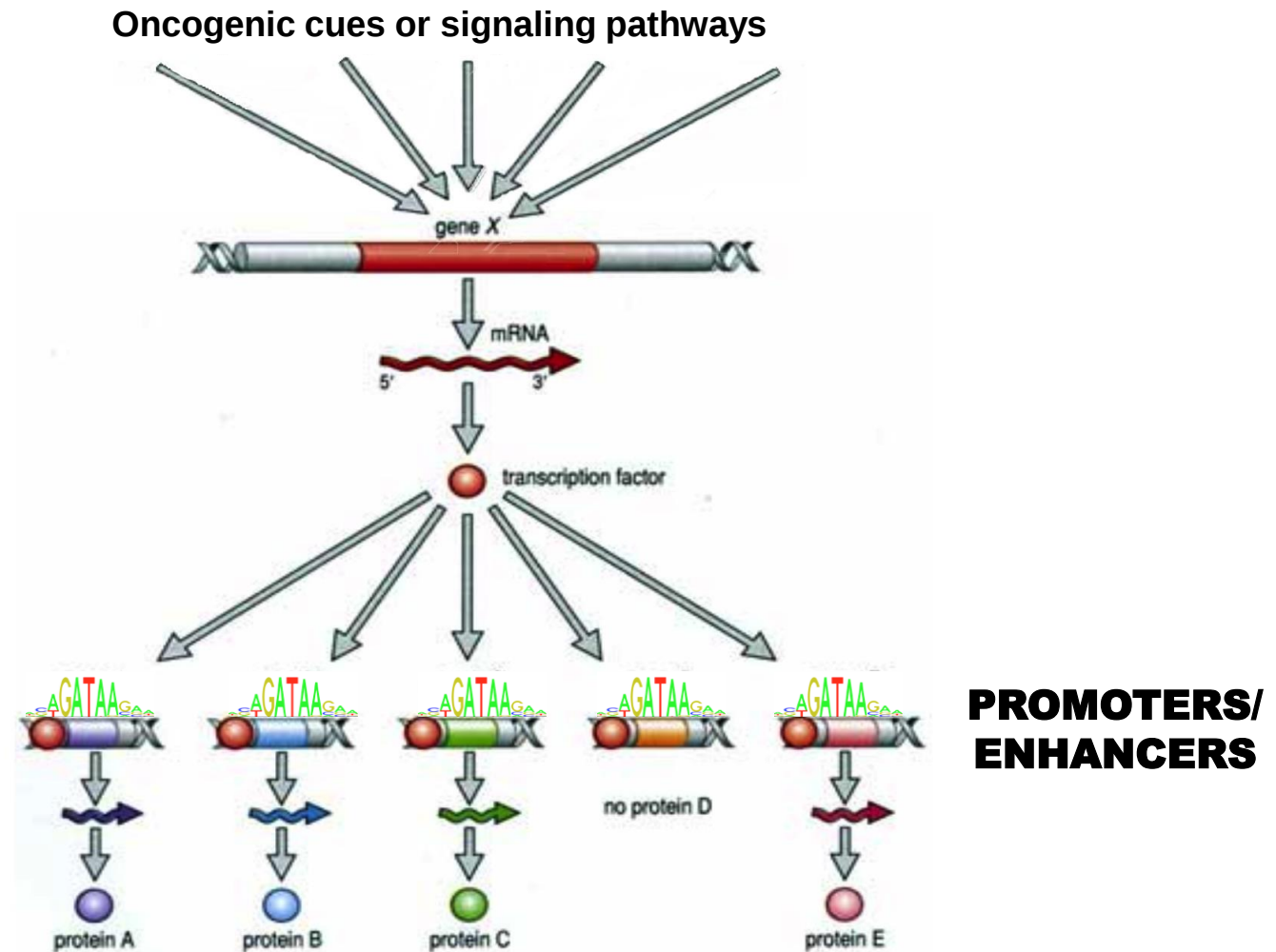
Feature	TFBS	Nucleosome positioning	Promoter, enhancer, insulator	Super-enhancer, PRC2 domain	Transcription unit	Nuclear topological domain
Genomic size	5-30bp	40-150bp	200-2,000bp	5-20kb	10-100kb	100kb-10Mb
ENCODE/Roadmap methods	ChIP-seq, DNase-seq	Mnase-seq, (ATAC-seq)	ChIP-seq, ATAC-seq, DNase-seq	ChIP-seq	ChIP-seq	Hi-C ChIA-PET

Mapping the chromatin structures of gene regulation (i.e. epigenomics)



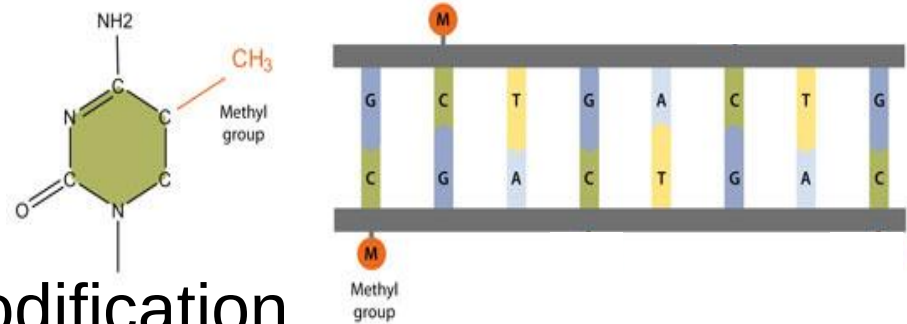
Feature	TFBS	Nucleosome positioning	Promoter, enhancer, insulator	Super-enhancer, PRC2 domain	Transcription unit	Nuclear topological domain
Genomic size	5-30bp	40-150bp	200-2,000bp	5-20kb	10-100kb	100kb-10Mb
ENCODE/Roadmap methods	ChIP-seq, DNase-seq	Mnase-seq, (ATAC-seq)	ChIP-seq, ATAC-seq, DNase-seq	ChIP-seq	ChIP-seq	Hi-C ChIA-PET

Transcription factors control downstream cancer networks through cis-regulatory elements



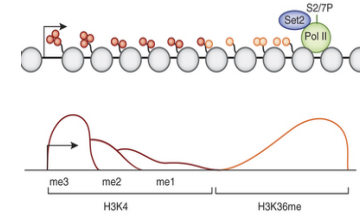
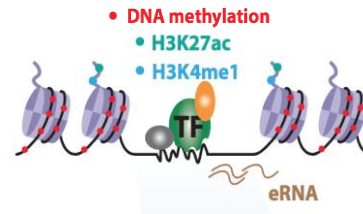
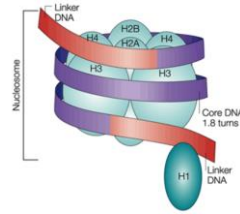
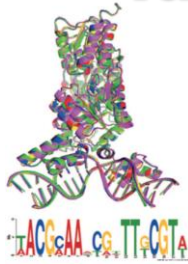
Modified from ScienceBlogs

DNA methylation is a popular chromatin biomarker, but often treated as a black box



- Highly stable covalent modification
 - Degraded DNA (e.g. FFPE) can be assayed with minimal loss of accuracy
 - Cell-free DNA can be detected in blood with high sensitivity
- Unlike other methods, bisulfite converts epigenetic states into Single Nucleotide Variants
 - Highly accurate estimation of percent cells methylated
 - Multiple “SNVs” per read in NGS provide even greater accuracy
 - SNP array technology (Infinium bead arrays)

Cancer chromatin structures that can be identified by DNA methylation



Feature	TFBS	Nucleosome positioning	Promoter, enhancer, insulator	Super-enhancer, PRC2 domain	Transcription unit	Nuclear topological domain
Genomic size	5-30bp	40-150bp	200-2,000bp	5-20kb	10-100kb	100kb-10Mb
ENCODE/Roadmap methods	ChIP-seq, DNase-seq	Mnase-seq, (ATAC-seq)	ChIP-seq, ATAC-seq, DNase-seq	ChIP-seq	ChIP-seq	Hi-C ChIA-PET
Has DNA methylation signature?	No?	Yes (2012)	Yes Promoters 1981 Enhancers 2011	Yes (2010-2013)	No?	Yes (2011)

Kelly et al.
Genome Research, 2012

Berman et al.
Nature Genetics, 2011

Berman et al.
Nature Genetics, 2011

Enhancer DNA methylation identifies cellular origins of pediatric medulloblastoma subtypes

538 | NATURE | VOL 510 | 26 JUNE 2014

2014

Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing

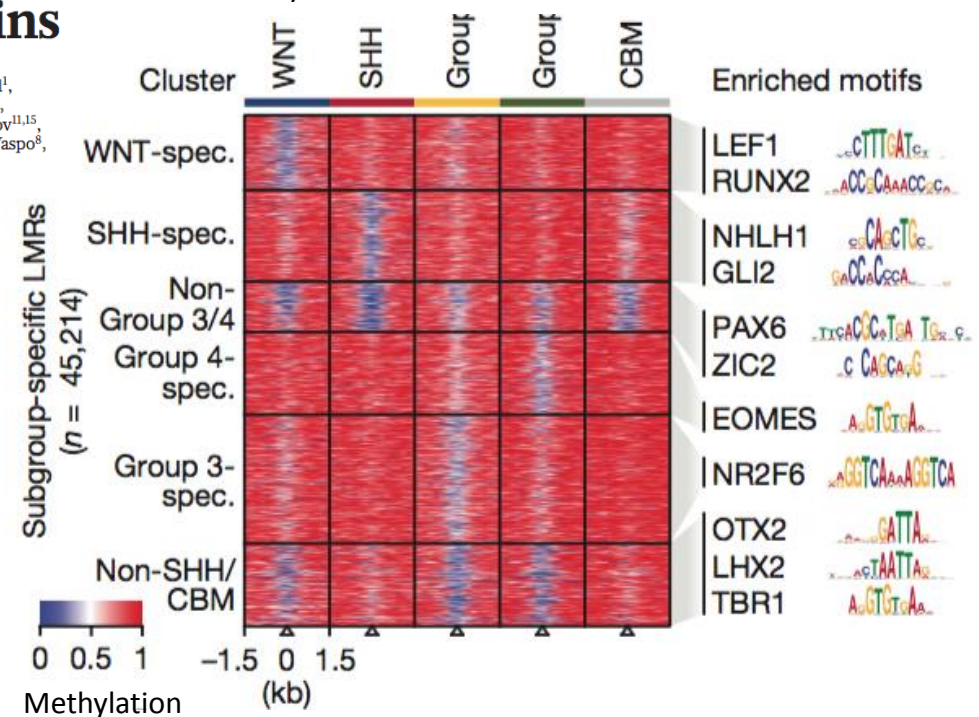
Volker Hovestadt^{1*}, David T. W. Jones^{2*}, Simone Picelli¹, Wei Wang¹, Marcel Kool², Paul A. Northcott², Marc Sultan³, Katharina Stachurski⁴, Marina Ryzhova⁵, Hans-Jörg Wernatz³, Meryem Ralser³, Sonja Brun⁶, Jens Bunt^{7,8}, Natalie Jäger⁹, Kortine Kleinheinz^{1,9}, Serap Erkek^{2,10}, Ursula D. Weber¹, Cynthia C. Bartholomae^{11,12}, Christof von Kalle^{11,12}, Chris Lawrenz¹³, Jürgen Eils¹³, Jan Koster⁸, Rogier Versteeg⁸, Till Milde^{14,15}, Olaf Witt^{14,15}, Sabine Schmidt¹⁶, Stephan Wolf¹⁶, Torsten Pietsch¹⁷, Stefan Rutkowski¹⁸, Wolfram Scheurle¹⁹, Michael D. Taylor^{20,21,22}, Benedikt Brors⁹, Jörg Felsberg^{23,24}, Guido Reifenberger^{23,24}, Arndt Borkhardt⁴, Hans Lehrach³, Robert J. Wechsler-Reya⁶, Roland Eils^{9,25,26,27}, Marie-Laure Yaspo³, Pablo Landgraf⁴, Andrey Korshunov^{28,29}, Marc Zapatka¹, Bernhard Radlwimmer¹, Stefan M. Pfister^{2,14} & Peter Lichter^{1,27}

Active medulloblastoma enhancers reveal subgroup-specific cellular origins

2016

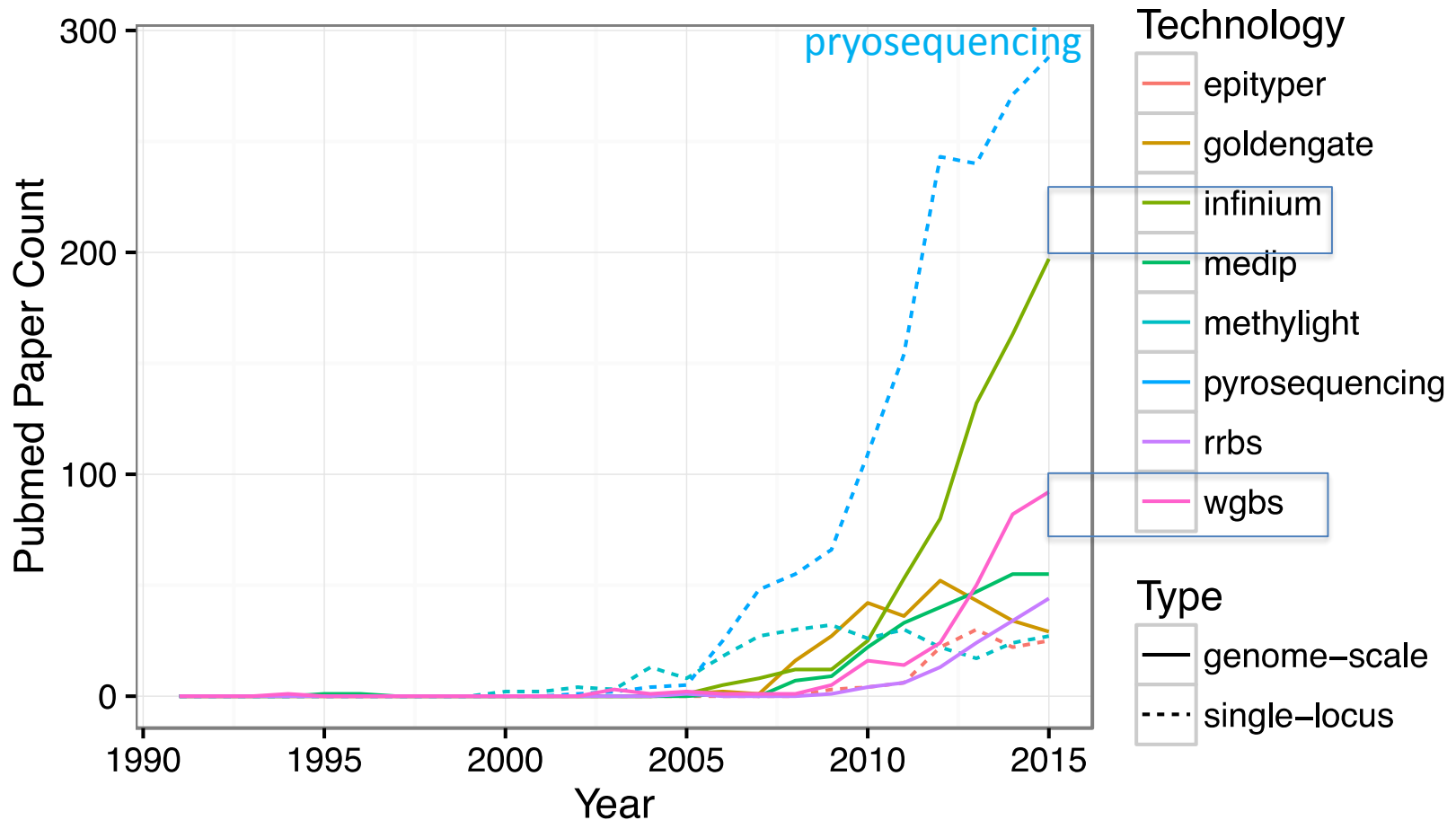
Charles Y. Lin^{1†*}, Serap Erkek^{2,3*}, Yilai Tong⁴, Linlin Yin⁵, Alexander J. Federation¹, Marc Zapatka⁶, Parthiv Haldipur⁷, Daisuke Kawauchi³, Thomas Risch⁸, Hans-Jörg Wernatz⁸, Barbara C. Worst³, Bensheng Ju⁹, Brent A. Orr¹⁰, Rhamy Zeid¹, Donald R. Polaski¹, Maia Segura-Wang², Sebastian M. Waszak², David T. W. Jones^{3,11}, Marcel Kool^{3,11}, Volker Hovestadt⁶, Ivo Buchhalter¹², Laura Sieber³, Pascal Johann³, Lukas Chavez², Stefan Gröschel¹³, Marina Ryzhova¹⁴, Andrey Korshunov^{11,15}, Wenbiao Chen⁵, Victor V. Chizhikov¹⁶, Kathleen J. Millen^{7,17}, Vyacheslav Amstislavskiy⁸, Hans Lehrach⁸, Marie-Laure Yaspo³, Roland Eils^{12,18}, Peter Lichter^{6,11}, Jan O. Korb², Stefan M. Pfister^{3,11,19}§, James E. Bradner¹§ & Paul A. Northcott^{3,4}§

Subtype-specific enhancers identified by WGBS of 34 MDB tumors



German Cancer Center (DKFZ)

Infinium array and WGBS are the most popular genome-scale methylation assays



Slide courtesy Hui Shen (Van Andel Research Institute)

Inferring altered transcription factor networks in cancer with expression + methylation

https://bioconductor.org/packages/release/bioc/html/ELMER.html



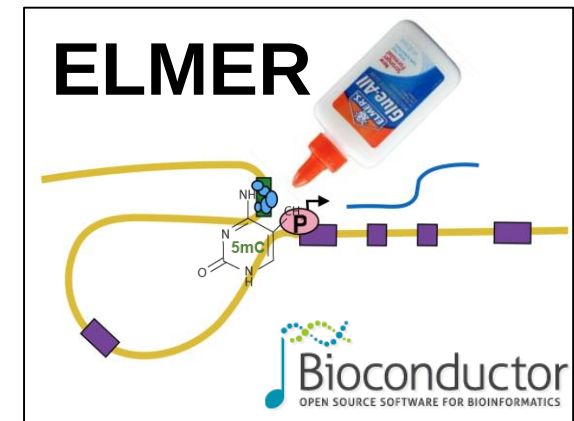
Home Install Help

Home » Bioconductor 3.3 » Software Packages » ELMER

ELMER

platforms all downloads top 20% posts 1 / 1 / 0 / 0 in Bioc 0.5 years

build error commits 1.83 test coverage unknown



Inferring Regulatory Element Networks Using Cancer

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*}

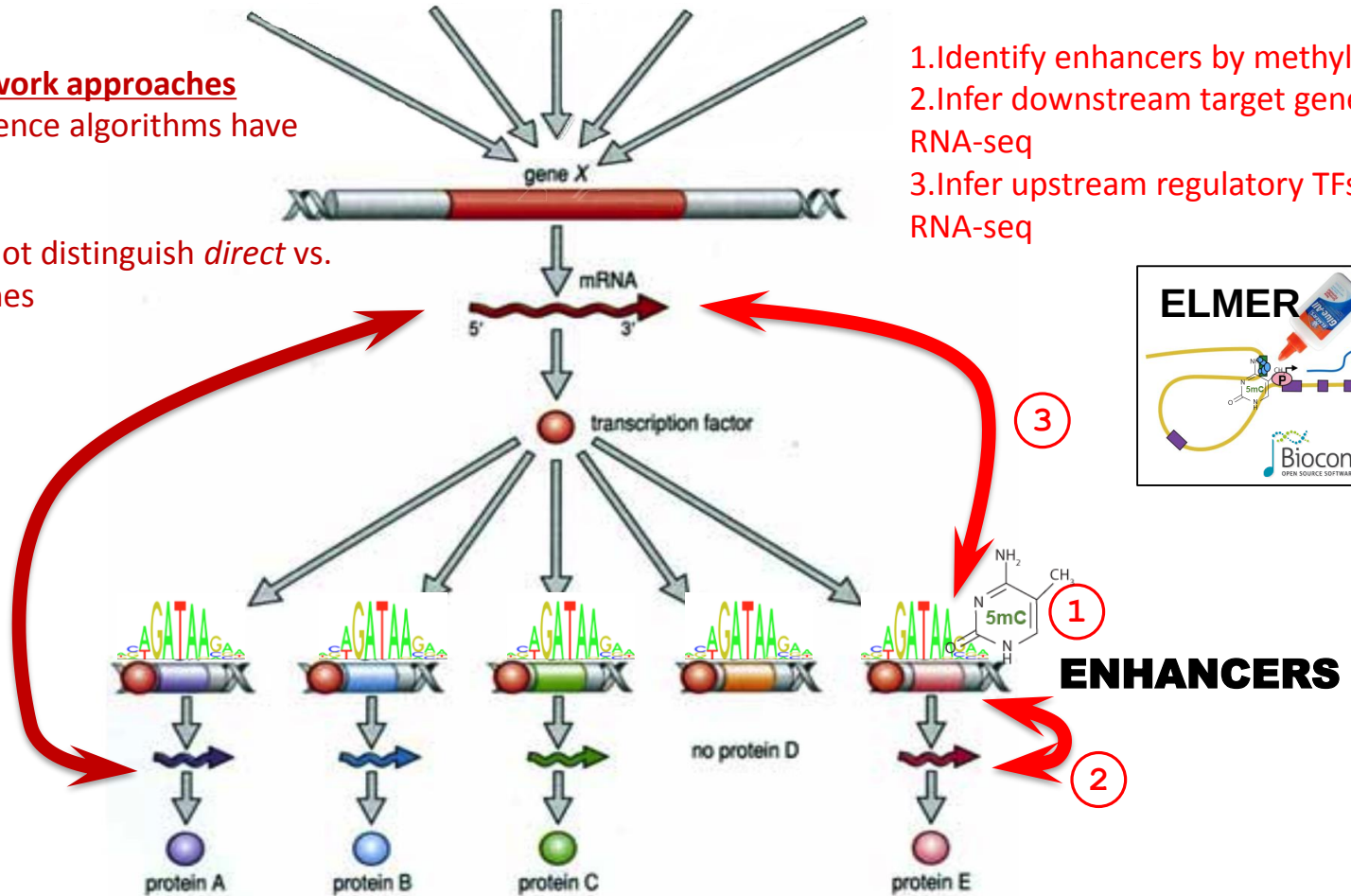
Inferring altered transcription factor networks in cancer with expression + methylation

Oncogenic cues or signaling pathways

Expression only network approaches

-Many network inference algorithms have been developed

-Problem: They can not distinguish *direct* vs. *indirect* TF target genes

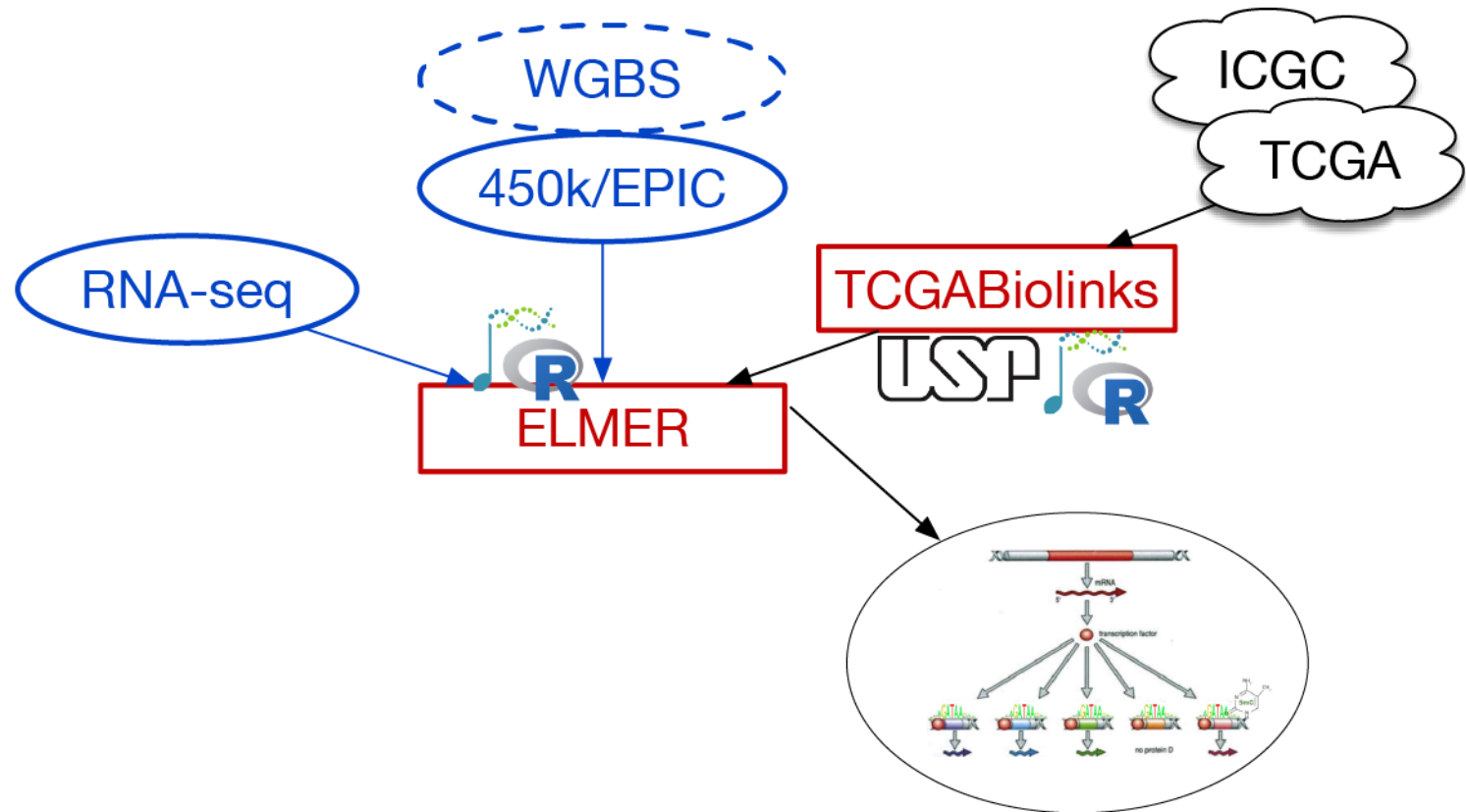


Modified from ScienceBlogs

Lijing Yao, Peggy Farnham, Ben Berman

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

BioConductor toolkit for DNA methylation analysis



TCGAbiolinks to import TCGA datasets

OXFORD JOURNALS

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURR

Institution: FMRP/USP/BIBLIOTECA CENTRAL Sign In as Personal Subscriber

Oxford Journals > Science & Mathematics > Nucleic Acids Research > Advance Access > 10.109

TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data

Antonio Colaprico^{1,2,*}, **Tiago C. Silva^{3,4,*}**, Catharina Olsen^{1,2}, Luciano Garofano^{5,6}, Claudia Cava⁷, Davide Garolini⁸, Thais S. Sabedot^{3,4}, Tathiane M. Malta^{3,4}, Stefano M. Pagnotta^{5,9}, Isabella Castiglioni⁷, Michele Ceccarelli¹⁰, Gianluca Bontempi^{1,2,*} and Houtan Noushmehr^{3,4,*}

+ Author Affiliations

*To whom correspondence should be addressed. Tel: +1 310 570 2362; Fax: +55 16 3315 0222; Email: houtan@usp.br

Correspondence may also be addressed to Gianluca Bontempi. Tel: +32 2 650 55 91; Fax: +32 2 650 56 09; Email: gbonte@ulb.ac.be

†These authors contributed equally to the paper as first authors.

Received October 22, 2015.
Revision received December 6, 2015.
Accepted December 10, 2015.



Universidade de São Paulo
Brasil



<http://bioconductor.org/packages/TCGAbiolinks/>

TCGAbiolinks

platforms all downloads top 5% posts 13 / 0.9 / 3 / 2 In Bloc 0.5 years
build ok commits 9.17 test coverage unknown

This is the **development** version of TCGAbiolinks; for the stable release version, see [TCGAbiolinks](#).

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

Bioconductor version: Development (3.4)

The aim of TCGAbiolinks is : i) facilitate the TCGA open-access data retrieval, ii) prepare the data using the appropriate pre-processing strategies, iii) provide the means to carry out different standard analyses and iv) allow the user to download a specific version of the data and thus to easily reproduce earlier research results. In more detail, the package provides multiple methods for analysis (e.g., differential expression analysis, identifying differentially methylated regions) and methods for visualization (e.g., survival plots, volcano plots, starburst plots) in order to easily develop complete analysis pipelines.

Collaboration with Tiago Silva and Houtan Noushmehr

TCGAbiolinks provides both programatic and GUI interfaces

- TCGA Data
- Clinical analysis
- Epigenetic analysis
- Transcriptomic analysis
- Genomic analysis
- Integrative analysis
- Starburst plot
- ELMER

Results table

Plots

chr14:94849579-96505661

1.66Mb

DICER1

CLMN

cg15862394

Create mee +

Data -
Select mee
Select results

Analysis +

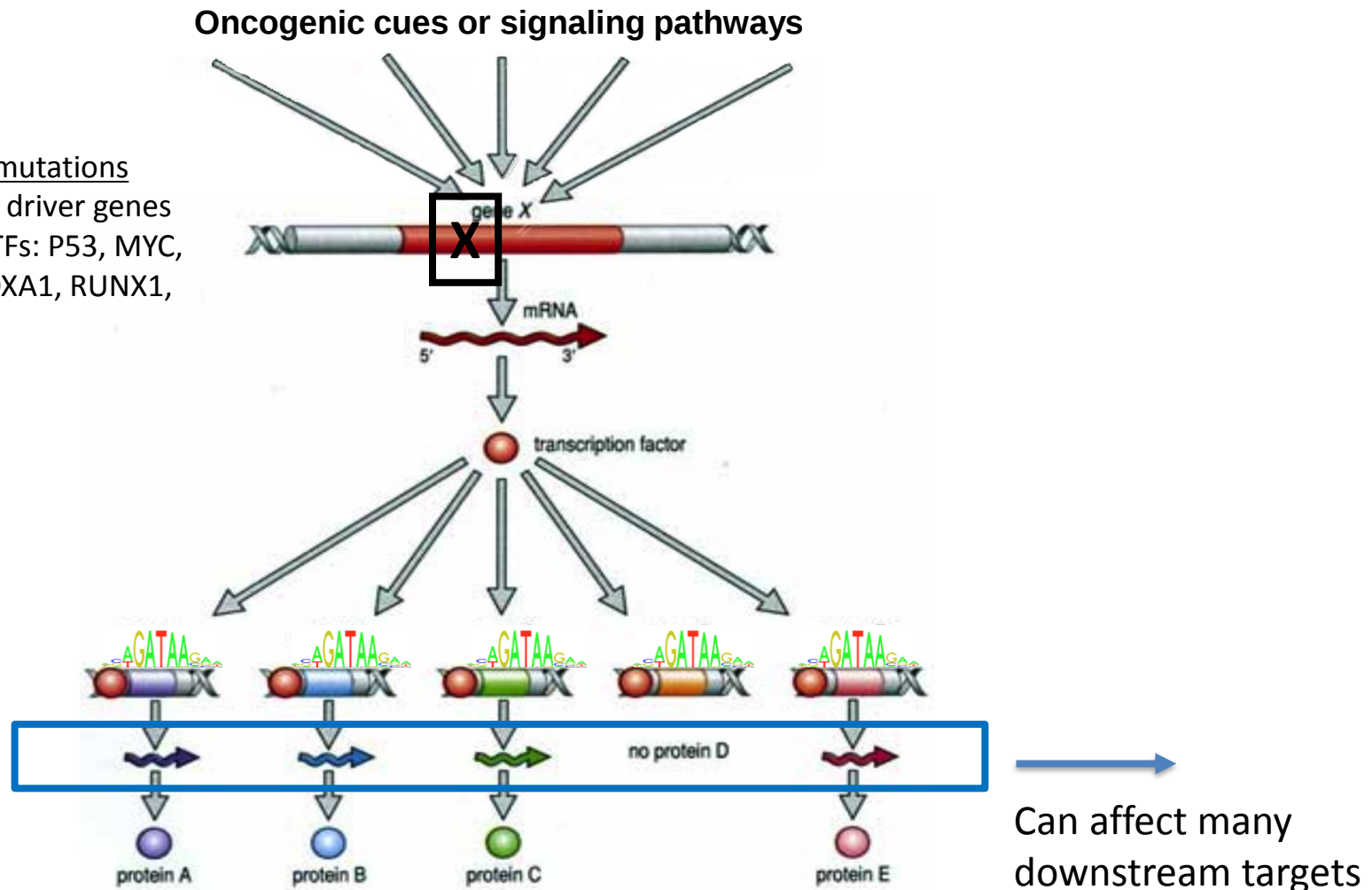
Plots -
Type selection -
Type of plot:
Schematic Plot
Schematic plot by:
Genes
Probes
Probes:
cg15862394
Size options +
Plot
Results table +

Tiago Silva

Understanding how genetic alterations affect transcriptional networks

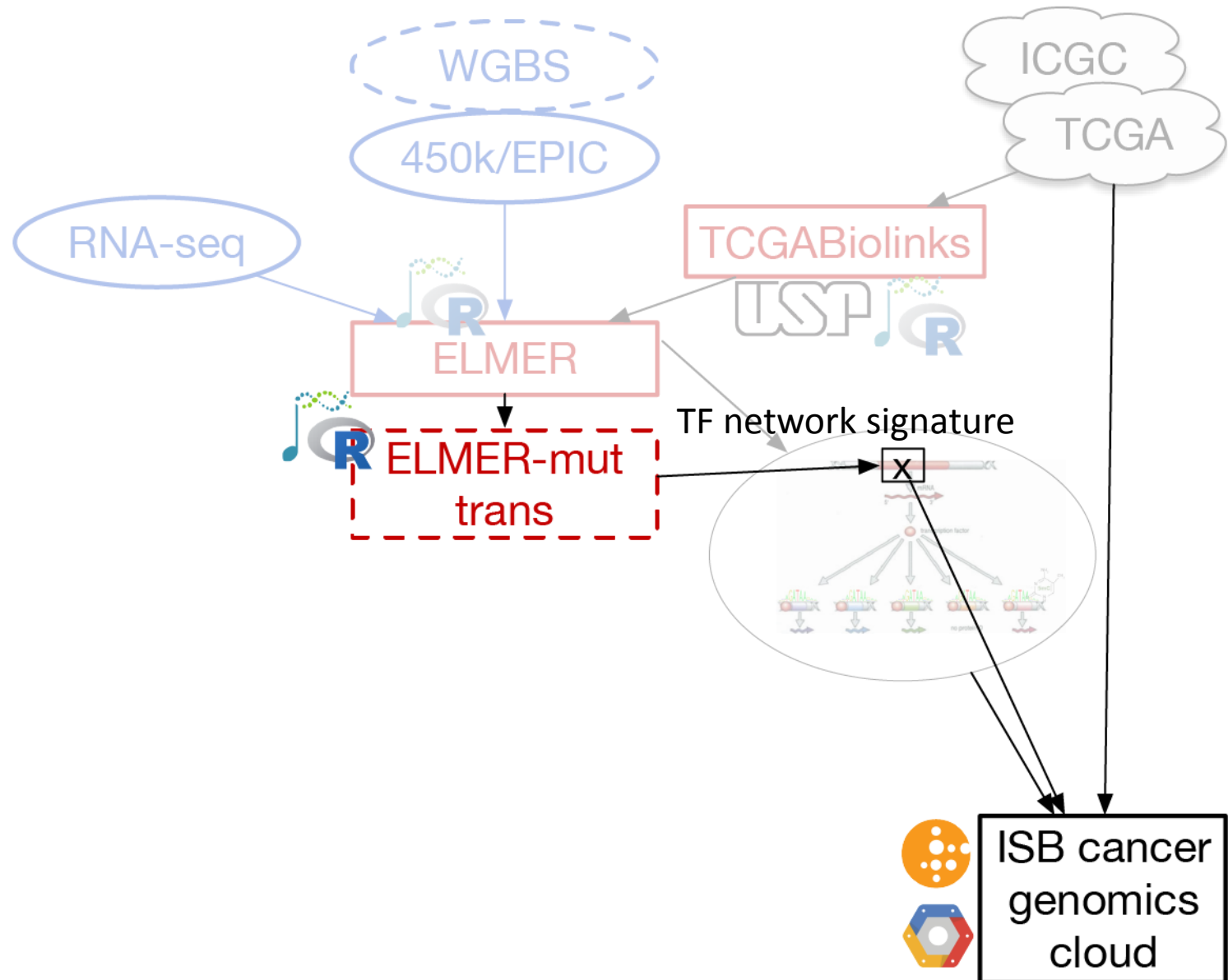
“Trans” cancer mutations

-~25% of cancer driver genes from TCGA are TFs: P53, MYC, WT1, GATA3, FOXA1, RUNX1, NFE2L2, etc.

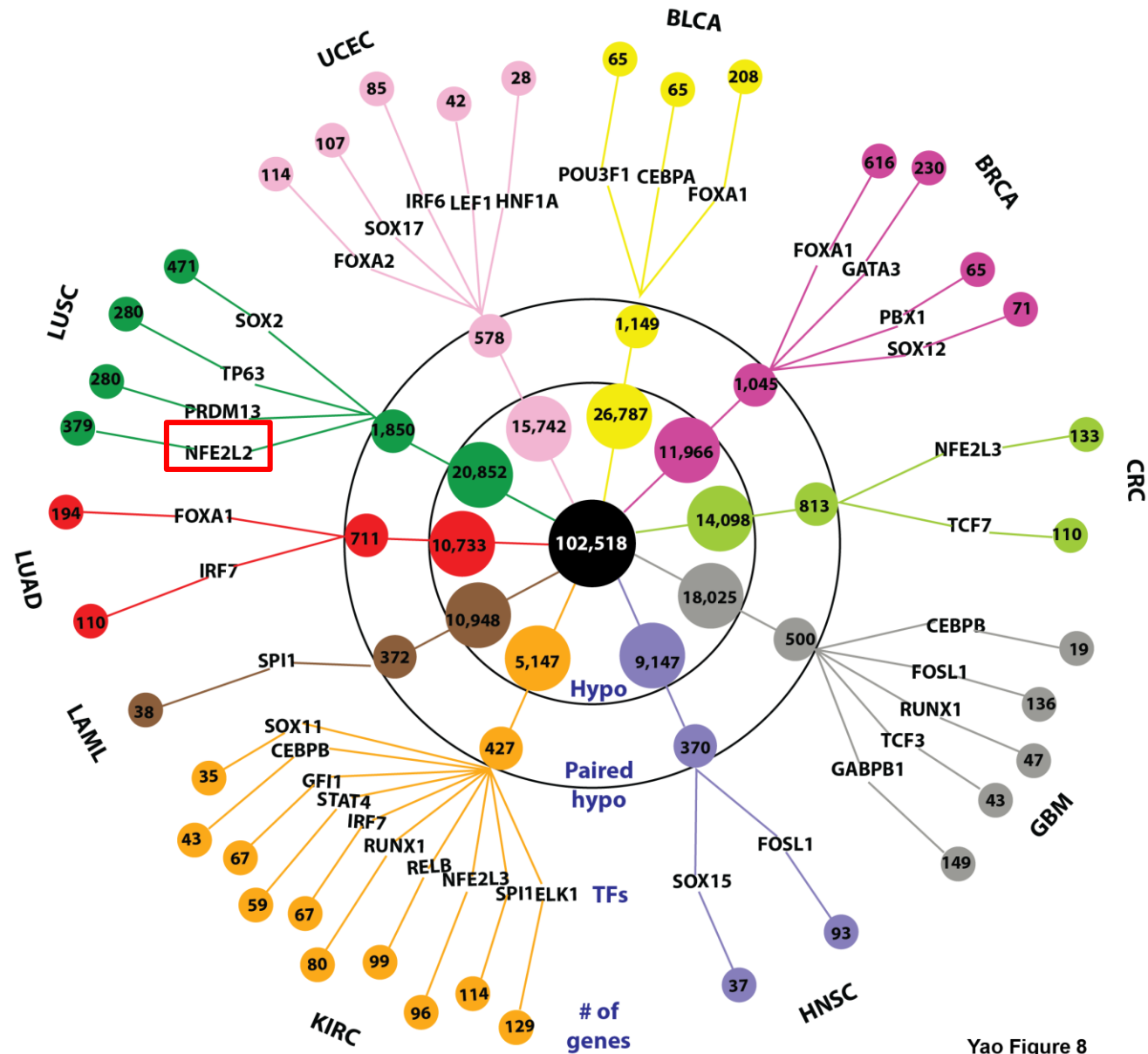


Modified from ScienceBlogs

Effects of *trans* TF mutations on cancer networks



Example: Lung squamous (LUSC) mutations in NFE2L2

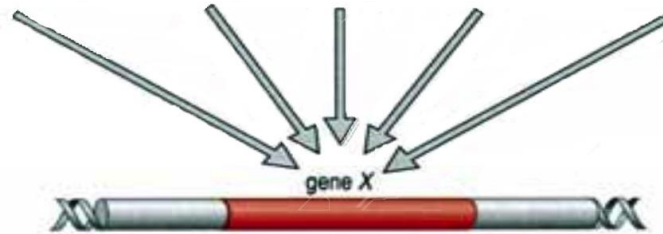


Yao Figure 8

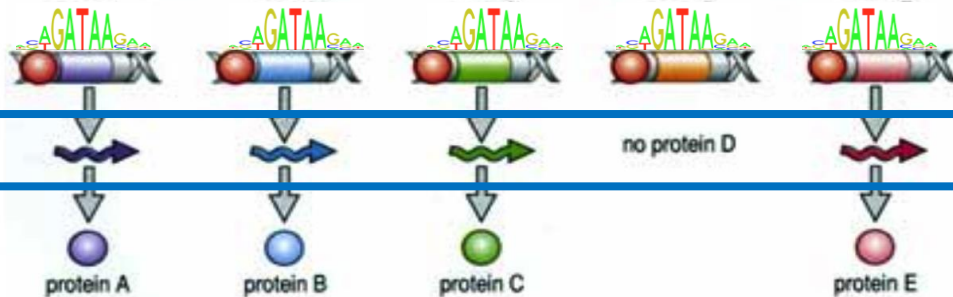
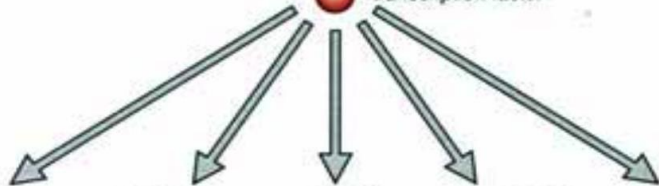
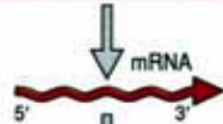
Lijing Yao

NFE2L2 network signatures in NFE2L2 pathway mutant cases

Oncogenic cues or signaling pathways



Trans TF mutation



Modified from ScienceBlogs

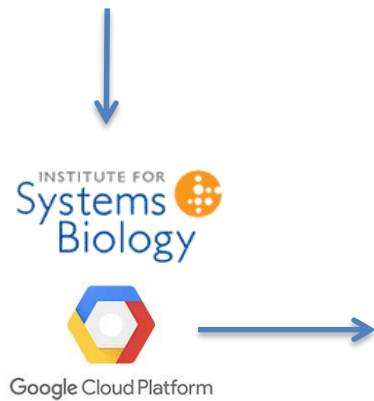
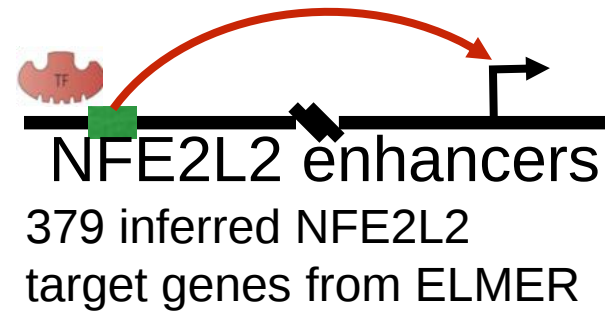
ISB Cancer Genomic Cloud

INSTITUTE FOR
Systems
Biology

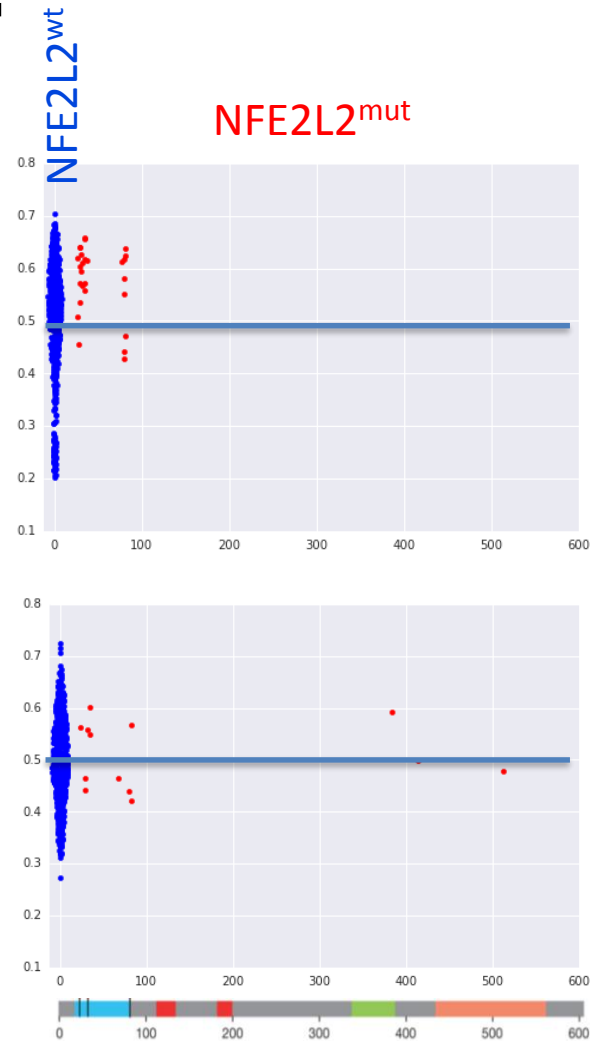
Google Cloud Platform

Expression signature of
ELMER-predicted targets

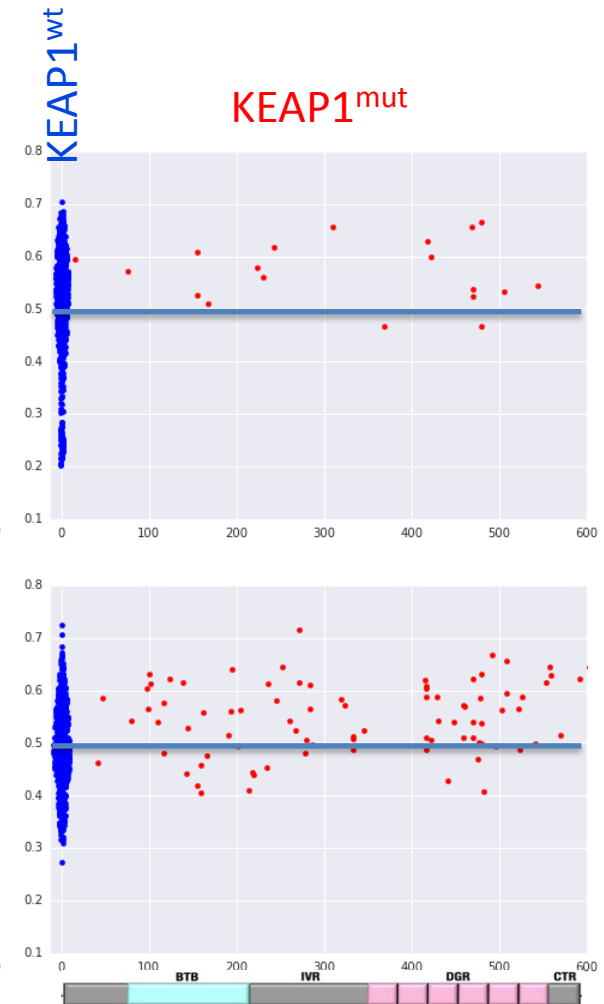
NFE2L2 network signatures in NFE2L2 pathway mutant cases



ELMER NFE2L2 expression signature



NFE2L2^{mut} AA position



KEAP1^{mut} AA position

538
LUSC
cases

510
LUAD
cases

Nicole Yeager and Lijing Yao

Cis alterations to transcriptional networks

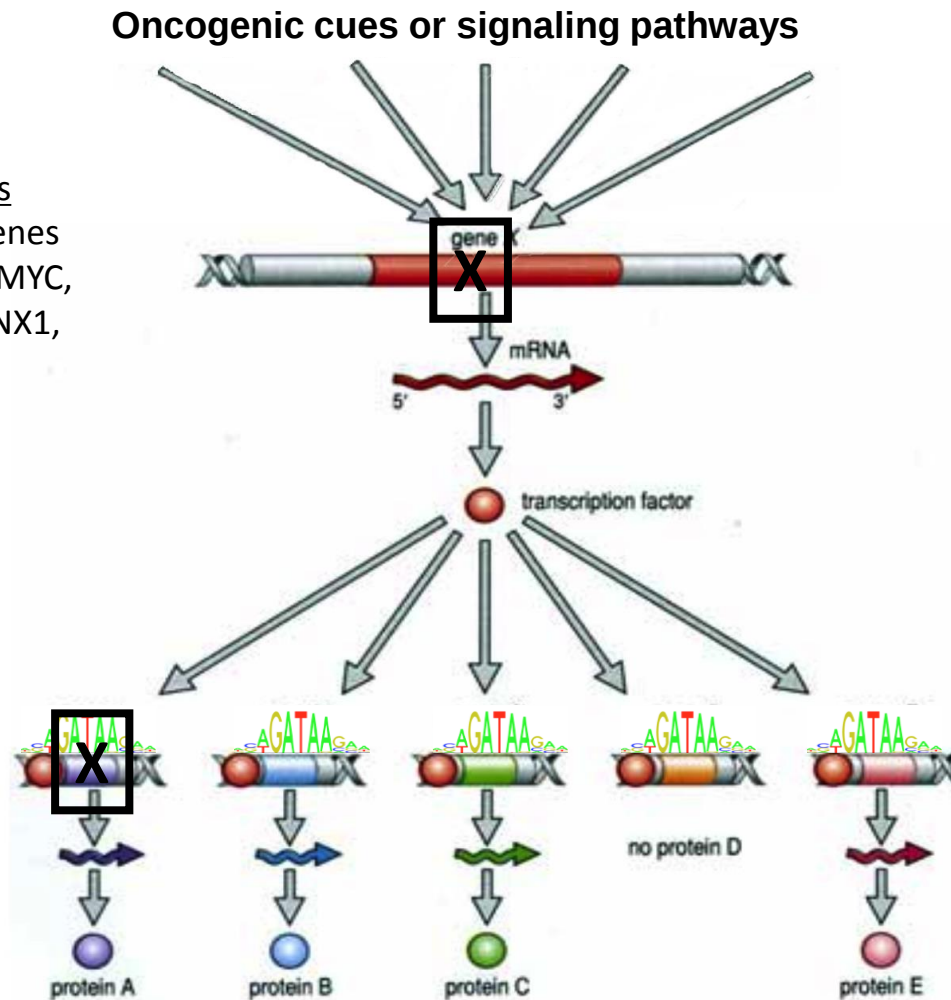
“Trans” cancer mutations

-~25% of cancer driver genes from TCGA are TFs: P53, MYC, WT1, GATA3, FOXA1, RUNX1, NFE2L2, etc.

“cis” cancer regulatory mutations

-Cancer GWAS hits fall in enhancer elements (MYC enhancer in CRC)

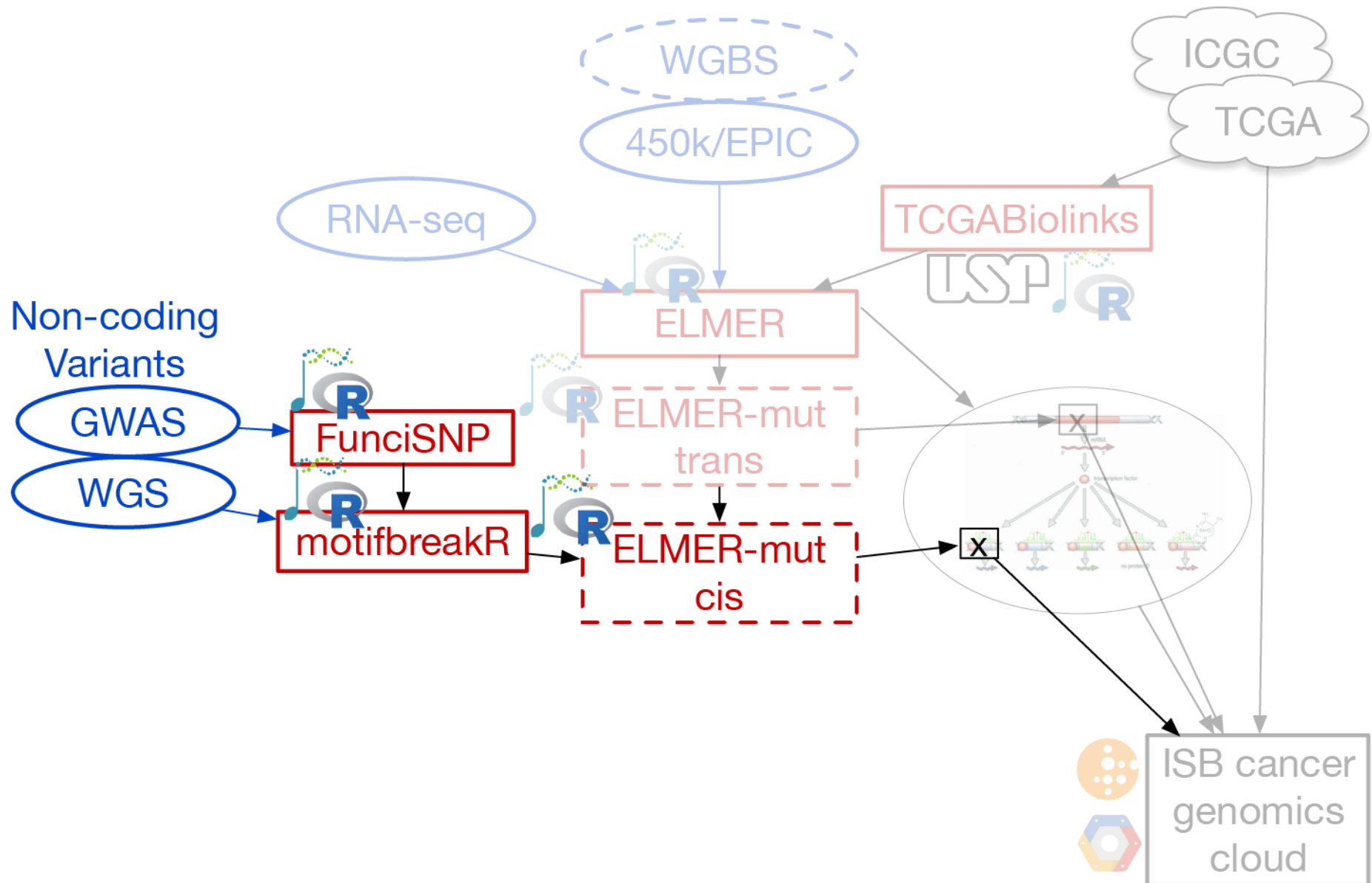
-Somatic regulatory mutations (TERT promoter in melanoma, PAX5 enhancer in CLL)



**PROMOTERS/
ENHANCERS**

Modified from ScienceBlogs

Transcription factor binding site analysis of *cis* mutations



FunciSNP to analyze GWAS *cis* variants

Published online 8 June 2012

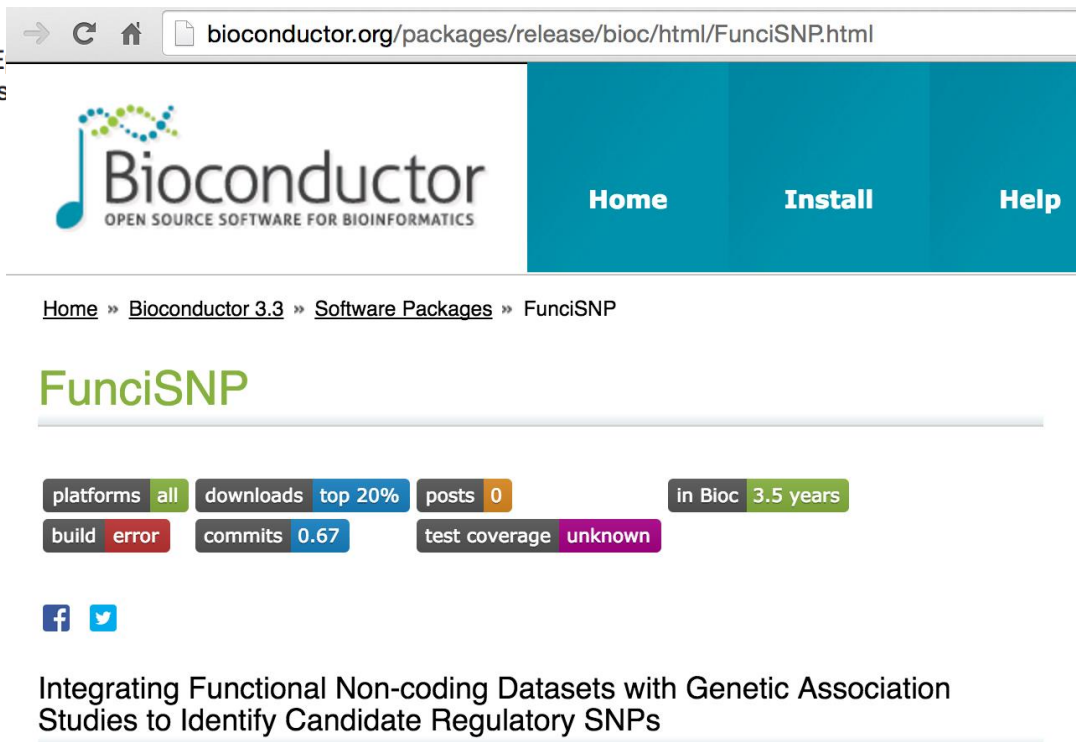
Nucleic Acids Research, 2012, Vol. 40, No. 18 e139
doi:10.1093/nar/gks542

FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs

Simon G. Coetzee^{1,2}, Suhn K. Rhie^{1,2}, Benjamin P. Berman^{1,2,3}, Gerhard A. Coetzee^{1,2,4,*}
and Houtan Noushmehr^{1,2,3,*}

¹Norris Cancer Center, ²Department of Preventive Medicine, ³Keck School of Medicine, University of Southern California, Los

Received March 20, 2012; Revised April 26, 2012; Accepted May 13, 2012



The screenshot shows the Bioconductor website interface. At the top, the URL is bioconductor.org/packages/release/bioc/html/FunciSNP.html. The Bioconductor logo is displayed, along with navigation links for Home, Install, and Help. Below the logo, the breadcrumb trail reads: Home » Bioconductor 3.3 » Software Packages » FunciSNP. The package name 'FunciSNP' is prominently displayed in green. A series of status badges are shown: 'platforms all' (green), 'downloads top 20%' (blue), 'posts 0' (orange), 'in Bioc 3.5 years' (green), 'build error' (red), 'commits 0.67' (blue), and 'test coverage unknown' (purple). Social media icons for Facebook and Twitter are present. The package description at the bottom reads: 'Integrating Functional Non-coding Datasets with Genetic Association Studies to Identify Candidate Regulatory SNPs'.

Simon Coetzee

motifbreakR to predict effects of TF binding site variants

→ ↺ 🏠 <https://www.bioconductor.org/packages/devel/bioc/html/motifbreakR.html>



Home

Install

Help



[Home](#) » [Bioconductor 3.4](#) » [Software Packages](#) » motifbreakR (development version)

motifbreakR

platforms **all** downloads **top 50%** **po**
build **ok** commits **2.67** **te**

Bioinformatics, 31(23), 2015, 3847–3849

doi: 10.1093/bioinformatics/btv470

Advance Access Publication Date: 12 August 2015

Applications Note

OXFORD

Sequence analysis

***motifbreakR*: an R/Bioconductor package for predicting variant effects at transcription factor binding sites**

Simon G. Coetzee¹, Gerhard A. Coetzee² and Dennis J. Hazelett^{1,*}

¹Bioinformatics and Computational Biology Research Center, Cedars-Sinai Medical Center, Los Angeles, CA, USA
and ²Department of Urology and Preventive Medicine, USC Norris Comprehensive Cancer Center, Los Angeles, CA, USA

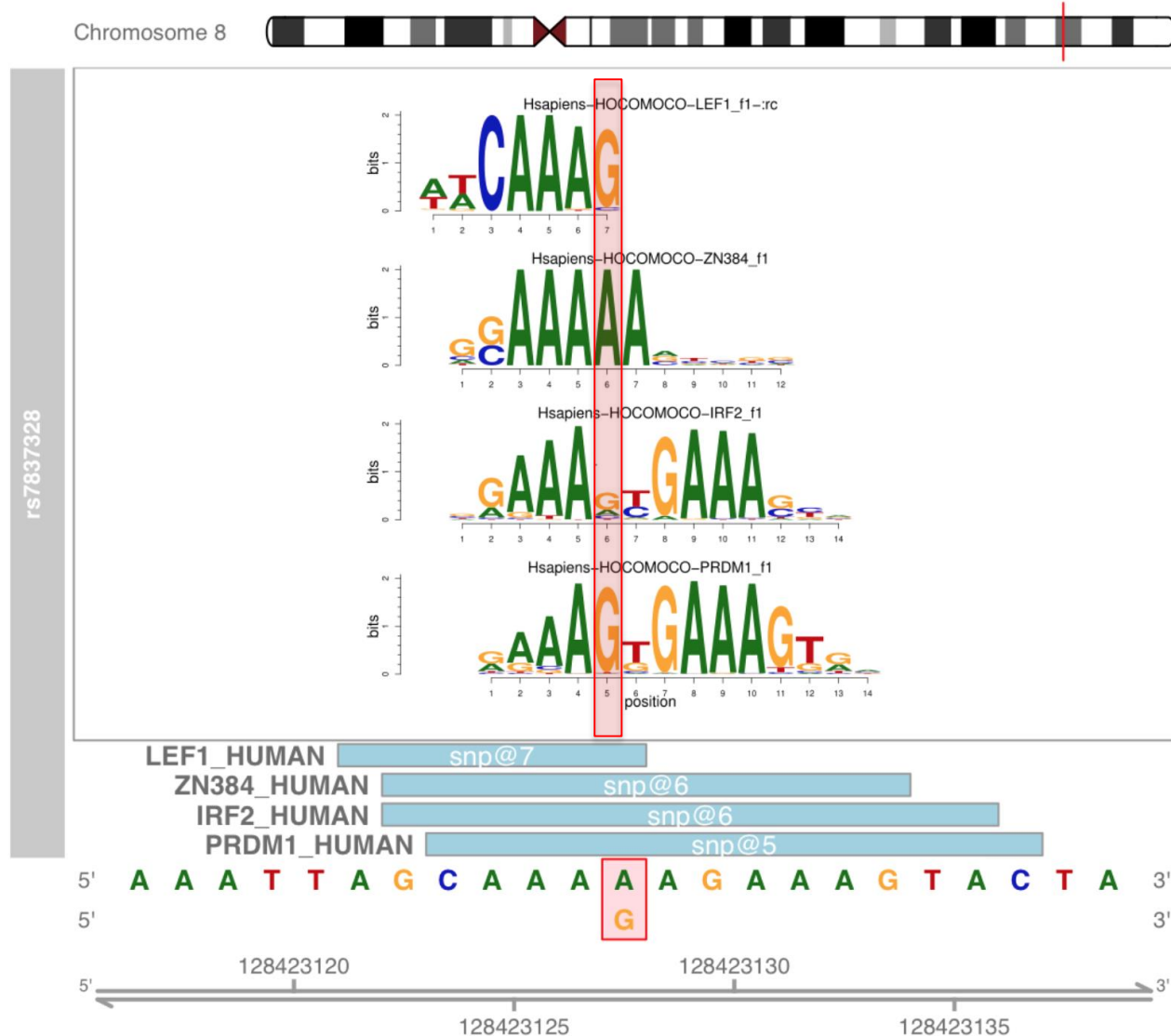
*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on May 12, 2015; revised on July 21, 2015; accepted on August 6, 2015

Simon Coetzee and Dennis Hazelett

motifbreakR output



Simon Coetzee and Dennis Hazelett

Often, DNA methylation changes are presented as a black box

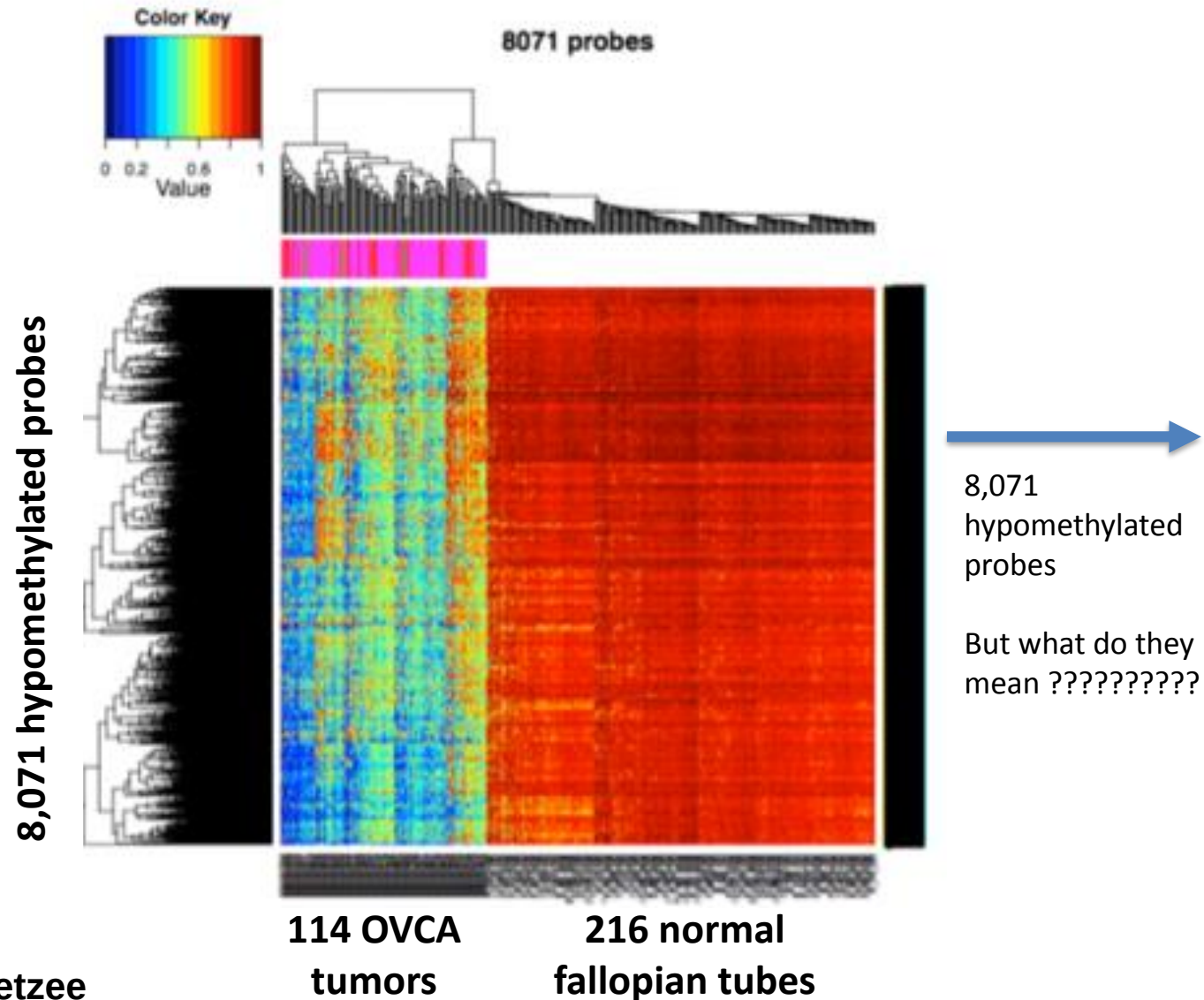
Data From

[Nature](#). 2015 May
28;521(7553):489-94. doi:
10.1038/nature14410.

“Whole-genome
characterization of
chemoresistant ovarian ca
ncer.”

[Nat Commun](#). 2016 May
24;7:11620. doi:
10.1038/ncomms11620.

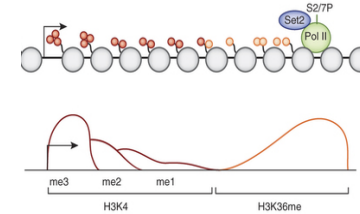
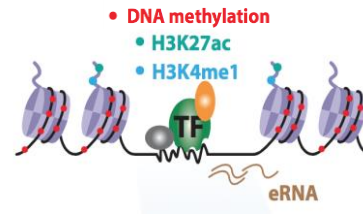
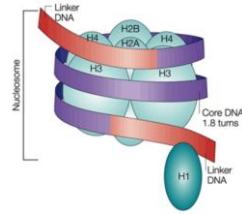
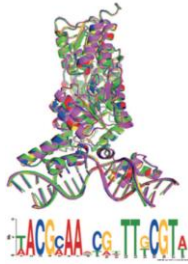
“Epigenetic
reprogramming of fallopian
tube fimbriae in BRCA
mutation carriers defines
early ovarian cancer
evolution.”



Huy Dinh and Simon Coetzee

FUNCTIONAL ANNOTATION OF DNA METHYLATION CHANGES

Using reference epigenomes to annotate methylation changes



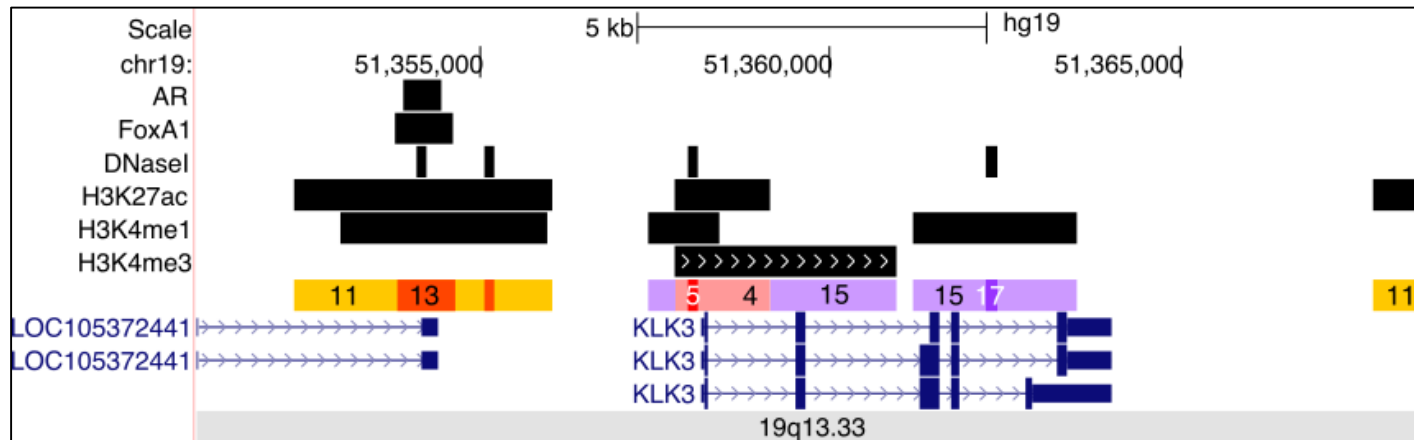
Feature	TFBS	Nucleosome positioning	Promoter, enhancer, insulator	Super-enhancer, PRC2 domain	Transcription unit	Nuclear topological domain
Genomic size	5-30bp	40-150bp	200-2,000bp	5-20kb	10-100kb	100kb-10Mb
ENCODE/Roadmap methods	ChIP-seq, DNase-seq	Mnase-seq, (ATAC-seq)	ChIP-seq, ATAC-seq, DNase-seq	ChIP-seq	ChIP-seq	Hi-C ChIA-PET
Has DNA methylation signature?	No?	Yes (2012)	Yes Promoters 1981 Enhancers 2011	Yes 2010-2013	No? (2016)	Yes (2011)

StateHub: A standard exchange format for combinatorial chromatin states

StateHub

- Rule-based states
- User-submitted states
- Fully versioned for reproducible and comparable results

TRS	100	100	000	100	000	111	000	100	Transcribed
HET	100	100	111	100	100	100	100	100	Heterochromatin
SCR	100	100	000	100	111	100	100	100	Silenced Chromatin
PAR	000	111	000	111	000	000	100	100	Promoter Active Region
PARC	000	111	000	111	000	000	111	100	Promoter Active Region Core
PPR	000	111	000	110	000	000	100	100	Promoter Poised Region
PPRC	000	111	000	110	000	000	111	100	Promoter Poised Region Core
EAR	101	110	000	111	000	000	100	100	Enhancer Active Region
EARC	101	110	000	111	000	000	111	100	Enhancer Active Region Core
EPR	111	110	000	110	000	000	100	100	Enhancer Poised Region
EPRC	111	110	000	110	000	000	111	100	Enhancer Poised Region Core
ER	111	110	000	101	000	000	100	100	Enhancer Region
ERC	111	110	000	101	000	000	111	100	Enhancer Region Core
PR	000	111	000	100	000	000	100	100	Promoter Region
PRC	000	111	000	100	000	000	111	100	Promoter Region Core
CTCF	000	000	000	000	000	000	100	111	CTCF Bound Region
CTCFC	000	000	000	000	000	000	111	111	CTCF Bound Region Core
RPS	100	100	000	100	000	000	111	100	Regulatory Putative Site
AR	100	100	000	111	000	000	100	100	Active Region
ARC	100	100	000	111	000	000	111	100	Active Region Core



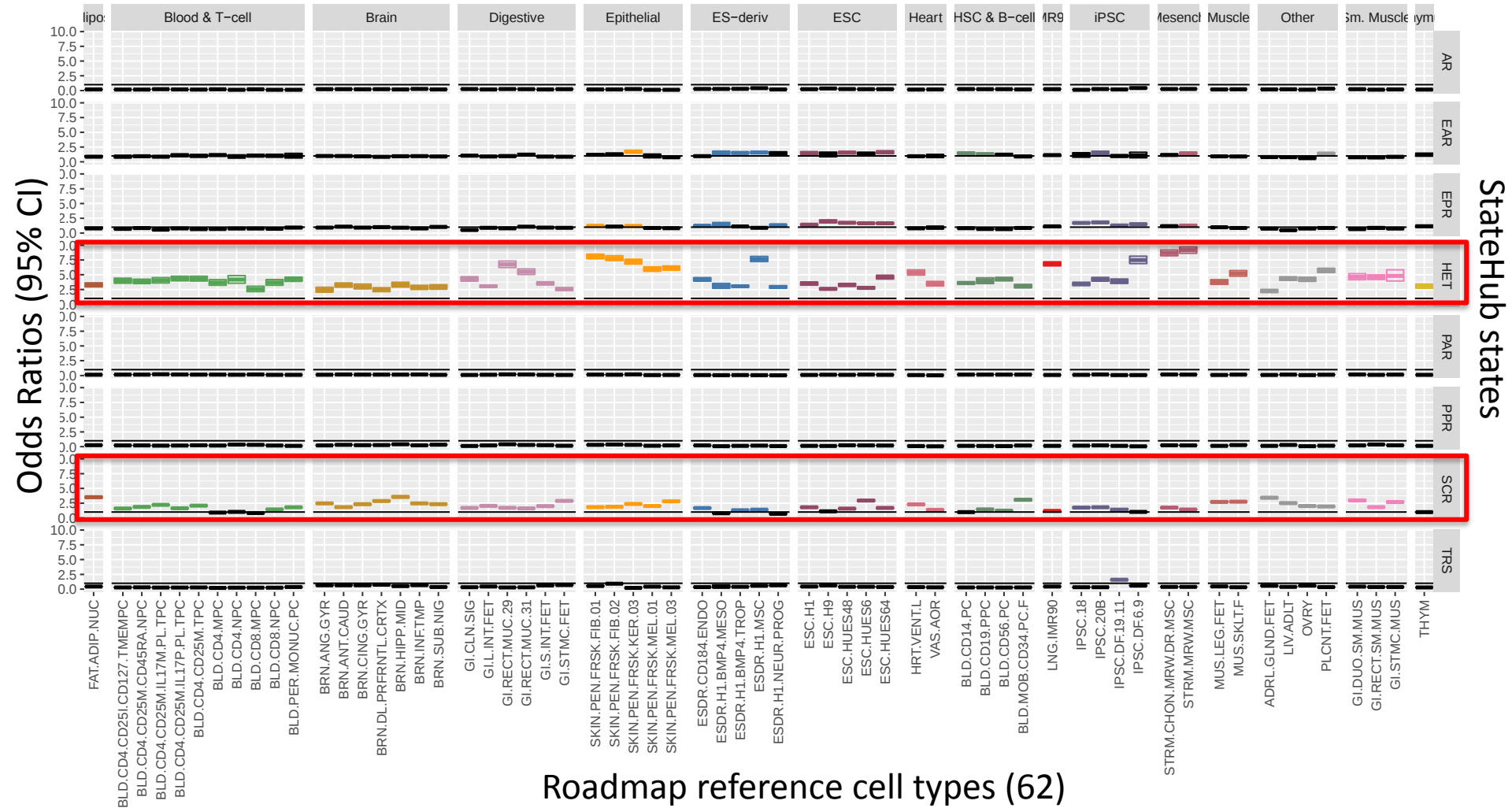
<https://github.com/Junkdnalab/StateHub>

<https://github.com/Junkdnalab/StatePaintR>

Simon Coetzee and Dennis Hazelett

FUNCTIONAL ANNOTATION OF DNA METHYLATION CHANGES

StatePaintR results for 8,071 hypo probes



	H3K9me3			H3K27me3						
HET	100	100	111	100	100	100	100	100	Heterochromatin	H3K9me3 broad peaks
SCR	100	100	000	100	111	100	100	100	Silenced Chromatin	H3K27me3 broad peaks

FUNCTIONAL ANNOTATION OF DNA METHYLATION CHANGES

Summary

- DNA methylation can identify cancer changes in transcriptional regulatory elements and regulatory networks (ELMER)
- Next step – tool integration to investigate *trans* and *cis* mutations affects on TF networks
- StateHub: A standard exchange format for combinatorial chromatin states, for annotating function of methylation changes
- All R/BioConductor links available at <https://www.bioconductor.org/>
- BioConductor, GitHub & Docker links available <http://junkdnalab.org>