1.- Personalized Medicine
2.- Epigenetics and Networks

Alfonso Valencia
Spanish National Cancer Research Centre
CNIO
valencia@cnio.es
@alfons_valencia

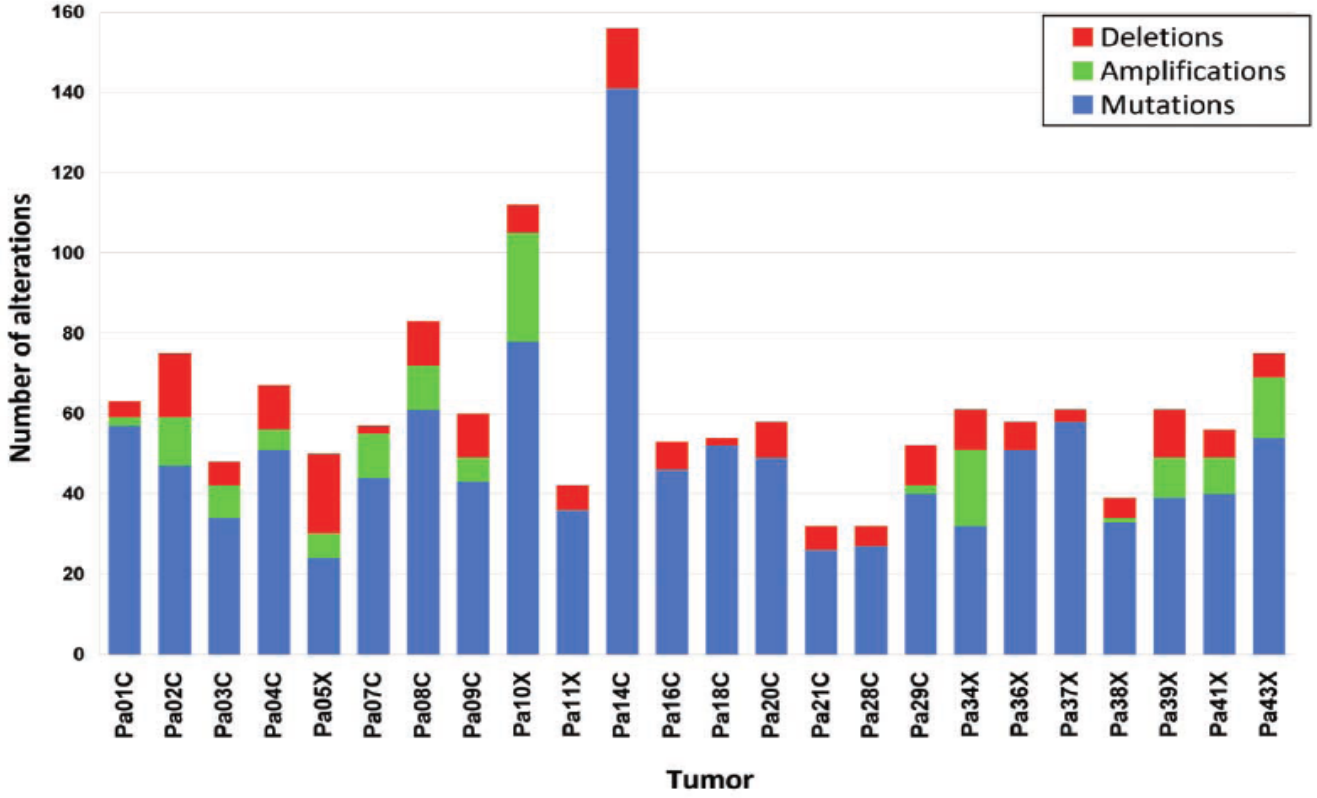
ESHG Satellite Symposium
"Machine Learning for Personalized Medicine"
Barcelona, May 2016

Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses

Siân Jones,[1*] Xiaosong Zhang,[1*] D. Williams Parsons,[1,2*] Jimmy Cheng-Ho Lin,[1*] Rebecca J. Leary,[1*] Philipp Angenendt,[1*] Parminder Mankoo,[3] Hannah Carter,[3] Hirohiko Kamiyama,[4] Antonio Jimeno,[1] Seung-Mo Hong,[4] Baojin Fu,[4] Ming-Tseh Lin,[4] Eric S. Calhoun,[1] Mihoko Kamiyama,[4] Kimberly Walter,[4] Tatiana Nikolskaya,[5] Yuri Nikolsky,[6] James Hartigan,[7] Douglas R. Smith,[7] Manuel Hidalgo,[1] Steven D. Leach,[1,8] Alison P. Klein,[1,4] Elizabeth M. Jaffee,[1,4] Michael Goggins,[1,4] Anirban Maitra,[1,4] Christine Iacobuzio-Donahue,[1,4] James R. Eshleman,[1,4] Scott E. Kern,[1,4] Ralph H. Hruban,[1,4] Rachel Karchin,[3] Nickolas Papadopoulos,[1] Giovanni Parmigiani,[1,9] Bert Vogelstein,[1] Victor E. Velculescu,[1]† Kenneth W. Kinzler[1]†

SCIENCE VOL 321 26 SEPTEMBER 2008

Manuel Hidalgo,
Director
Clinical Programme
CNIO



PALB2
Identified as breast cancer susceptibility gene whose protein is closely associated with BRCA2 and is essential for BRCA2 anchorage to nuclea structures. [2008]

PALB2 mutations have been previously reported in patients with familial breast cancer, and the PALB2 protein is a binding partner for BRCA2. [2009]

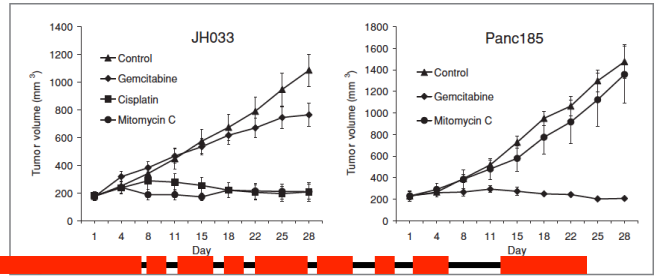
Published OnlineFirst December 6, 2010; DOI:10.1158/1535-7163.MCT-10-0893

Molecular Cancer Therapeutics

Spotlight on Clinical Response

Personalizing Cancer Treatment in the Age of Global Genomic Analyses: *PALB2* Gene Mutations and the Response to DNA Damaging Agents in Pancreatic Cancer

Maria C. Villarroel[1,2], N.V. Rajeshkumar[1,2], Ignacio Garrido-Laguna[1,2], Ana De Jesus-Acosta[1,2], Sián Jones[1,2], Anirban Maitra[1,2], Ralph H. Hruban[1,2], James R. Eshleman[1,2], Alison Klein[1,2], Daniel Laheru[1,2], Ross Donehower[1,2], and Manuel Hidalgo[1,2]

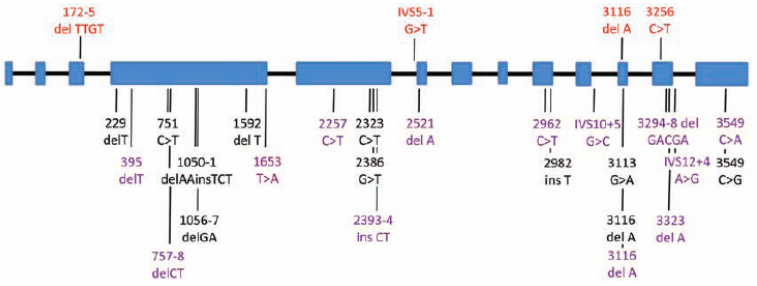


IVS10 +2 C>T Splice site

Exomic Sequencing Identifies *PALB2* as a Pancreatic Cancer Susceptibility Gene

Siân Jones,[1,2] Ralph H. Hruban,[2,3] Mihoko Kamiyama,[3] Michael Borges,[3] Xiaosong Zhang,[1,2] D. Williams Parsons,[1,2] Jimmy Cheng-Ho Lin,[1,2] Emily Palmisano,[2] Kieran Brune,[2] Elizabeth M. Jaffee,[2,3] Christine A. Iacobuzio-Donahue,[2,3] Anirban Maitra,[2,3] Giovanni Parmigiani,[2,3] Scott E Kern,[2,3] Victor E. Velculescu,[1] Kenneth W. Kinzler,[1] Bert Vogelstein,[1,2] James R. Eshleman,[2,3]* Michael Goggins,[2,3,4]* Alison P. Klein[2,3]*†

www.sciencemag.org SCIENCE VOL 324 10 APRIL 2009



EL PAÍS.com | Salud

Lunes, 21/3/2011, 08:22 h

Inicio Internacional España Deportes Economía Tecnología Cultura Gente y TV Sociedad Opinión Blogs In English buscar

Educación | Salud | Ciencia | El Viajero | El País semanal | Domingo

ELPAIS.com > Sociedad > Salud

Un cáncer de páncreas curado gracias al trasplante a un ratón

Transferir el tumor al animal permite investigar con los fármacos y dar con una solución personalizada - Un médico español lidera esta técnica

D. ALANDETE / E. DE BENITO - Washington / Madrid - 11/01/2011

Los sufridos ratones de laboratorio pueden ser los futuros catavenenos de los enfermos de cáncer. Al menos, eso es lo que está investigando uno de los recientes fichajes del Centro Nacional de Investigaciones Oncológicas (CNIO), Manuel Hidalgo. El médico, que hasta hace poco estaba en el Johns Hopkins de Baltimore (EE UU), ha conseguido con la ayuda de estos roedores lo que él cree que se podría calificar como "el primer caso de un cáncer avanzado de páncreas que se ha curado".

El sistema que se aplicó -de momento a un único paciente- es una combinación de análisis genético para saber qué fármacos pueden funcionar y de experimentación directa sobre el propio tumor. "Tuvimos suerte y acertamos, porque no había margen de error".

Los ratones desempeñaron un papel relevante: se le trasplantó el tumor que se había quitado al paciente, y así se pudo ensayar en él qué fármaco funcionaba, sin que tuviera que exponerse a

Importación de cerebros

Adoptar hábitos de vida saludables ayudaría a evitar el 40% de los cánceres

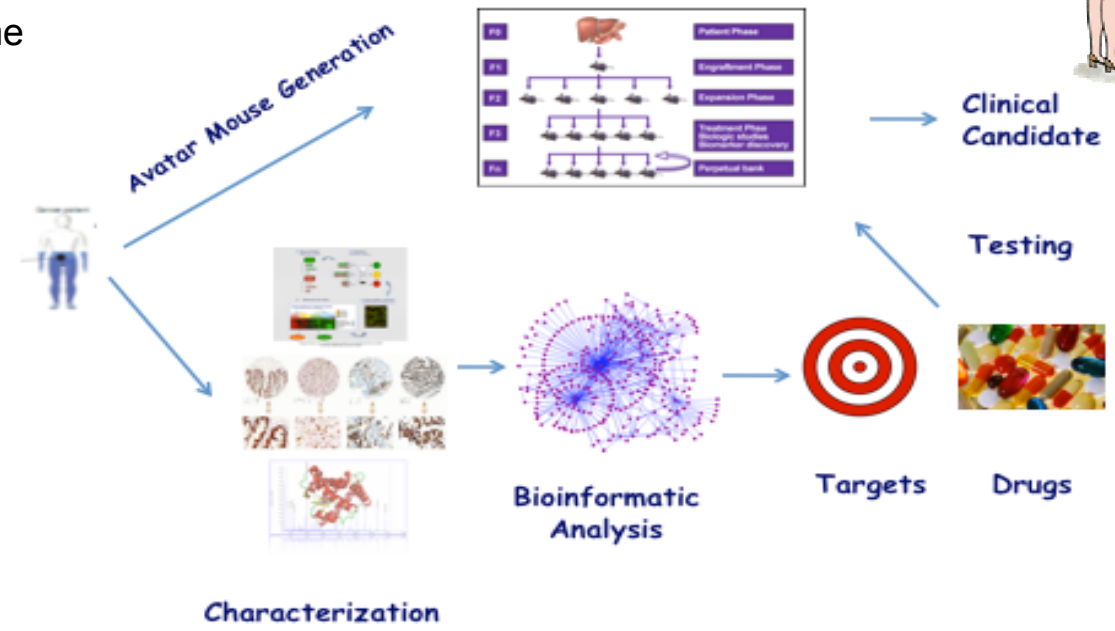
La noticia en otros webs
- webs en español
- en otros idiomas

Mark Gregoire (en el centro) rodeado de su familia después de superar el cáncer de páncreas.-

CNIO co-clinical cancer initiative

Manuel Hidalgo,
Director
Clinical Programme
CNIO

Clinical Trial



Genomic alterations

N#	PATIENT ID	TUMOR TYPE	N# MUTS	RELEVANT MUTATED GENES	N# CNV	RELEVANT CNV	PUTATIVE TARGETS
1	PGDX4	Neuroendocrine tumor	5	CREB3L3, ITPR2, MYO5B	0	0	CREB3L3
2	PGDX30	Glioblastoma	63	EPHA3, NF1, PTPN11 ,FAS , CDKN2A	0	0	NF1
3	PGDX7	High grade pancreatic neuroendocrine tumor	62	ARID1A, ARID1B, JAKMIP2, JARID2 PIK3C2A, PIK3CA, SSTR2, DDR2, TP53	6	GNG11	PI3KCA, DDR2
4	PGDX11	Pancreatic Adenocarcinoma	38	KRAS, UBA1, FAM83H, SMAD4, SLC15A2, PIWIL3, SLC3A2, SLC22A17, TP53.	10	0	None
5	PGDX61	Melanoma Uveal	5	GNA11, TAOK3	0	0	GNA11
6	PGDX68	Colon cancer	71	APC, DICER 1, TP53, CHEK1, SOS1	63	0	CHEK1
7	PGDX76	Melanoma	952	BRCA1, EZH2, FGFR2, FN1, IGF1R, KDR, KRAS, MET, MPL, PRKCB, PIK3C2G, PTK2B.	0	0	FGFR2, IGF1R, PIK3C2G, MET, BRCA1
8	PGDX135	Melanoma	29	BAI3, DNAH5, MDN1, NRAS	2	SKT19	NRAS
9	PGDX368	Pancreatic Adenocarcinoma	18	SMAD4, KRAS, ERBB2IP	0	0	ERBB2IP
10	PGDX331	Pancreatic Adenocarcinoma	21	KRAS, XPC, P53	0	0	XPC
11	PGDX369	Pancreatic Adenocarcinoma	29	KRAS. P53,SMAD4	3	0	None
12	PGDX379T2	Renal Carcinoma	25	BAP1	965	ZAP70, FGFR3, NOTCH1, TERT, STK11, GNA11, ZNF668, SOCS1, IRS2	BAP1, FGFR3, NOTCH1, STK11, GNA11.
13	PGDX330	Glioblastoma	64	MLLT10, PBRM1	27	EGFR, CDKN2A, ERRFI1	EGFR
14	PGDX27	NSCLC	45	EGFR, EZH2, TPR, TP53, RHOH	0	0	EGFR
15	PGDX17	NSCLC	69	ARID4B, PIK3R6, PTPRC, TP53, TOPO2A	13	0	PTPRC, KIF5B-RET Fusion STK11 Whole Gene Deletion
16	PGDX140	NSCLC	391	BUB1B, CYLD, EPHB6, EGFR, KRAS, KEAP1, MSH6, NTRK1, NTRK3, MUTYH, RUNX1T1, TP53	12	CDKN2A	EGFR
17	PGDX24	SCLC	32	PAPPA2	Not Done	Not Done	None
18	PGDX60	NSCLC	9	NOCHT2, MML3	6	0	NOCTH2
19	PGDX3	NSCLC	38	DLK1, JAK3	Not Done	Not Done	JAK3
20	PGDX48	Duodenal cancer	127	CARD11, CHL1, FANCD2, IRS1, NRAS, SMARCA4, TP53, PIK3R1	25	0	NRAS, PIK3R1
21	PGDX39	Colon cancer	172	APC, EGFR, FN1, GRM1, TP53, TSC2	0	0	EGFR, FN1,TSC2
22	PGDX327	Colon cancer	108	APC, AXIN2, EGFR, KDM6A, PTEN, PI3KCA, TP53	32	RECQL4	EGFR, PI3KCA, PTEN
23	PGDX310	Esophageal cancer	96	PIK3R1, NF1, TP53	65	CCND1	NF1, PIK3R1

Garralda et al.

Integrated Next Generation Sequencing and Avatar Mouse Models for Personalized Cancer Treatment.

Clinical Cancer Research 2014

Case 3: High grade pancreatic neuroendocrine tumor

- 44 years old male.
- High grade neuroendocrine carcinoma with disseminated lymph node disease
- Prior treatment with gemcitabine-Oxaliplatin: PR (Partial Response).
- When Progression Disease (PD): Fresh tumor specimen from lymph node was obtained for exome sequencing and for Avatar/xenograft generation.

Exome sequencing analysis:
- 64 somatic relevant mutations
- 6 Copy number variations

Point mutation in *PIK3CA* gene: 909F>C

PI3K inhibitors and MEK inhibitors

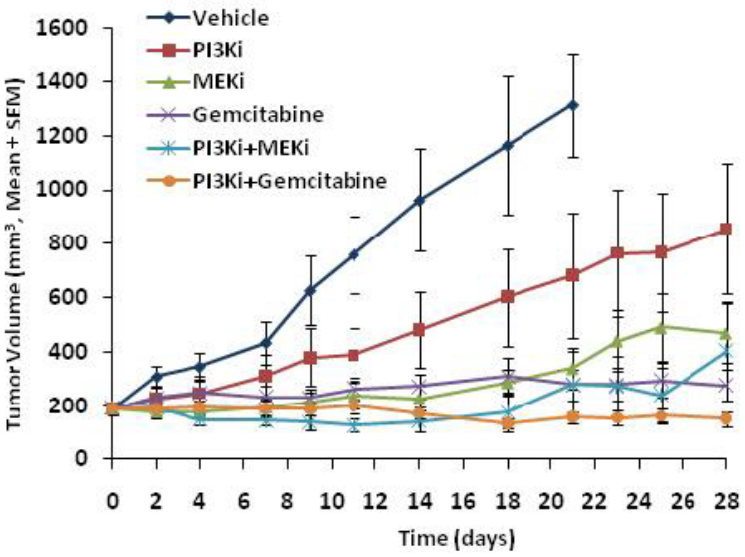
Point mutation in DDR2 protein: 381P>A

DDR2 mut - 4% of Squamous Cell Lung Cancer. DDR2 mut associated with sensitive to Dasatinib (*Hammerman et al. Cancer Discovery 2011*:)

Case 3: High grade pancreatic neuroendocrine tumor

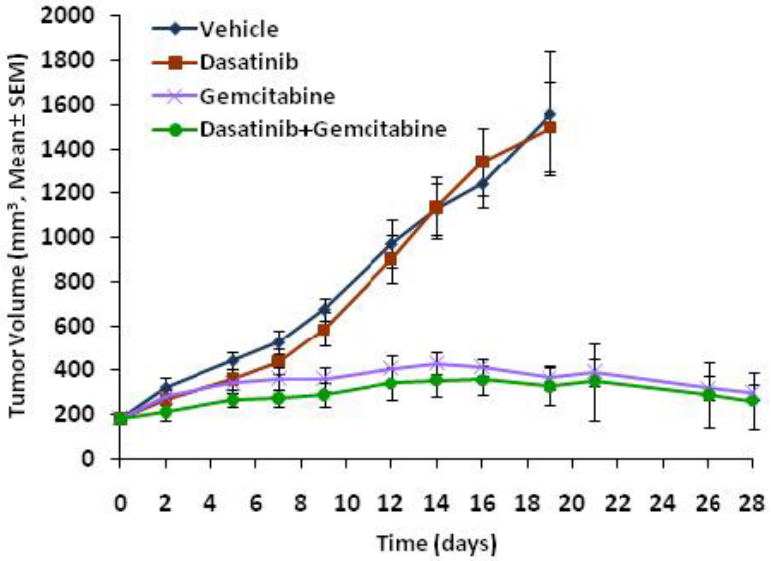
Avatar PDX- Panc 1 test

PI3Ki, MEKi and Gemcitabine



Mutation in PI3K

Dasatinib + Gemcitabine



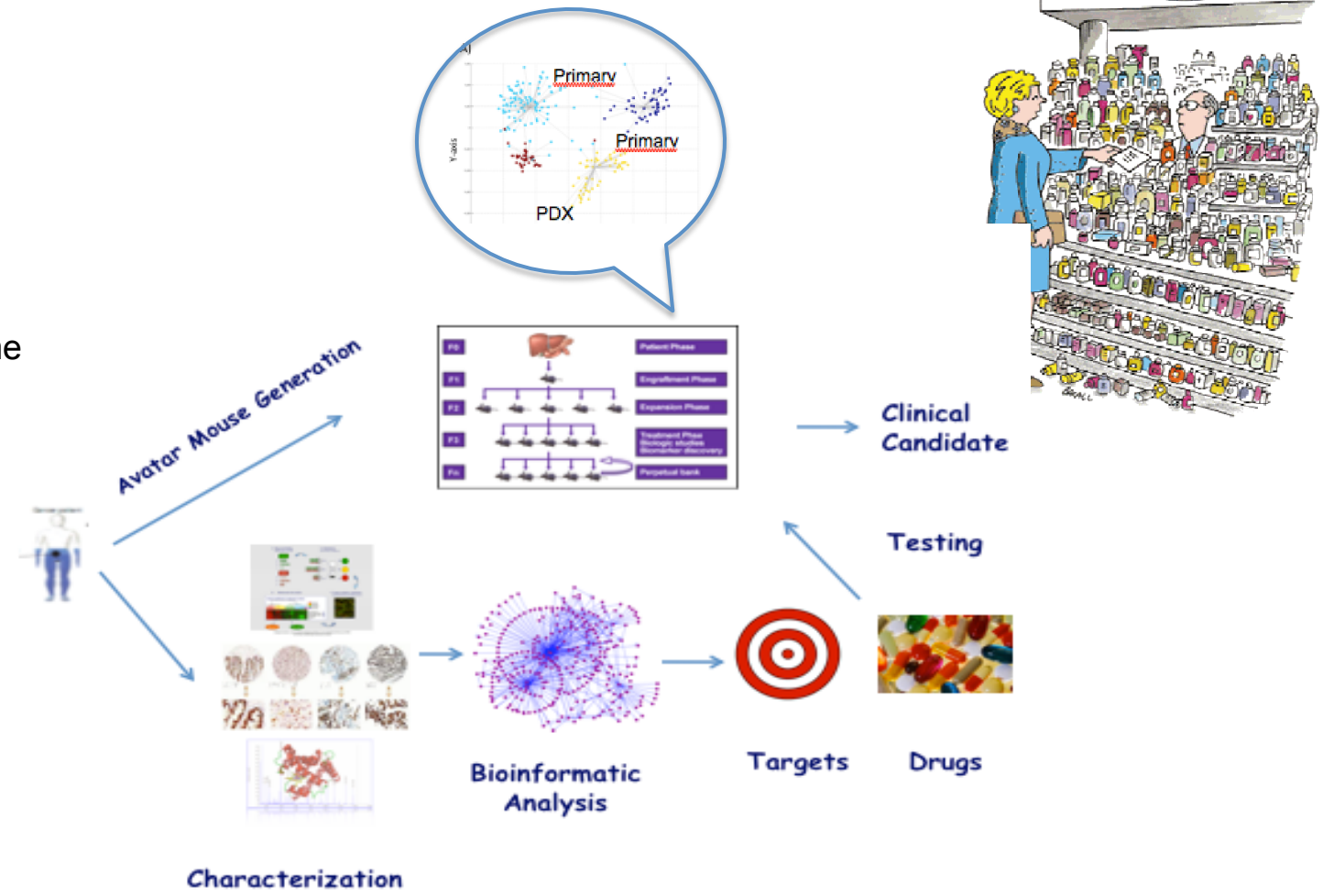
Mutation in DDR2

Patient's treatment

Why it did not work?

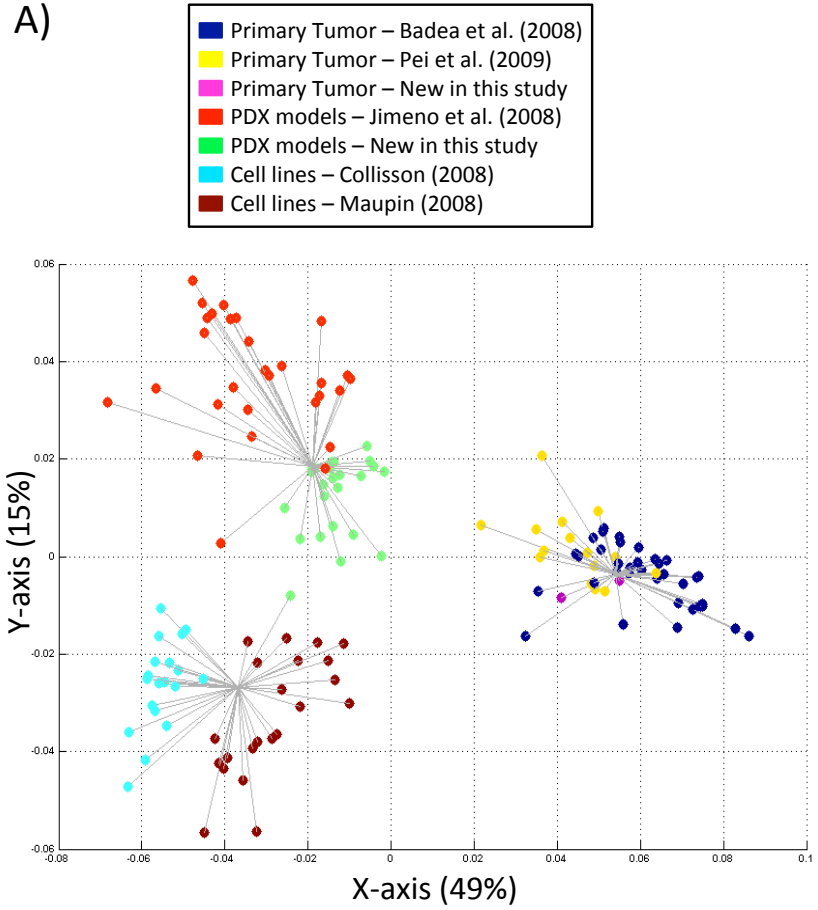
CNIO co-clinical cancer initiative

Manuel Hidalgo,
Director
Clinical Programme
CNIO

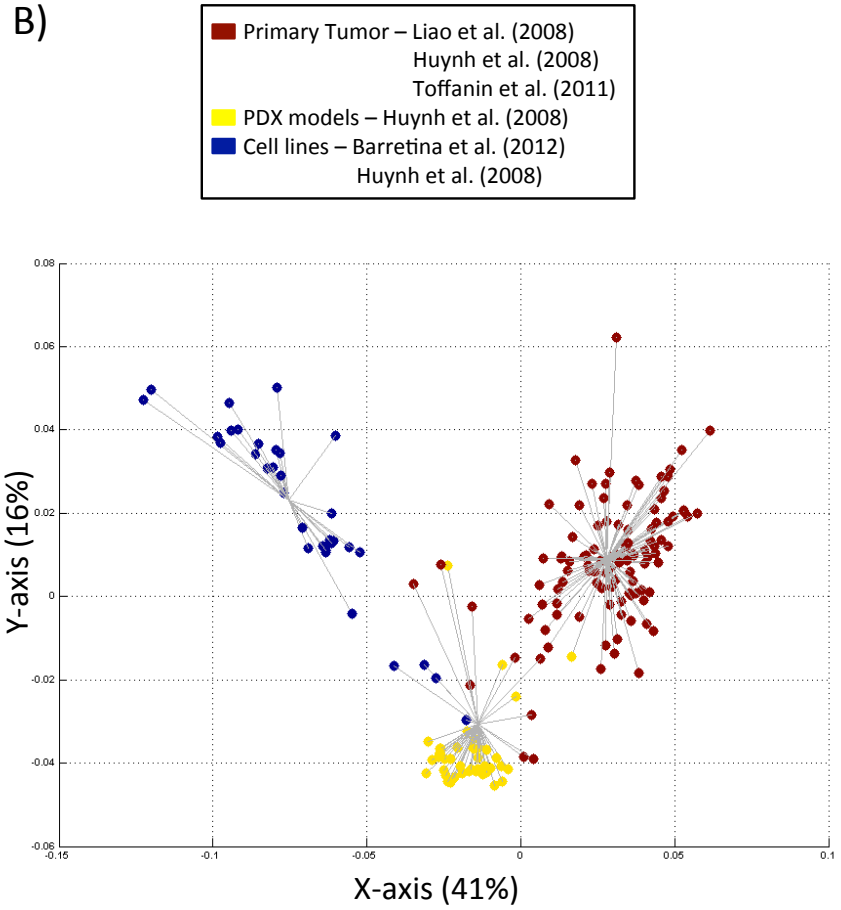


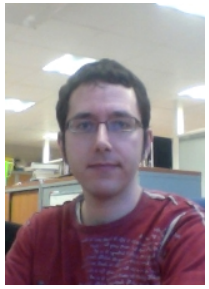
Pancreas PDXs conserve key features of their tissue of origin

A)



B)

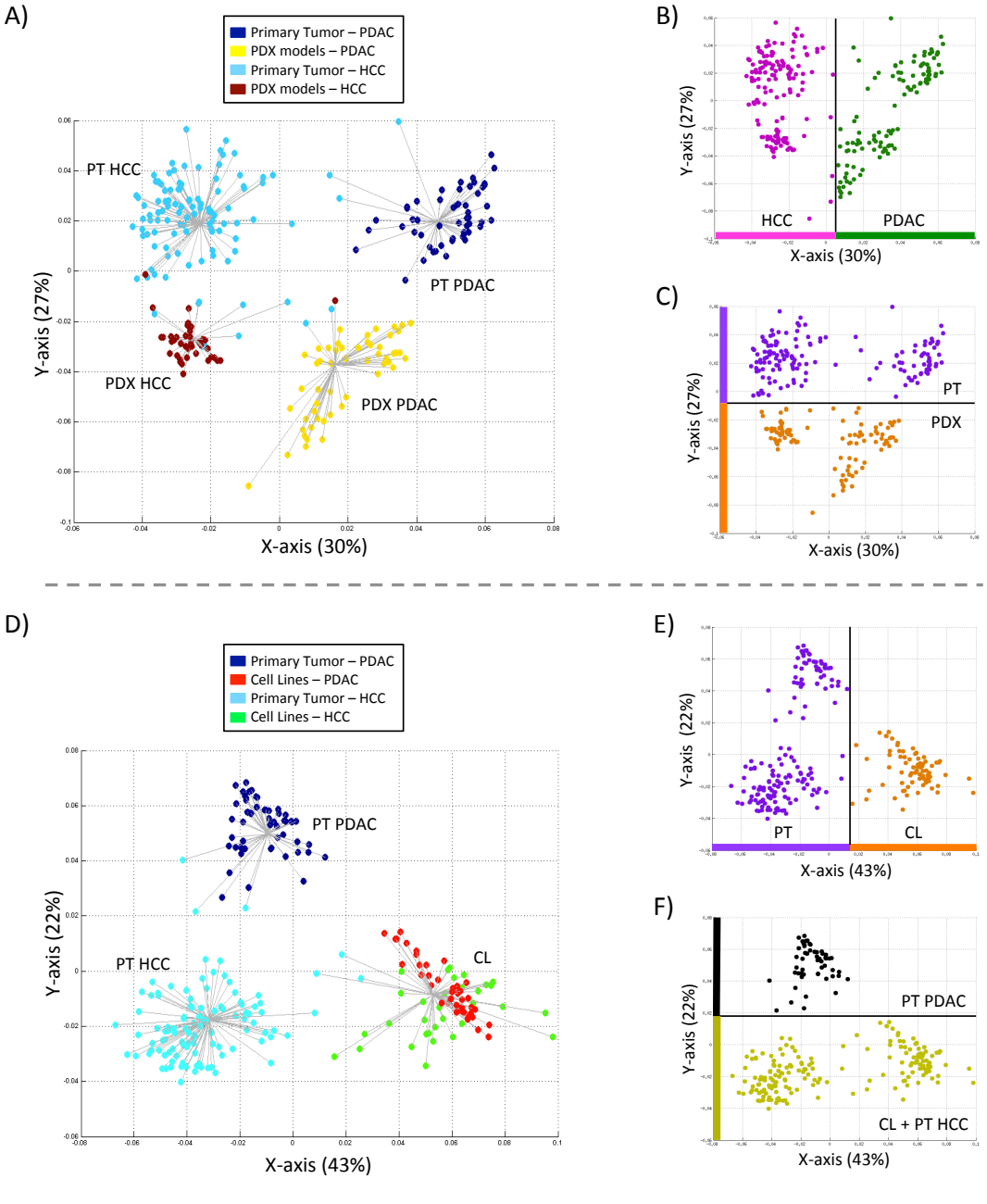




Daniel Rico

Raquel Martinez

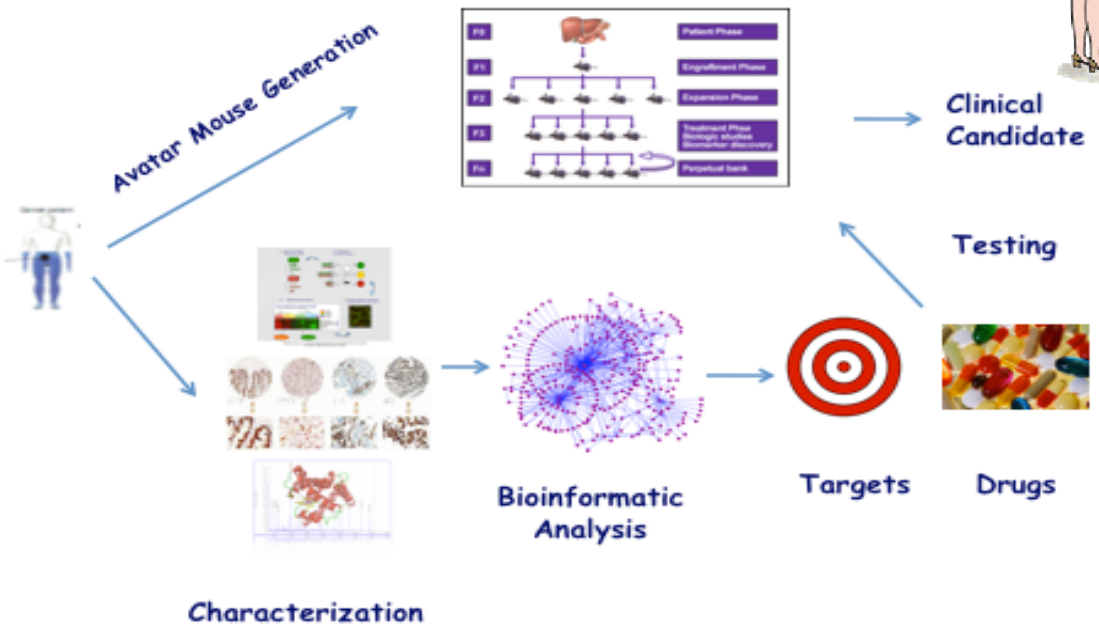
Martinez-Garcia et al., Genome Med 2014



Pancreas PDXs conserve key features of their tissue of origin

Martinez-Garcia et al., Genome Med 2014

CNIO co-clinical cancer initiative



CREDIT: Amparo Garrido/CNIO/Fátima Al-Shahrour
Fátima Al-Shahrour

Email Article
Email Editor
Discuss in Forum
Related Articles
Print Article

Special Issue

Applying Bioinformatics to Precision Medicine

By Ma
So ca res Wh dis Tra pr (C

A new breed of scientist

According to Valencia, the job that Al-Shahrour does requires a very wide range of knowledge and skills; he emphasizes her "biological background, capacity to develop bioinformatics methods, deep understanding of genomics, good communication skills and proved record in team management." Also important, he adds, is her clear understanding of the limitations of the experimental and computational techniques.

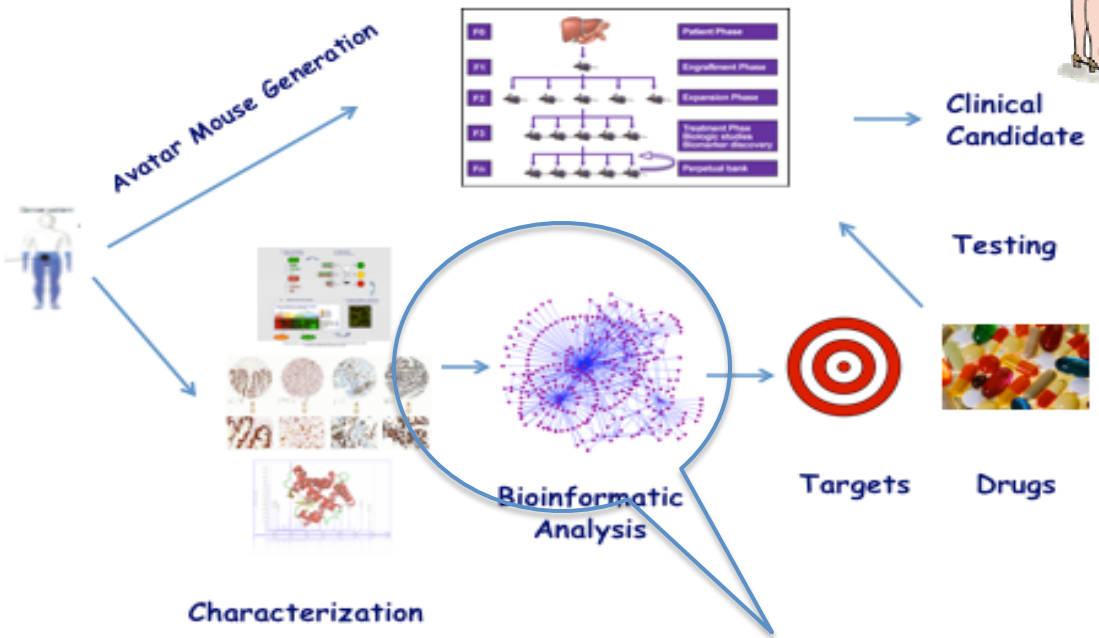
Early-career scientists who wish to follow in her footsteps must be ready to embrace the training challenges. Tamayo writes: "My advice to them is to study mathematics (not only old statistics but also advanced probability), information theory, machine learning, programming, numerical methods, chemistry, physics, cellular biology and biochemistry. It is important not only to be able to talk to multiple domain experts, and develop a solid hard-core analytical mind frame to cast problems, but also to have access to a rich set of paradigms about how to deal with complexity." Cancer pharmacogenomics is "a particularly demanding field that requires a lot of flexibility and adaptability in terms of what problems one solves over time and in requiring to learn from many fields of expertise," he adds.

CNIO personalized Medicine Initiative

- Workflow of PerMed
  - Constant flow of cases
  - Short response time (days)
  - Weekly meeting with clinicians
  - Accumulating statistics to derive patters

- *still … this is a research project (all PerMed are research projects)*

CNIO co-clinical cancer initiative



Mardis Genome Medicine 2010, 2:84
http://genomemedicine.com/content/2/11/84

Genome Medicine

MUSINGS

The $1,000 genome, the $100,000 analysis?

Elaine R Mardis*

Clinical Genetics Has a Big Problem That's Affecting People's Lives

The Atlantic

Unreliable research can lead families to make health decisions they might regret.

Daniel MacArthur at Massachusetts General Hospital found a similar trend in a study of over 60,000 people, the results of which have been uploaded to a pre-print server. On average, each of these volunteers is walking around with 53 gene variants that are classified as "pathogenic" in two widely-used databases. When the team took a closer look at 200 of these variants, they found enough evidence to classify just *nine* of them as pathogenic.

"Reproducibility problems in clinical genetics ... have massive and real-time consequences for thousands of families."

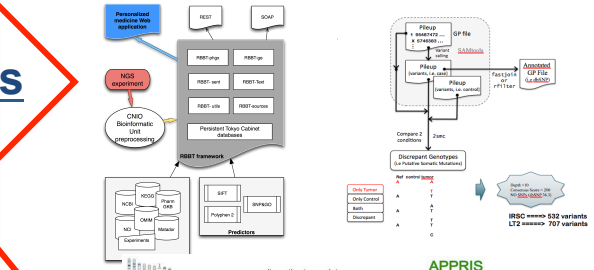
Bioinformatics for personalize medicine

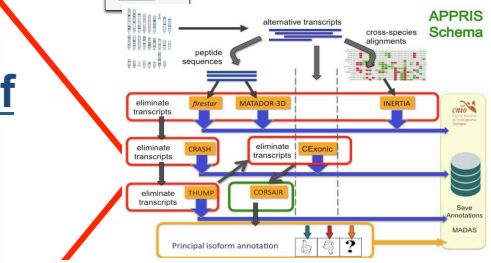
Genome Analysis

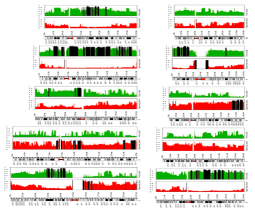
Consequences of alterations:
SNPs, CNVs, miRNA, Epigenetics

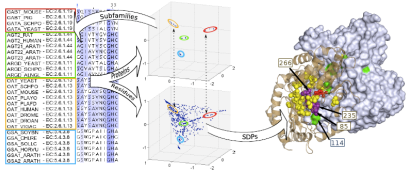
Phenotypic and pathways Analysis

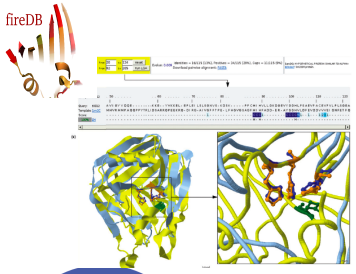
Drug relation
Proteins, genes pathways

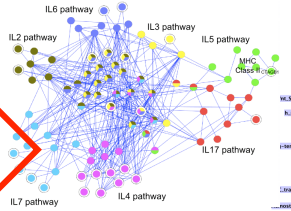


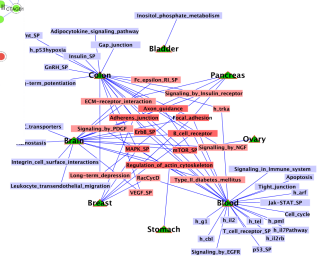


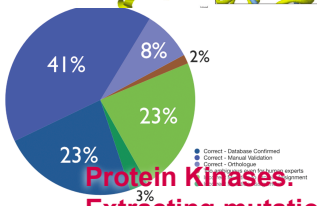




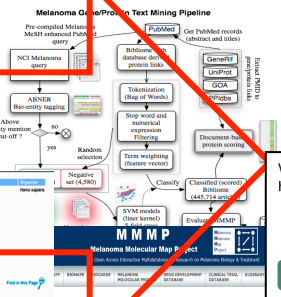








Protein Kinases.
Extracting mutations from text



Valencia and Hidalgo *Genome Medicine* 2012, 4:61
http://genomemedicine.com/content/4/7/61

Genome Medicine

REVIEW

Getting personalized cancer genome analysis into the clinic: the challenges in bioinformatics

Alfonso Valencia* and Manuel Hidalgo

Alternative Splicing

In allowing for the generation of a diverse range of mature RNAs from the same gene, alternative splicing allows has considerable theoretical potential to expand the range of cellular proteins.

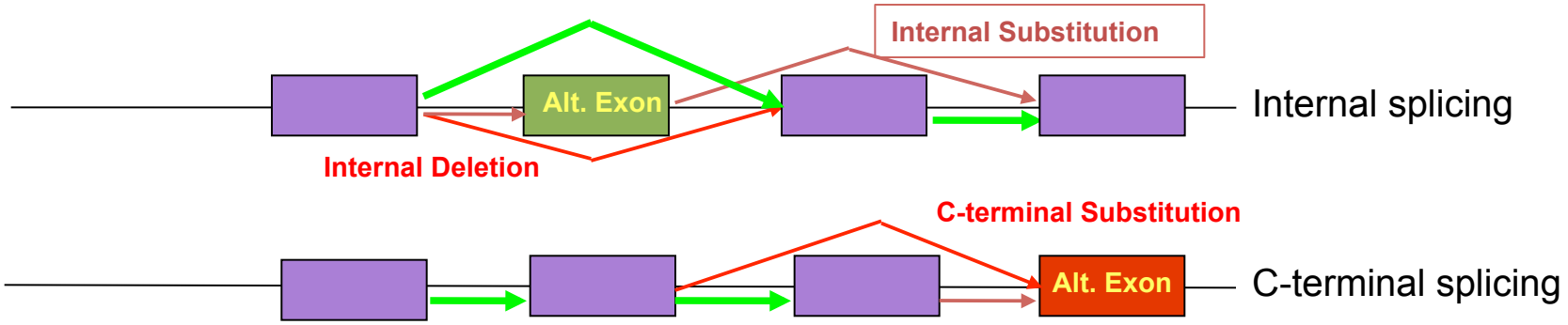
Recent studies estimate that 40–80% of multi-exon human genes can produce differently spliced mRNAs.

There are many studies that implicate alternative transcripts in biological processes such as development.

This has lead to the hypothesis that alternative splicing can explain the apparent lack of correlation between organism complexity and numbers of genes.

In the past few years, it has become clear that a phenomenon called alternative splicing is <u>one reason human genomes can produce such complexity with so few genes.</u>*

*Science, July 2005



by Michael Tress, CNIO

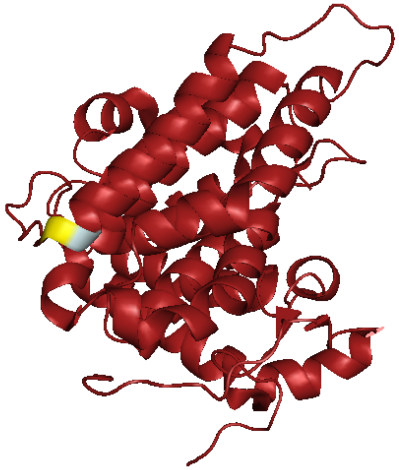
APPRIS - principal functional isoforms

http://appris.bioinfo.cnio.es/

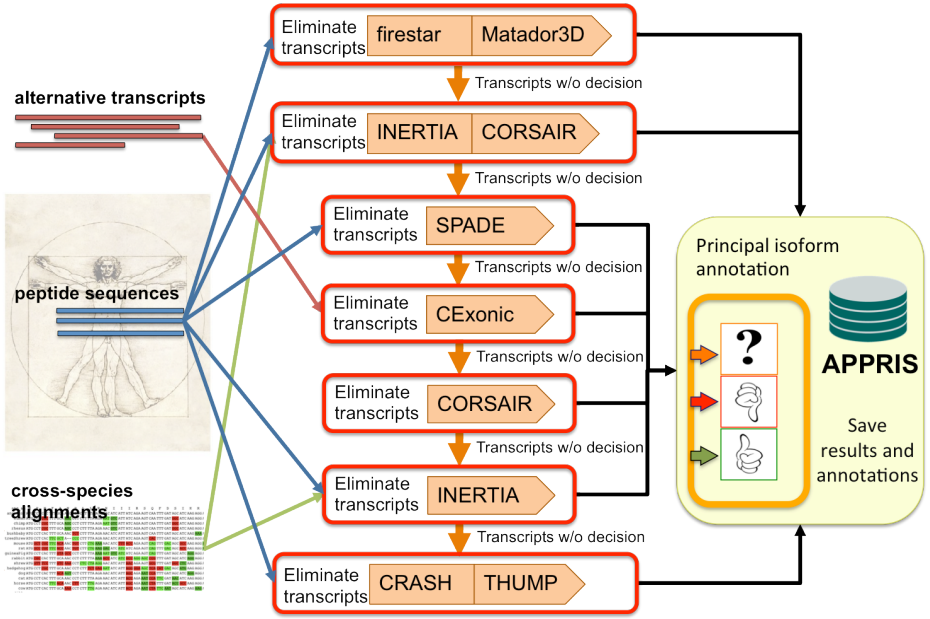
The UniProt database picks one variant and clusters the others around it.

The APPRIS database select **principal isoforms based on:**

- protein structural information,
- functionally important residues,
- protein functional domains and
- evidence of cross- species conservation.



APPRIS Schema



Assumption 1: genes ***have just one principal isoform***

Assumption 2: the principal isoform is the ***oldest in evolutionary terms.***

APPRIS: annotation of principal and alternative splice isoforms.
Rodriguez JM *et al.* Nucleic Acids Res. 2013, 2015
Tress ML *et al.* *Bioinformatics*. 2008

by Michael Tress

2 or more protein isoforms for 0.67% of the human genome

GENETICS

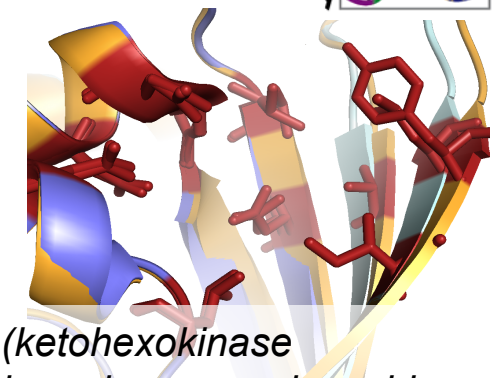
Isoform Identification

The differential reconnection of transcribed exons, termed alternative splicing, has the potential to result in one gene encoding multiple protein isoforms. The degree to which alternatively spliced transcripts are translated into functional proteins, however, is not well understood. Ezkurdia *et al.* used data across multiple mass spectrometry experiments to investigate the degree to which genes with alternative transcripts gave rise to protein isoforms. Comparison of the predicted proteins from the gene and genetic variant database of ENCODE (GENCODE) to the Swiss-Prot database allowed for the identification of 150 human genes that encoded at least one protein isoform and 13 with three or more, with the caveat that identification was biased toward those most likely to be detected. Heterogeneous nuclear ribonucleoproteins, which are involved in the regulation of alternative splicing, showed enrichment in alternative isoforms. Furthermore, the majority of differences detected among all predicted isoforms differed in sequence by the insertion/deletion of a single amino acid. Investigation into the *Drosophila* and mouse proteomes revealed similar patterns. Together, these results suggest that alternative splicing is under selective constraint. — LMZ

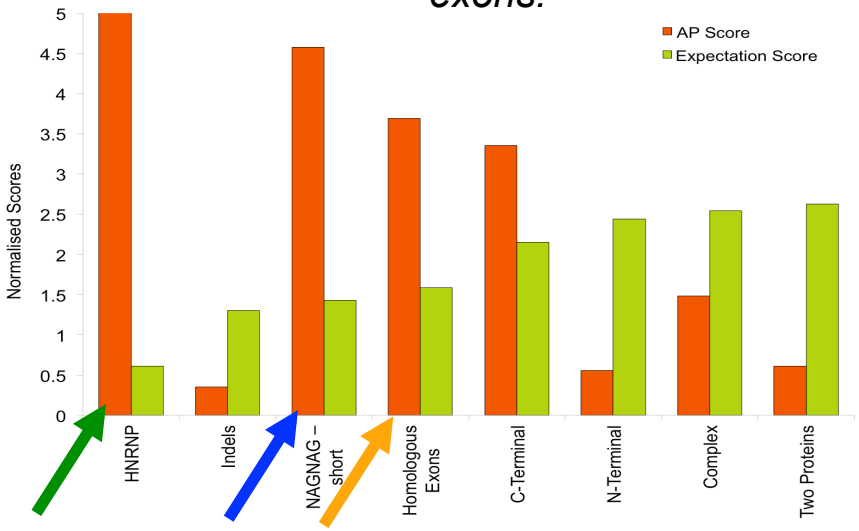
Mol. Biol. Evol. **29**, 10.1093/molbev/mss100 (2012)

Science

Isoforms with small differences are significantly over-represented



(ketohexokinase homologous replaceable exons.



MBE Advance Access published April 17, 2012

Comparative Proteomics Reveals a Significant Bias Toward Alternative Protein Isoforms with Conserved Structure and Function

Iakes Ezkurdia,†,[1] Angela del Pozo,†,[1] Adam Frankish,[2] Jose Manuel Rodriguez,[1] Jennifer Harrow,[2] Keith Ashman,[3] Alfonso Valencia,*,[1] and Michael L. Tress*,[1]

PLoS Comput Biol. 2015

Michael Tress

Staff Scientist
National Center for Cancer Research (CNIO)
2006 – Present (7 years)

Research on protein structure prediction, protein fu
alternative splicing.

Post-Doctoral Researcher
CNB
March 2002 – April 2006 (4 years 2 months)

Researcher in bioinformatics/computatioal biology

Education

University of Warwick
1997 – 2001

University of York
MSc, Biological Cmputation
1989 – 1990

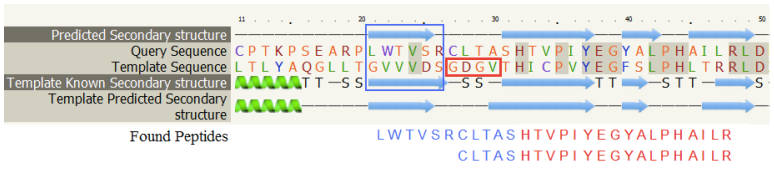
Bath Spa University
BSc, Biochemistry
1983 – 1987

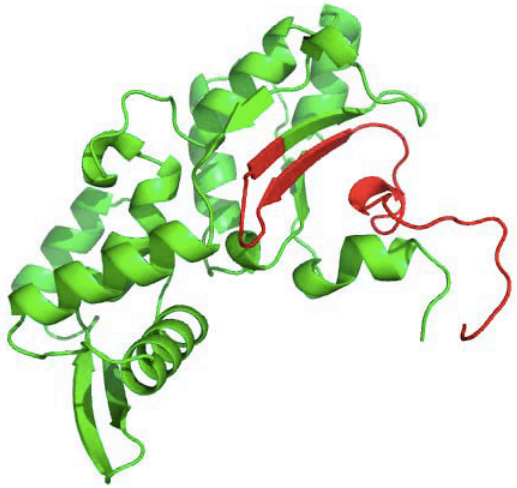
A single isoform (principal isoform) per gene

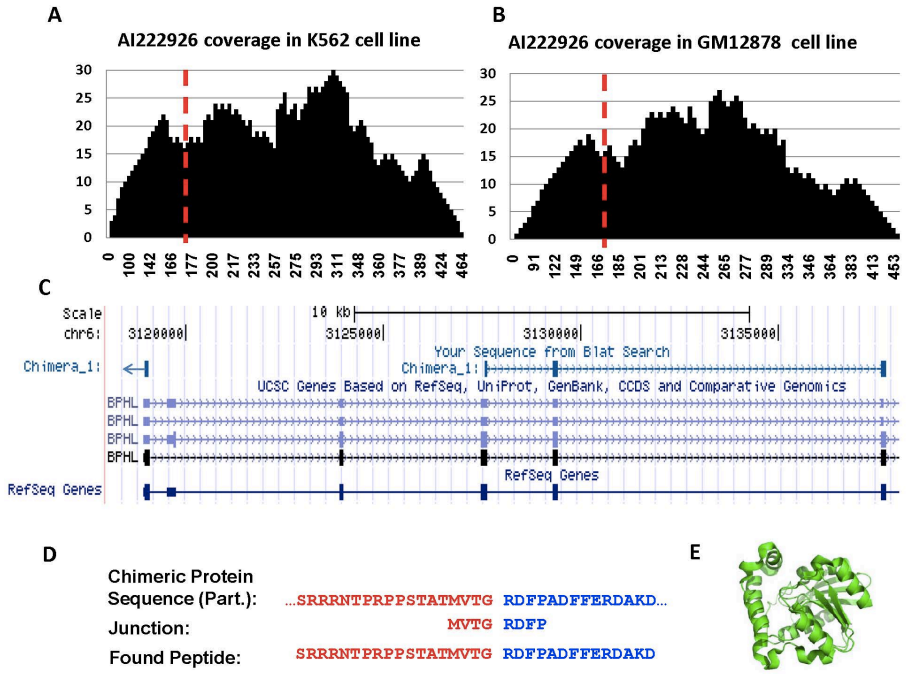
Alterations in isoform composition in cancer

More than anticipated trans-splicing events

Chimeras confirmed at the RNA and protein levels







Chimera of **actin**, ACTG1 and **ribosomal protein** RPL13A

Overlapping unique peptides from 18 mass spectrometry experiments (Pvalue<10^{-52}) confirming transcript from ChimerDB (Kim et al., 2006, Kim et al. 2010) and a few further confirmed by SRM

The motif 'GDGV' (a red rectangle) is the ATP-binding site missing in the chimera

Transplicing/ chimeric mRNAs

POST TRANSCRIPTIONAL REGULATION

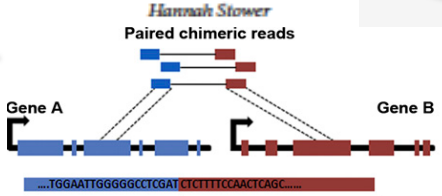
Chimeric protein production

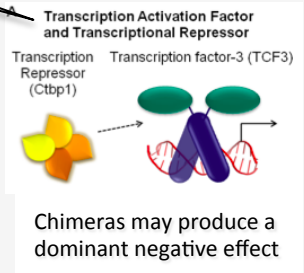
Chimeric transcripts arise from the joining of exons from two or more different genes. Reporting of such transcripts has become more widespread as data from genome-wide transcriptional analyses has increased. However, the number of known chimeric transcripts far outnumbers the reported number of chimeric proteins. Here, Frenkel-Morgenstern et al increase the number of identified translated chimeric transcripts and describe features indicative of their biological functionality.

Firstly the authors analysed previously reported human tissue RNA-seq datasets for the presence of chimeric reads, that is, those that do not align to annotated transcripts and include a chimeric exon-exon junction. This approach confirmed the expression of 175 out of 7,424 previously reported chimeric transcripts from 16 human tissues. Analysis of the level of expression of these transcripts revealed that while chimeric transcripts are themselves expressed at a low level, they incorporate transcripts that are normally expressed at a high level; they are also expressed in a highly tissue-specific manner.

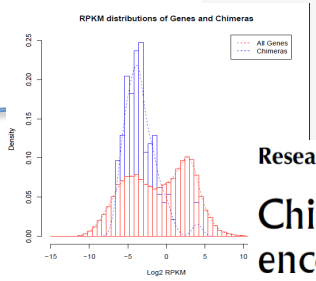
To confirm the translation of these transcripts, the authors both searched mass spectrometry databases and generated their own shotgun mass spectrometry datasets. They initially searched for peptides that spanned a junction site of the human chimeric transcripts and they found 12 chimeric proteins. Following this,they generated targeted mass spectrometry data and this confirmed the expression of a further three novel chimeric proteins. The 175 expressed chimeras are enriched in signal and transmembrane peptides suggesting that generating chimeric transcripts is a mechanism for altering the localisation of the translated protein. Thus the authors' data suggests a potential function for chimeric proteins. The search is on to elucidate precise biological roles.

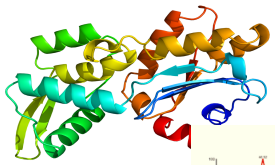
Nat Rev Genet



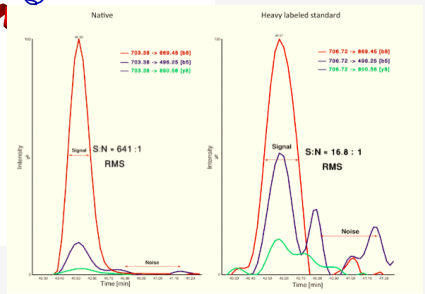


Chimeras may produce a dominant negative effect





Caption



Research

Proof Only

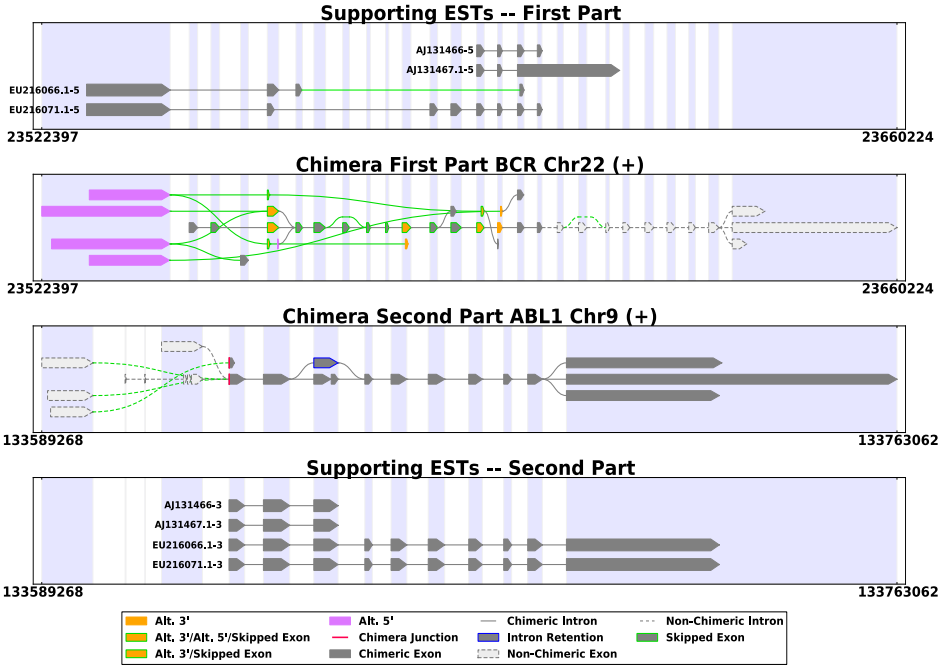
Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts

Milana Frenkel-Morgenstern,[1] Vincent Lacroix,[2] Iakes Ezkurdia,[1] Yishai Levin,[3] Alexandra Gabashvili,[3] Jaime Prilusky,[4] Angela del Pozo,[1] Michael Tress,[1] Rory Johnson,[5] Roderic Guigo,[5] and Alfonso Valencia[1,6]

Genome Research
www.genome.org

Two overlapping mass-spec peptides, (18 experiments, Pvalue<10-7, FDR<1%)
Dr. Levin, Weizmann Institute

ChiTaRS: chimeric transcripts/proteins Database



Research

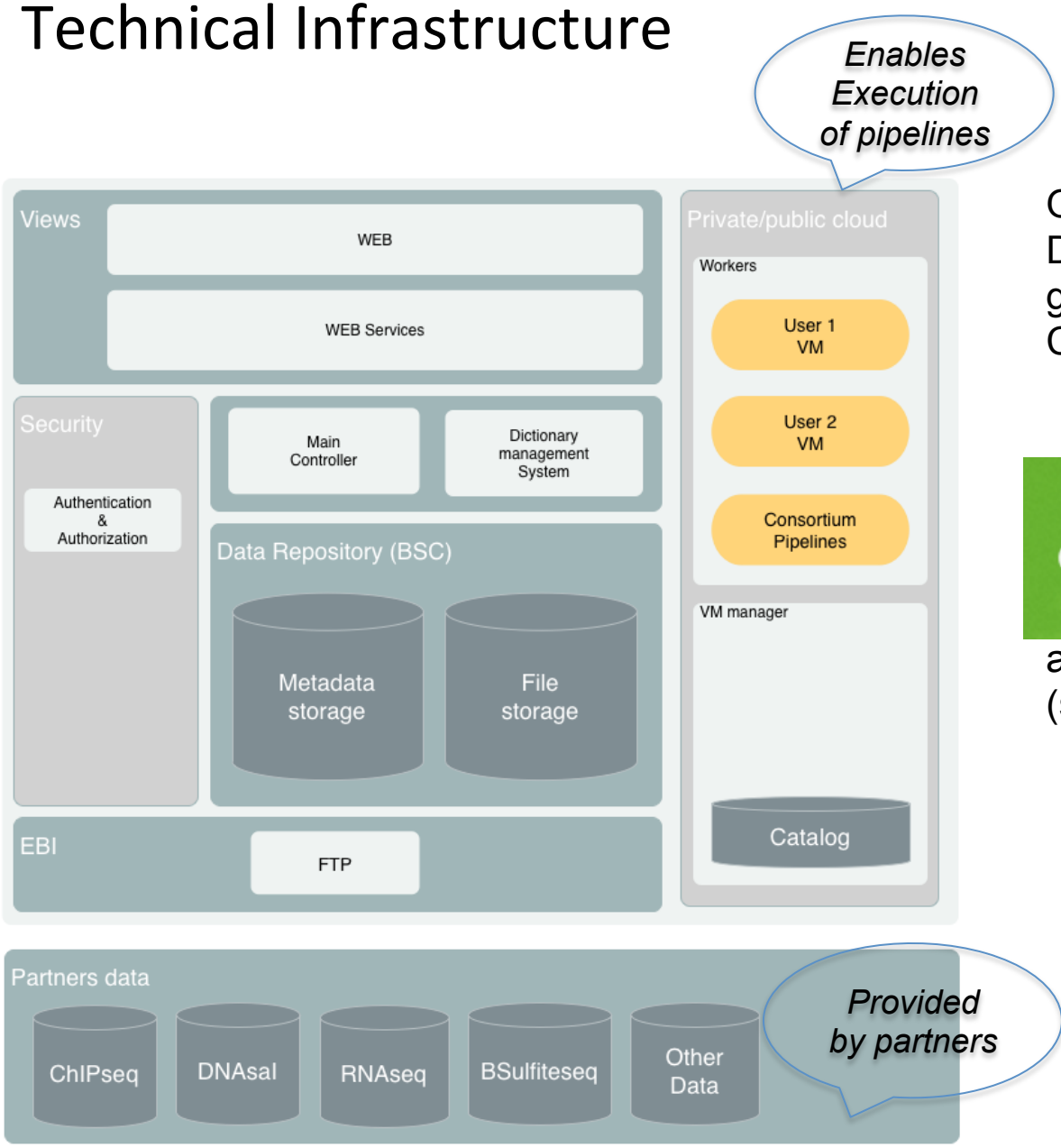
Chimeras taking shape: Potential functions of proteins encoded by chimeric RNA transcripts

Milana Frenkel-Morgenstern,[1] Vincent Lacroix,[2] Iakes Ezkurdia,[1] Yishai Levin,[3] Alexandra Gabashvili,[3] Jaime Prilusky,[4] Angela del Pozo,[1] Michael Tress,[1] Rory Johnson,[5] Roderic Guigo,[5] and Alfonso Valencia[1,6]

Genome Research
www.genome.org

Milana Frenkel-Morgenstern et al., NAR 2013. 2015

Technical Infrastructure



Rbbt

Ruby bioinformatics toolbox

Operation at file system level
Developed CNIO, accessible github, will be installed at CNAG.

elasticsearch.

as in ICGC
(similar to MongoDB)

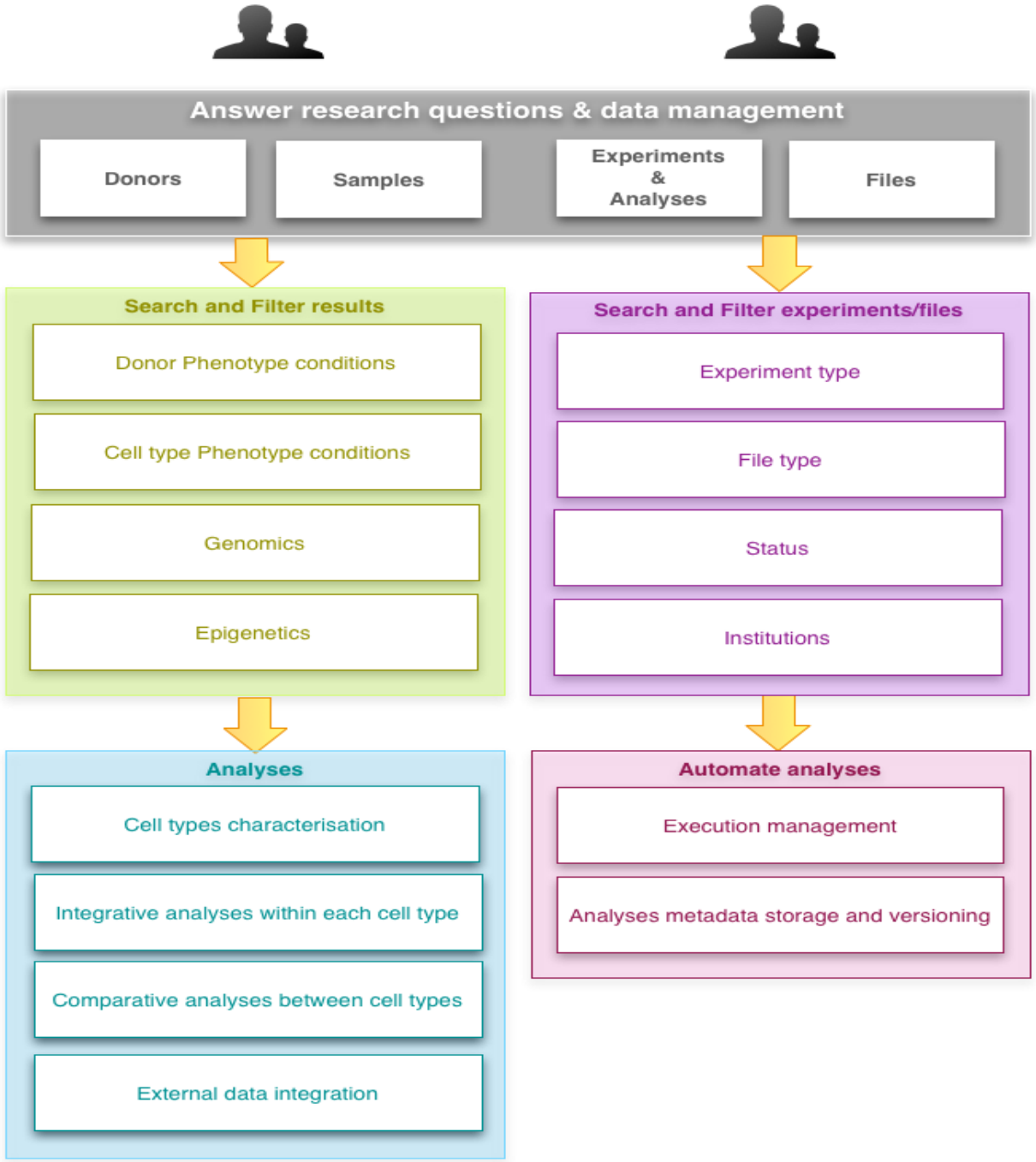
ANGULARJS by Google

as in ICGC frontend

Use cases are divided in:

1. Data management
2. Answer research questions

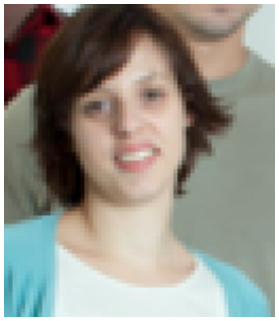
The data portal provides functionalities to satisfy both categories.



Bioinformatics Core Unit

Gonzalo Gómez

Osvaldo Graña



Miriam Rubio

David G. Pisano

Spanish National Bioinformatics Institute

Victor de la Torre

Miguel Vazquez

José María Fernández-González

INB Instituto Nacional de Bioinformática

elixir SPAIN

A Network Biology Approach to Information Processing in Epigenetic Regulation

Alfonso Valencia
Spanish National Cancer Research Centre
CNIO

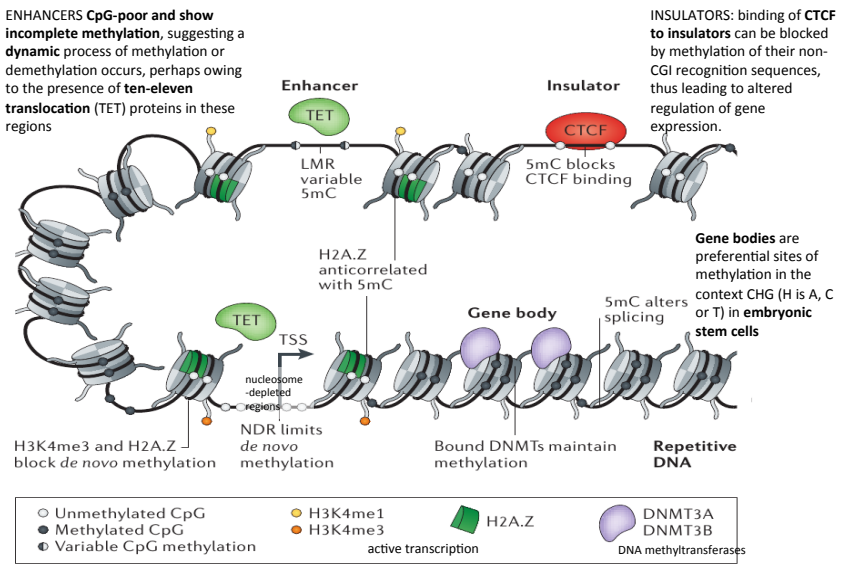
valencia@cnio.es
@alfons_valencia

ESHG symposium 2016: Machine Learning for Personalized Medicine

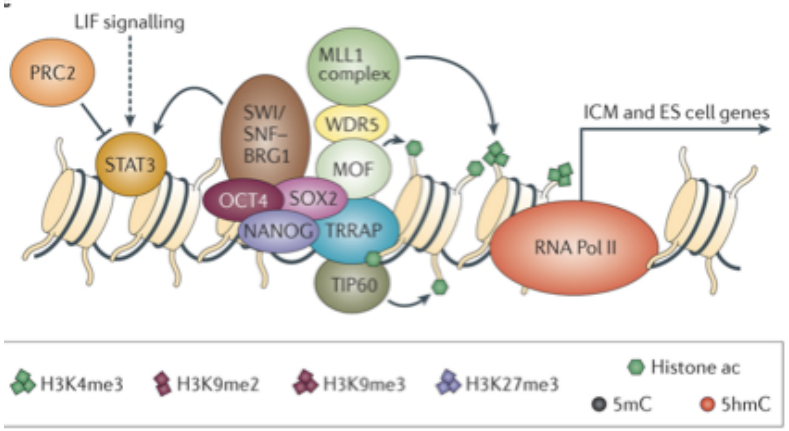
Barna May 2016

EPIGENETICS/EPIGENOMICS

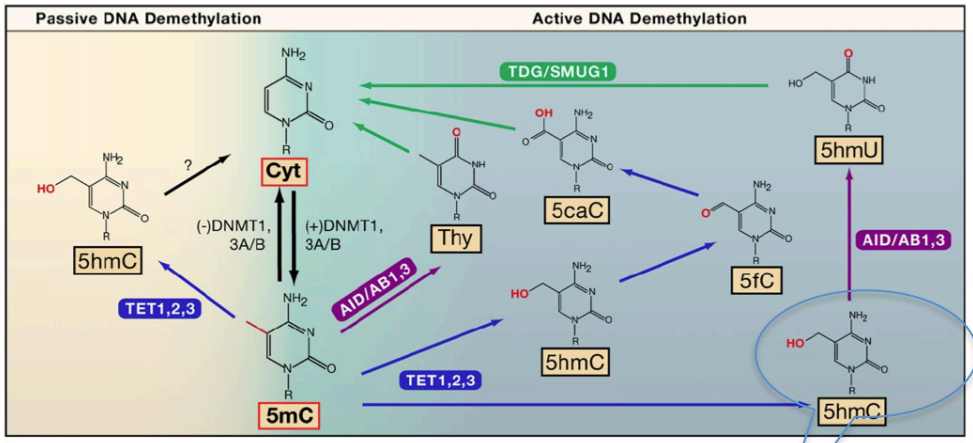
Complex relationship between Cytosine modifications, Histone marks and Chromatin Binding Proteins



Nat Rev Genet. 2012 May 29;13(7):484-92



Nature Reviews | Genetics



Stem cells

Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1

Mamta Tahiliani[1], Kian Peng Koh[1], Yinghua Shen[2], William A. Pastor[1], Hozefa Bandukwala[1], Yevgeny Brudno[2], Suneet Agarwal[3], Lakshminarayan M. Iyer[4], David R. Liu[2], L. Aravind[4], and Anjana Rao[1]

Science. 2009 May 15; 324(5929): 930–935. doi:10.1126/science.1170116.

The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain

Skirmantas Kriaucionis and Nathaniel Heintz

Science **324**, 929 (2009);

DOI: 10.1126/science.1169786

Science AAAS

Segmenting the Genome in Functional Regions

Mouse embryonic stem cells
"Core" epigenomic features (basic scaffold)

Histone modifications

H3K4me3 --- ENCODE, GSE11724, GSE12241, GSE36114
H3K4me2 --- ENCODE, GSE11172, GSE36114
H3K4me1 --- ENCODE, GSE11172, GSE36114
H3K27Ac ---- ENCODE, GSE36114,
H3K9me3 --- ENCODE, GSE12241, GSE18371
H2Aub1 --- GSE34518
H3K27me3 --- GSE12241, GSE41589, GSE36114
H3K36me2 --- GSE41589
H3K36me3 --- ENCODE, GSE11724, GSE12241, GSE34518, GSE41589, GSE36114
H3K79me2 --- GSE11724
H4K20me3 --- GSE12241
H3K9Ac --- ENCODE
H2AZ --- GSE36114

Cytosine modifications

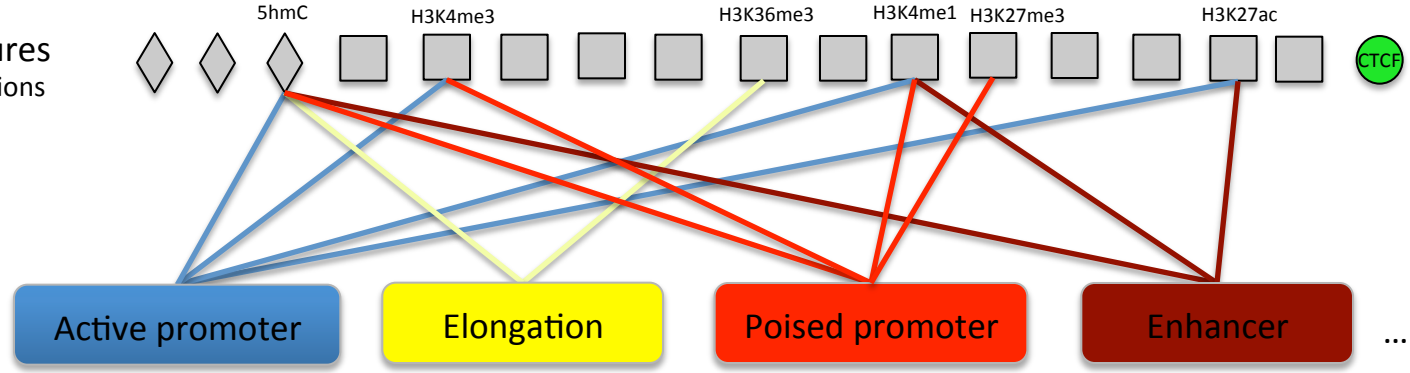
5mC --- GSE28682
5hmC --- GSE28682
5fC --- GSE40148

Insulator

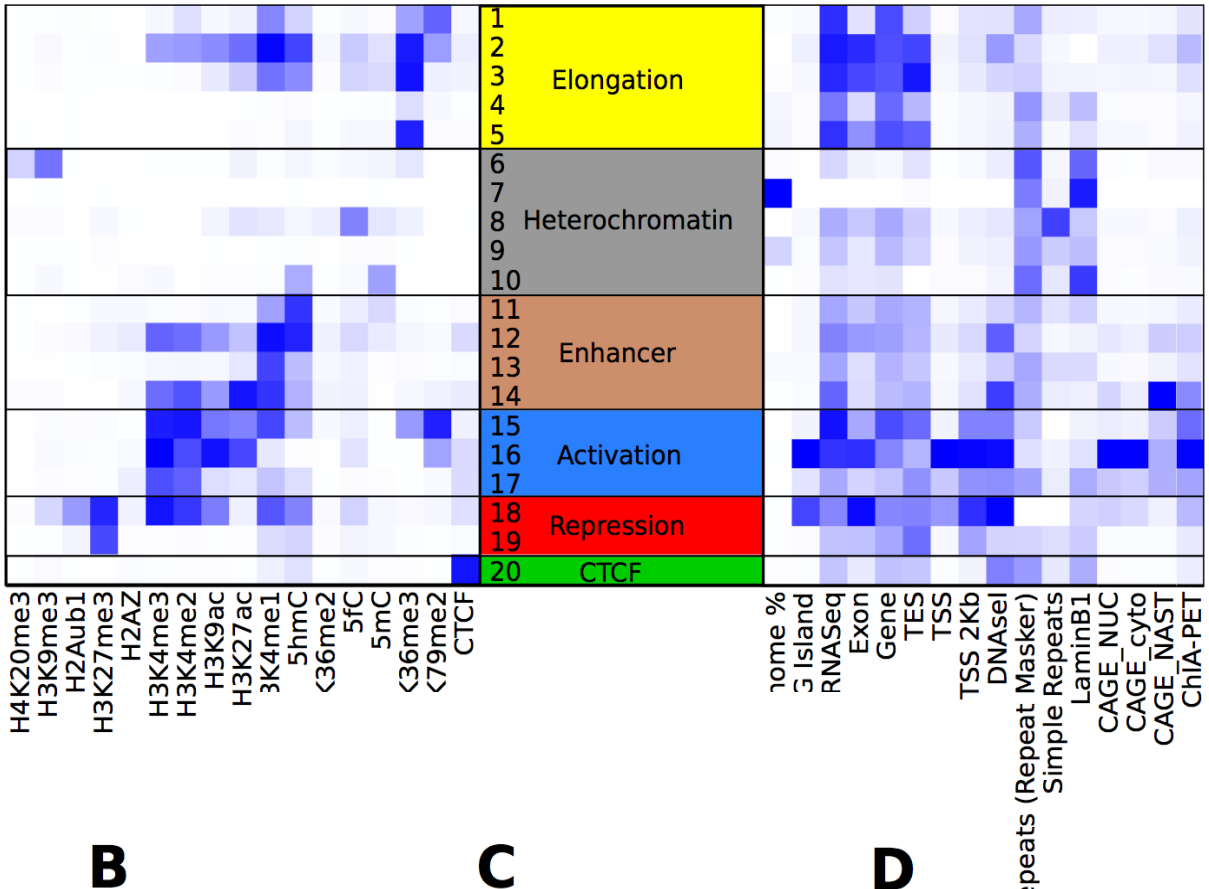
CTCF --- GSE11431, GSE25777, GSE28247

17 “Core” epigenomic features
CTCF, cytosine & histone modifications

20 Chromatin states
Combinations of core features



ChromHMM



20 Chromatin states obtained by combining ChipSeq data of 17 core features

B C D

Chromatin Related Proteins (data from mouse ESCs)

Polycomb

CoREST --- GSE27841
REST --- GSE27841
RING1B --- GSE34518, GSE42466
SIN3A --- GSE24843
SUZ12 --- GSE11724, GSE42466, 41589, 11431
PHF19 --- GSE41589, GSE41609

ESC Transcription factors

C-MYC --- GSE11431
E2F1 --- GSE11431
MAX --- GSE48175
NANOG --- GSE11431, GSE11724
N-MYC --- GSE11431
OCT4 --- GSE11431, GSE11724
SMAD1 --- GSE11431
SOX2 --- GSE11431, GSE11724
STAT3 --- GSE11431
KLF4 --- GSE11431
Tcfcp2l1 --- GSE11431

Enhancer

P300 --- GSE11431, GSE28247

Cohesin complex

NIPBL --- GSE22562
SMC1 --- GSE22562
SMC3 --- GSE22562
RAD21 ---- GSE25777

Others

CBX3 --- GSE4424
CBX7 --- GSE42466
HCFC1 --- ENCODE
MAFK ---- ENCODE
ZC3H11A --- ENCODE
ZNF384 --- ENCODE
ESRRB – GSE11431
KAP1 --- GSE41903
LAMINB --- GSE28247
MI2B --- GSE27841
OGT --- GSE39154
TCF3 --- GSE11724
BRG1 --- GSE14344

Chromatin remodelers

HDAC1 --- GSE27841
HDAC2 --- GSE27841
KDM2A --- GSE40860
KDM2B ---- GSE40860
LSD1 --- GSE18515, GSE27841
TET1 --- GSE24843
MBD1A --- GSE39610
MBD1B --- GSE39610
MBD2A --- GSE39610
MBD2T --- GSE39610
MBD3A --- GSE39610
MBD4 --- GSE39610
MECP2 --- GSE39610
MLL2 --- GSE48172
SETDB1 –- GSE18371

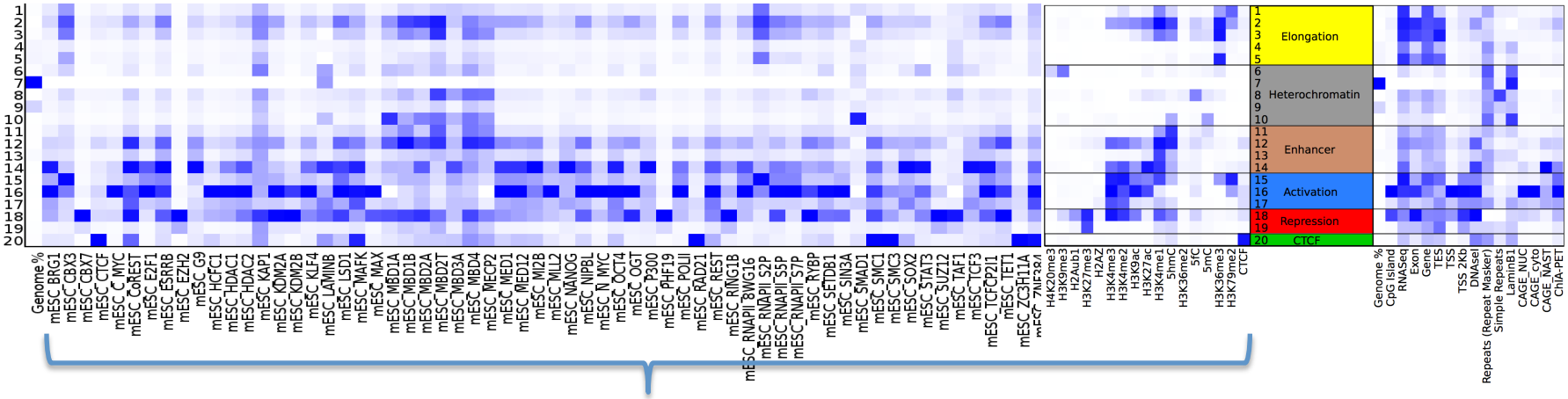
Transcription related

POLII --- GSE12241, GSE28247
RNAPII-8WG16--- GSE34518
RNAPII-S2P --- GSE34518
RNAPII-S5P ---- GSE34518
RNAPII-S7P ---- GSE34518
MED12 --- GSE22562
MED1 --- GSE22562
TAF1 --- GSE36114

Epigenomic Mouse Stem Cell co-localization Network

Core epigenomic features

60 Chromatin-related Proteins (CrPs)



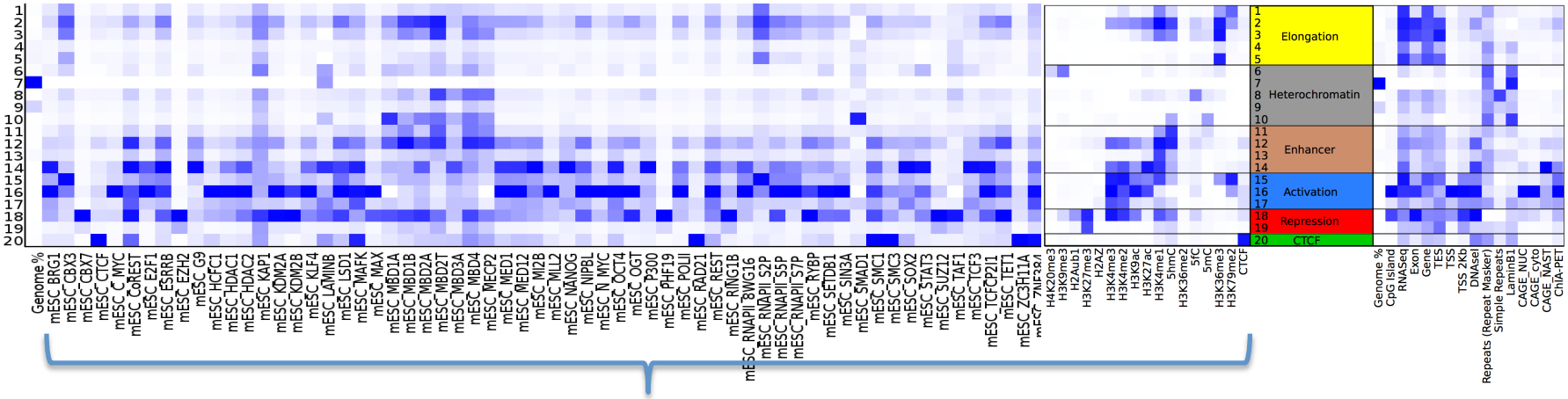
Distribution of CRPs / Epi features in 20 Heterochromatin states

Heterochromatin state dependent Co-localization Networks

Epigenomic Mouse Stem Cell co-localization Network

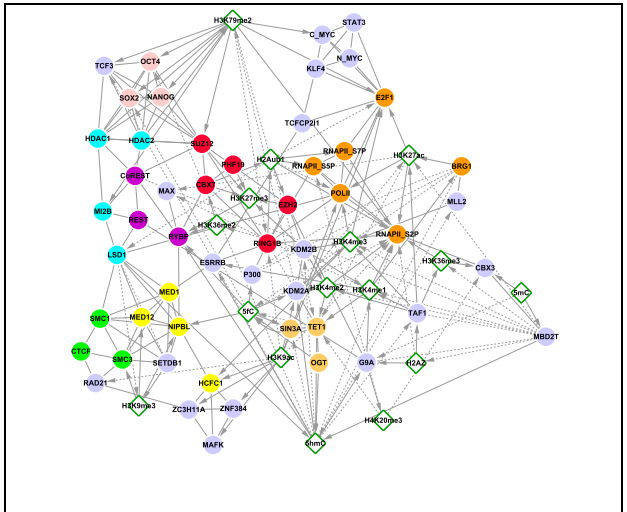
60 Chromatin-related Proteins (CrPs)

Core epigenomic features

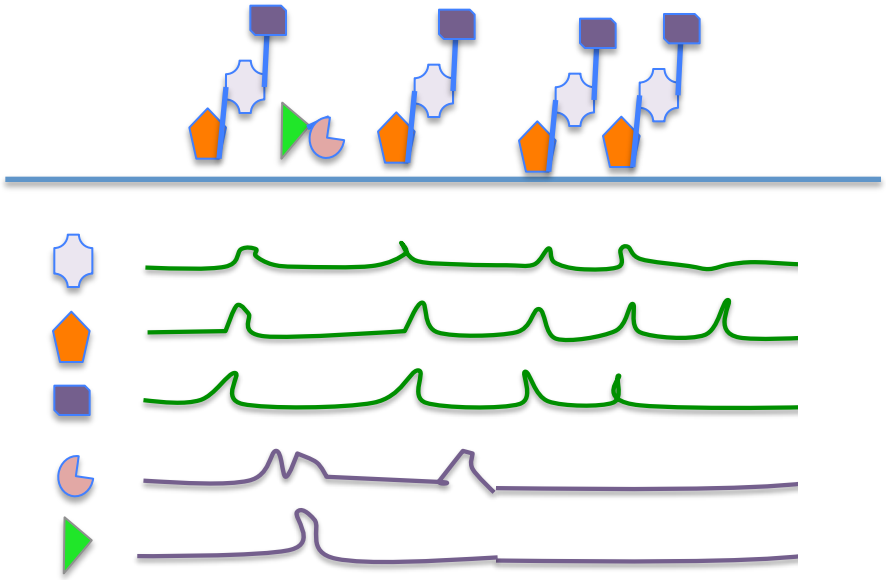


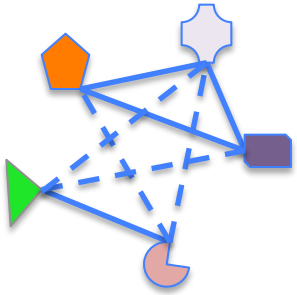
Distribution of CRPs in 20 Heterochromatin states

Heterochromatin state dependent Co-localization Networks



Inference of chromatin **DIRECT co-localiztion networks** from ChIP-seq data





Co-localization Correlation network
With **positive** interactions and
mutual exclusion (**negative**) interactions

Sparse Partial Correlation Networks
(Lasserre et al. 2013)

Regularized linear regression with Elastic Nets
(Perner et al. 2014)

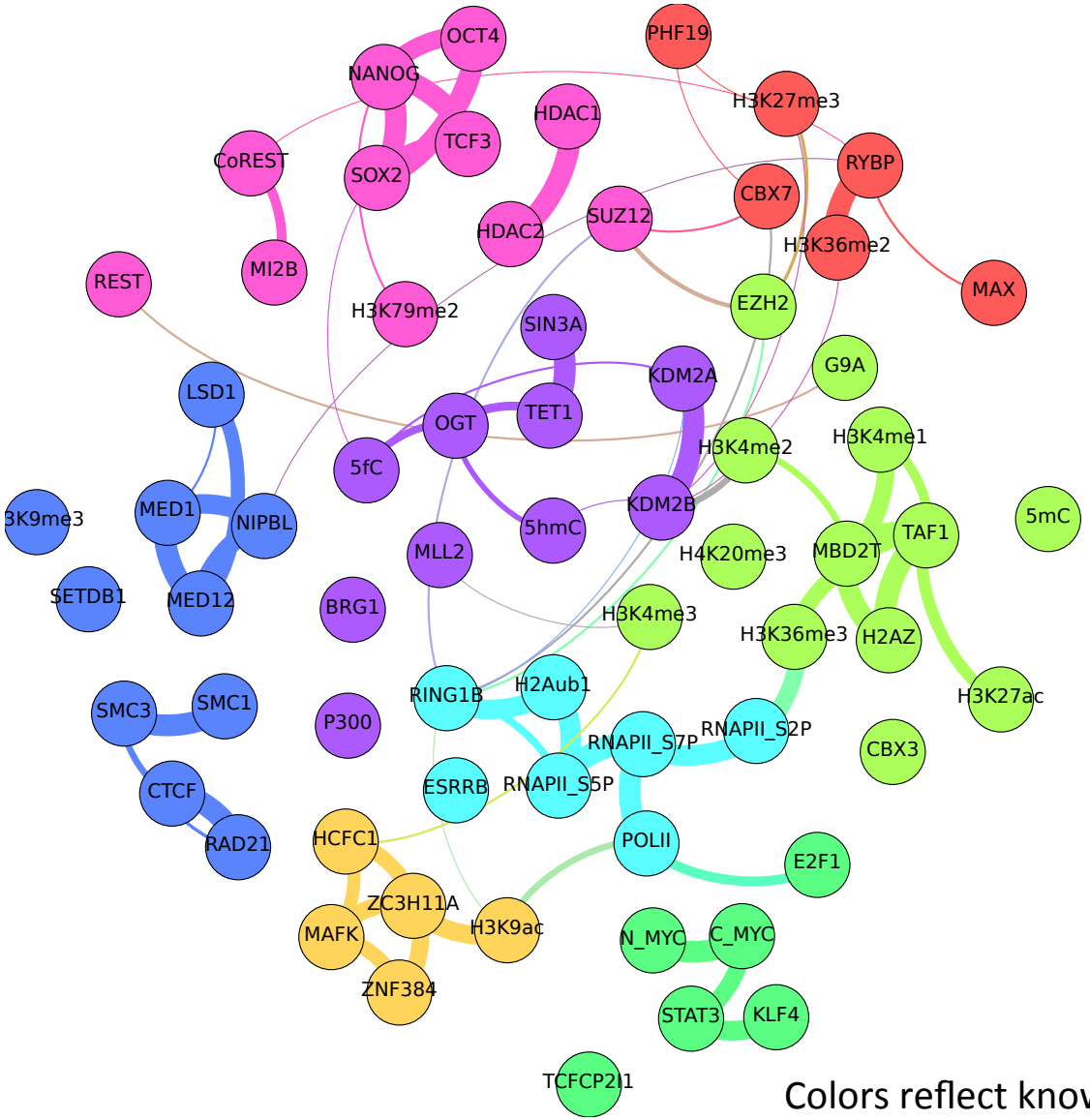
MAX-PLANCK-GESELLSCHAFT

Juliane Perner

Martin Vingron

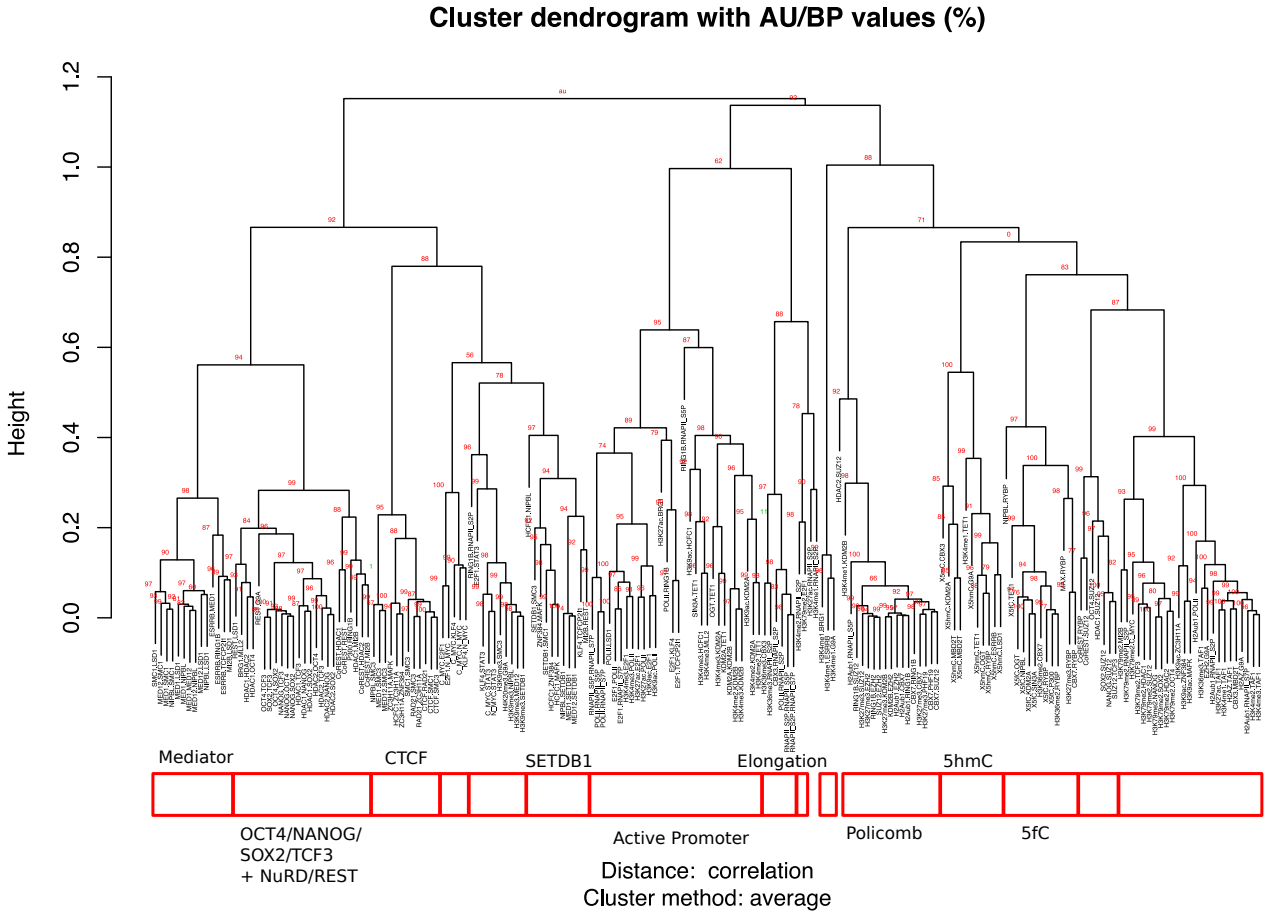
Ho-Ryun Chung

State 18 (Polycomb promoter)

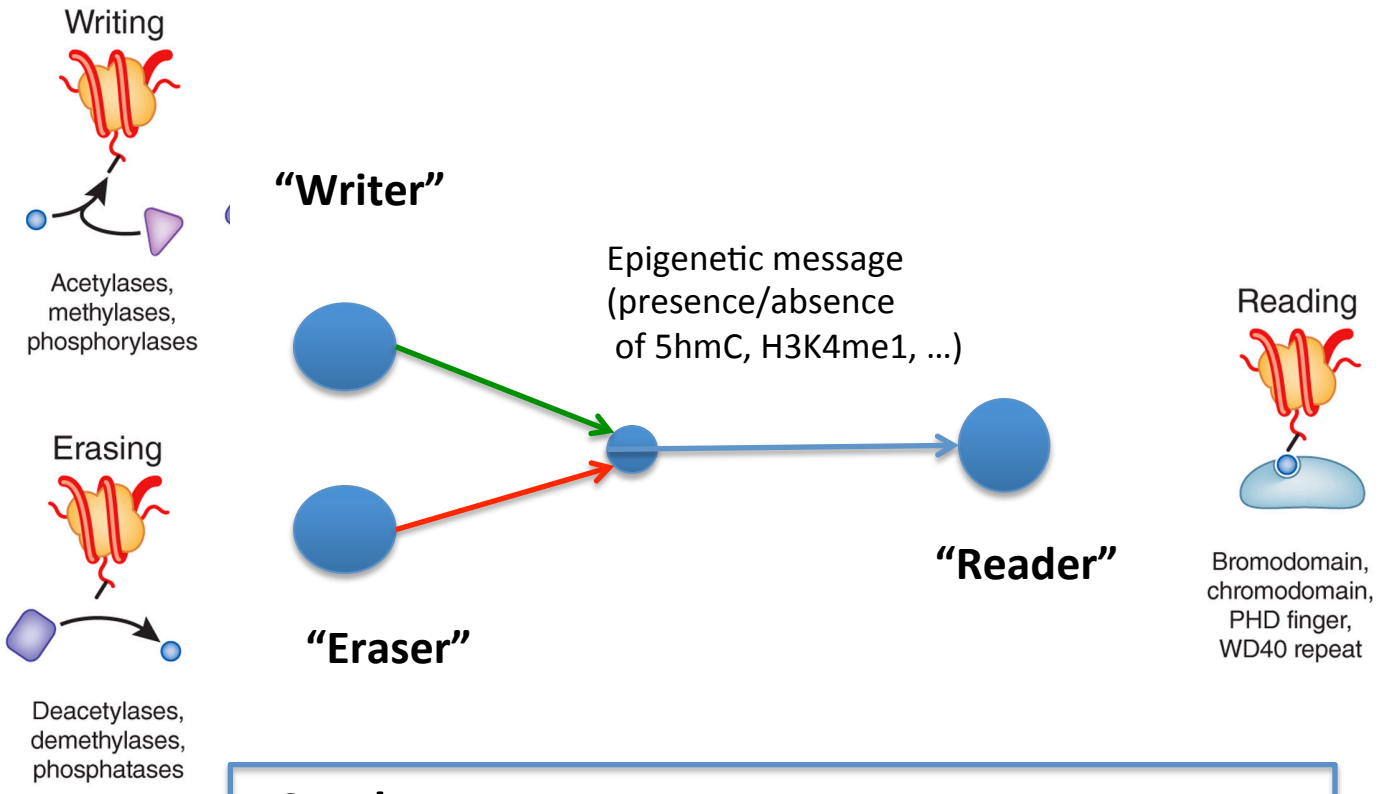


Colors reflect known protein complexes

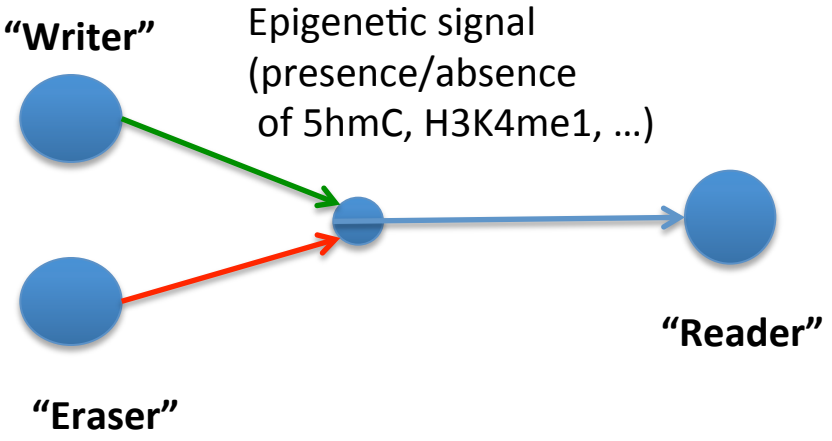
Global Co-localization network reflects organization in protein complexes



Epigenetics as a communication system



From “Location” to “Communication” Network



Sender → **Signal** → **Receiver**

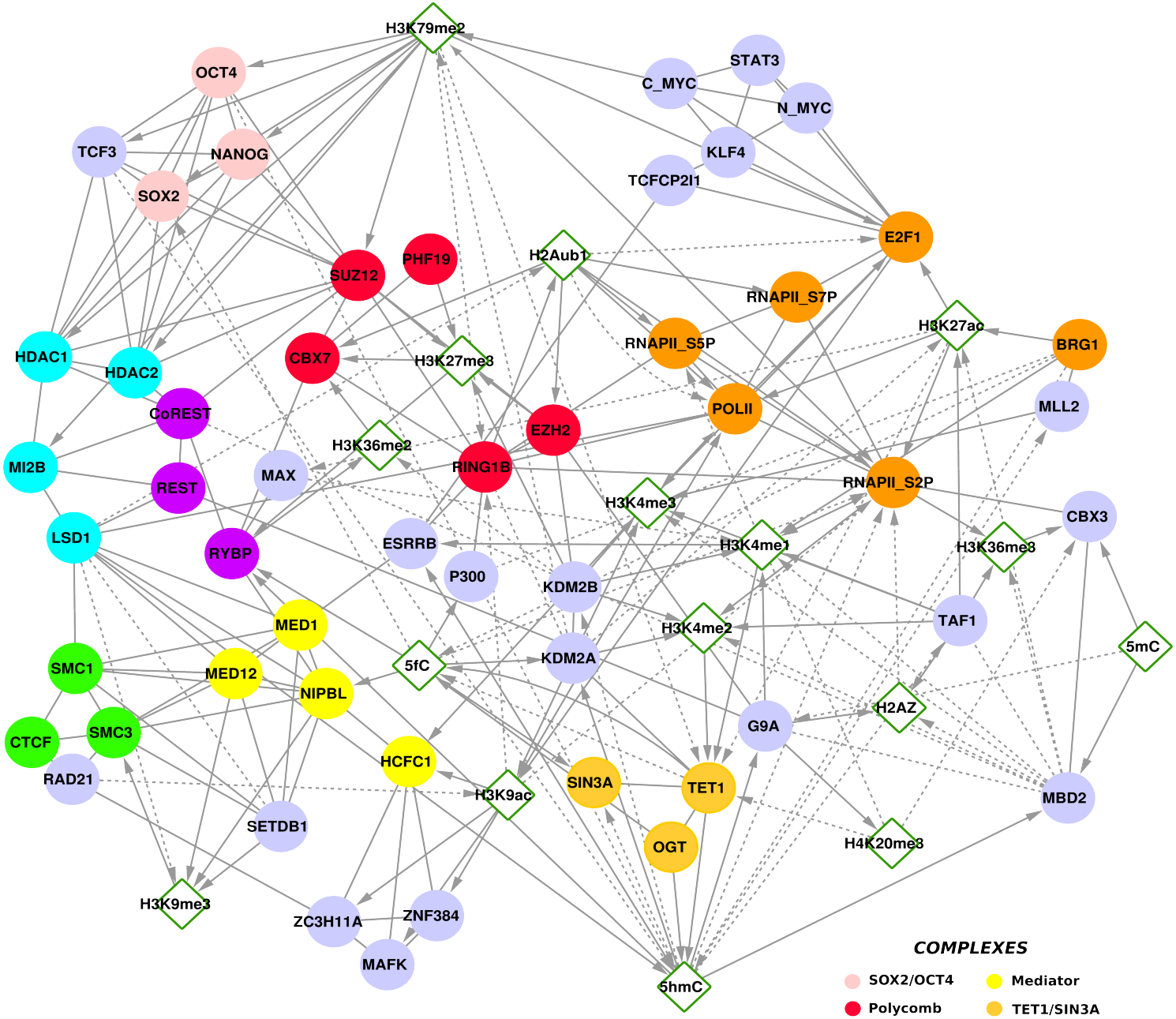
Sender
1.-Writer/Eraser from experiments
2.-Influences signal location
(eg. KOs)

Signal
Histone marks (H3K4me1, …).
Cytosine modifications
(5mC, 5hmC, 5fC).

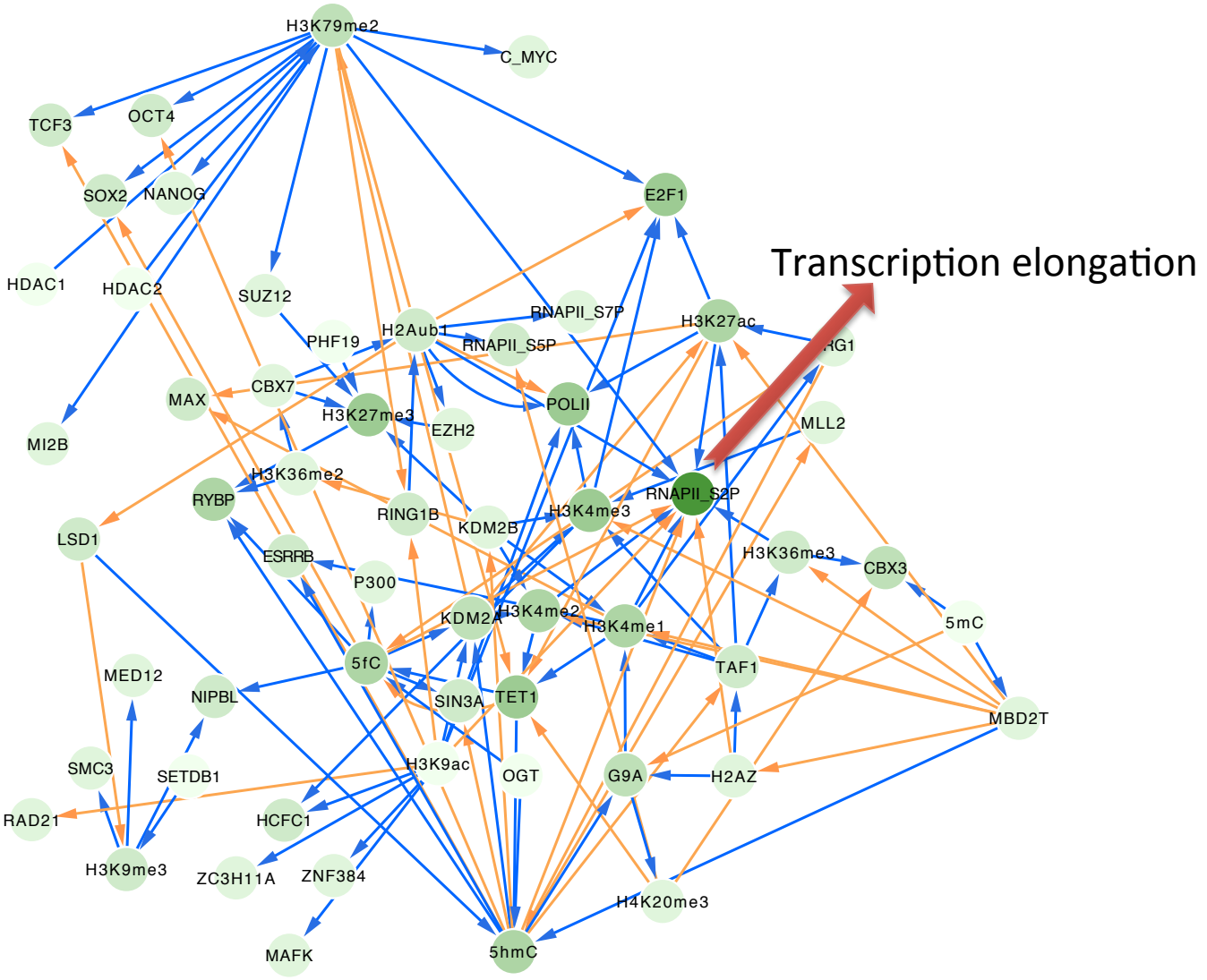
Receiver
Not-sender signal interactors

KNOWLEDGE FROM THE LITERATURE -> EDGE DIRECTIONALITY

124 (52.5%) directional interactions: 56 emitter / signal, 68 receiver/signal
18 CrP nodes act as emitters or receivers of different signals

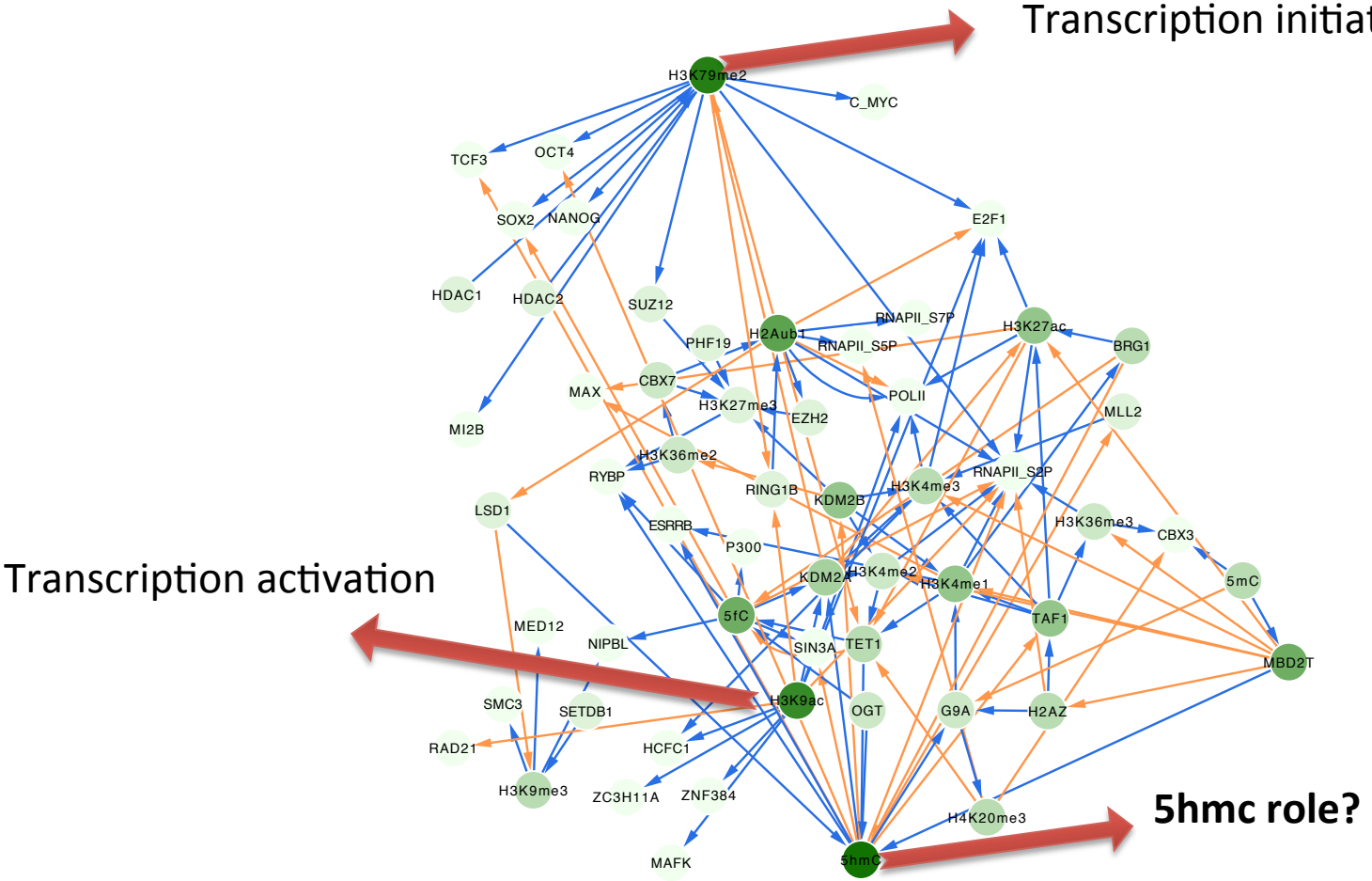

H3K79me2
OCT4
STAT3
C_MYC
N_MYC
TCF3
NANOG
KLF4
SOX2
TCFCP2I1
E2F1
SUZ12
PHF19
H2Aub1
RNAPII_S7P
H3K27ac
BRG1
HDAC1
HDAC2
CBX7
H3K27me3
RNAPII_S5P
CoREST
POLII
MLL2
EZH2
H3K36me2
MI2B
RING1B
REST
MAX
RNAPII_S2P
H3K4me3
CBX3
LSD1
RYBP
ESRRB
H3K4me1
H3K36me3
P300
KDM2B
MED1
H3K4me2
TAF1
5mC
SMC1
MED12
5fC
KDM2A
NIPBL
H2AZ
CTCF
SMC3
G9A
HCFC1
SIN3A
TET1
MBD2
RAD21
H3K9ac
SETDB1
OGT
H4K20me3
H3K9me3
ZC3H11A
ZNF384
MAFK
5hmC
COMPLEXES
SOX2/OCT4
Mediator
Polycomb
TET1/SIN3A
CoREST/REST
Mi-2/NuRD
POL II
Cohesin
No complex assign

Indegree (Transcription)

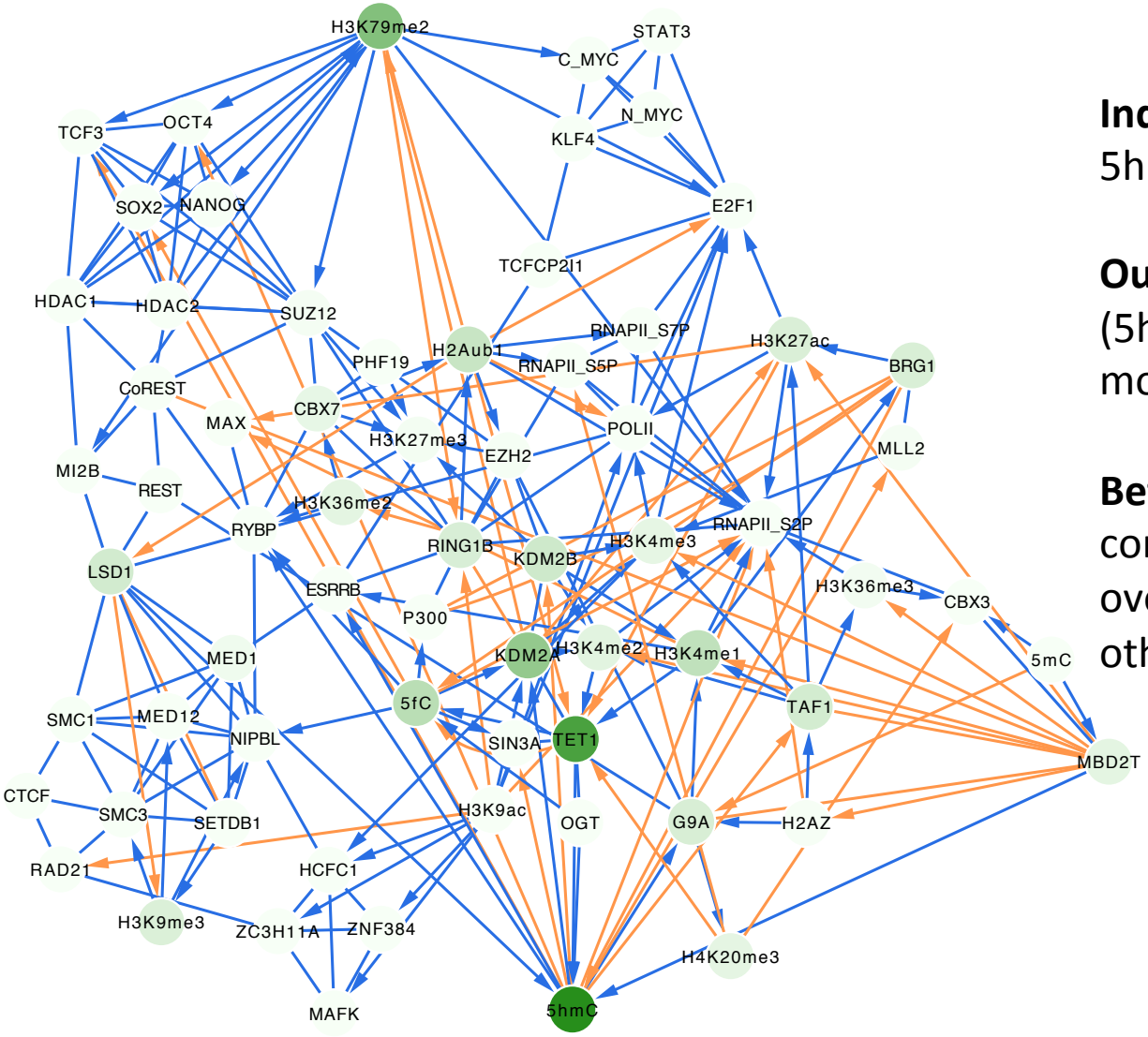


Outdegree

(5hmC is the message with more receivers)



Co-localization network



Indegree (Transcription, 5hmc)

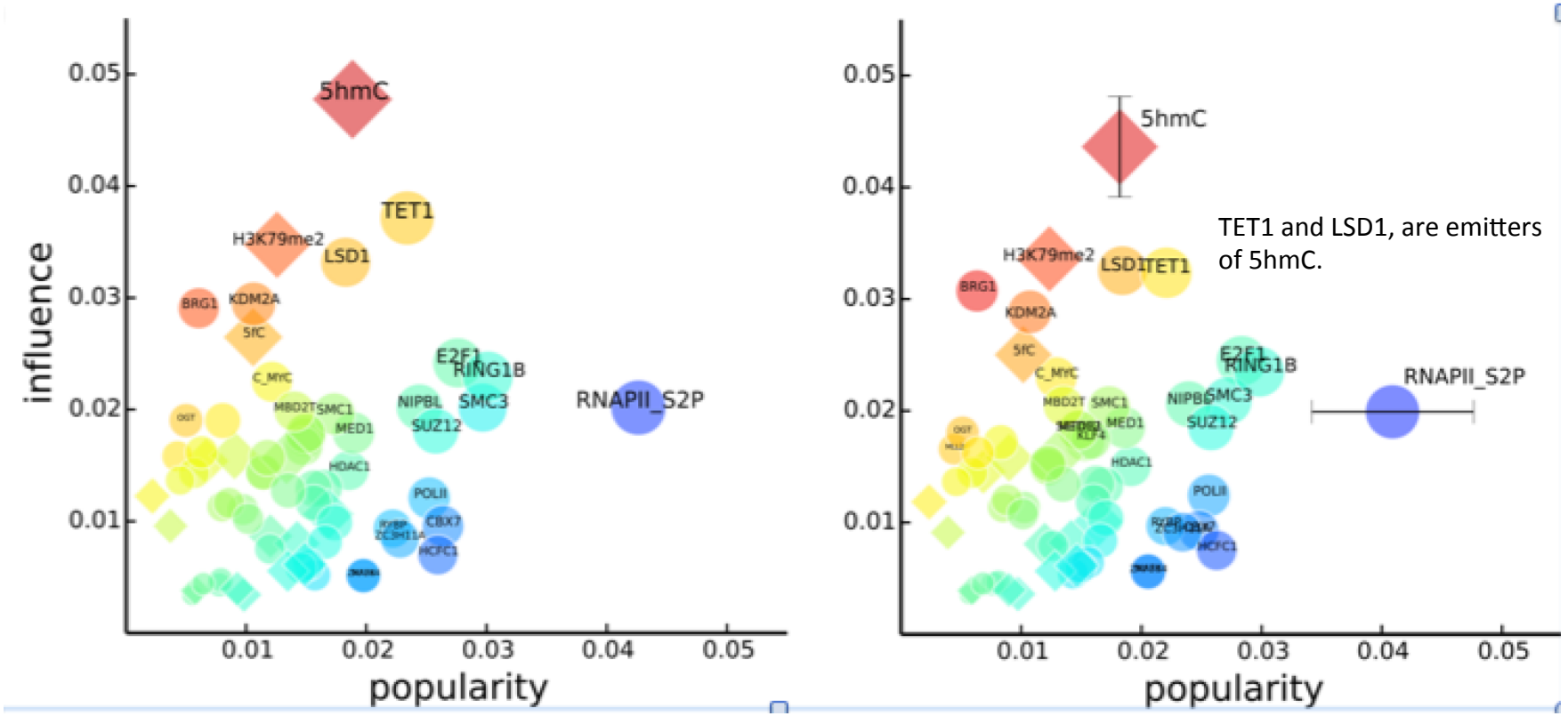
Outdegree
(5hmC is the message with more receivers)

Betweennes/Centrality
control that this node exerts over the interactions of other nodes in the network

$$C_b(n) = \sum_{s \neq n \neq t} (\sigma_{st}(n) / \sigma_{st})$$

5hmC is the most central node

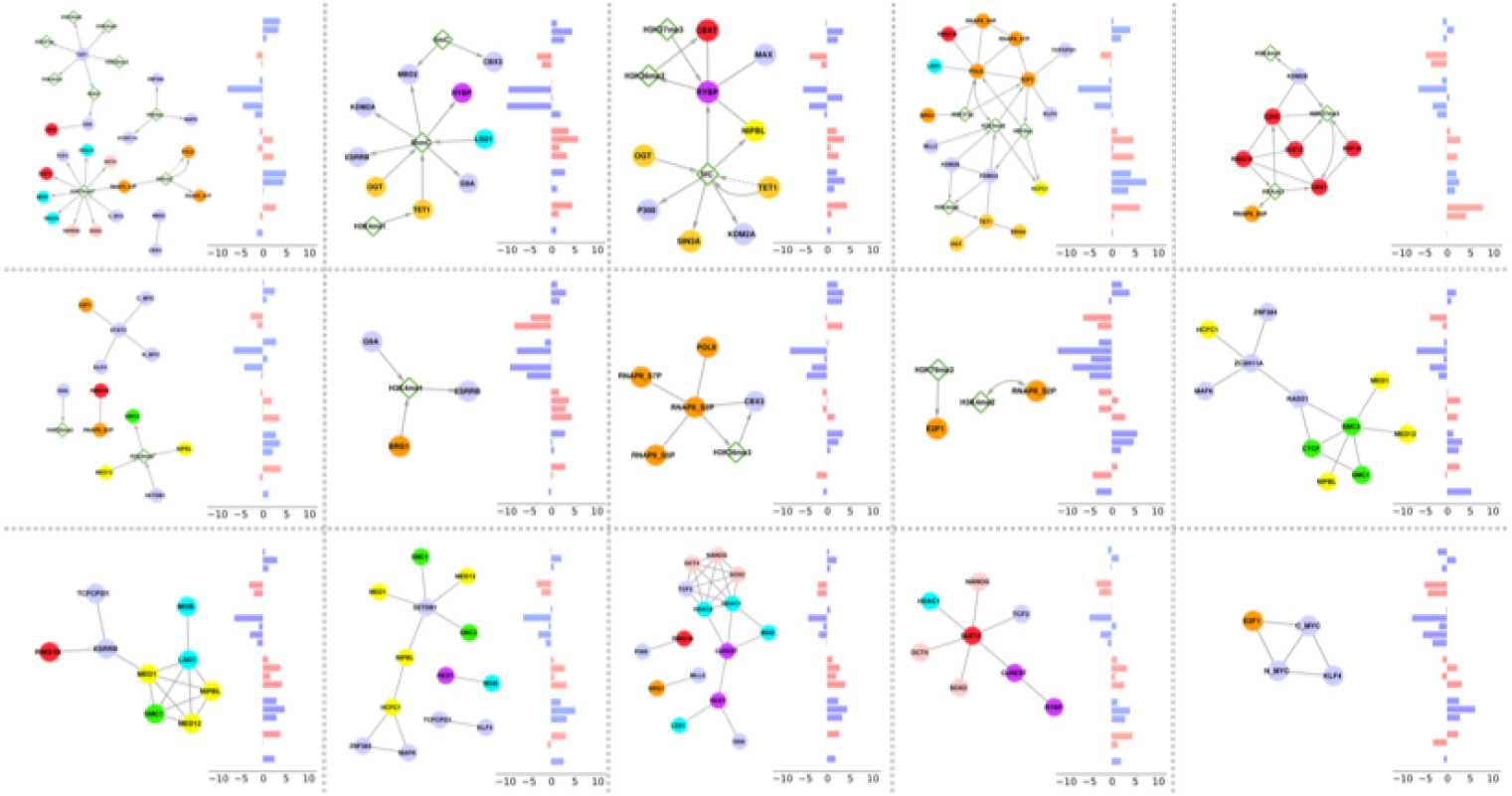
Influence/popularity



- **Influential nodes: information easily spreads out to the rest of the network,**
- **Popular nodes gather information from many regions of the network.**

Effect of directionality miss-assignments and random edges removal

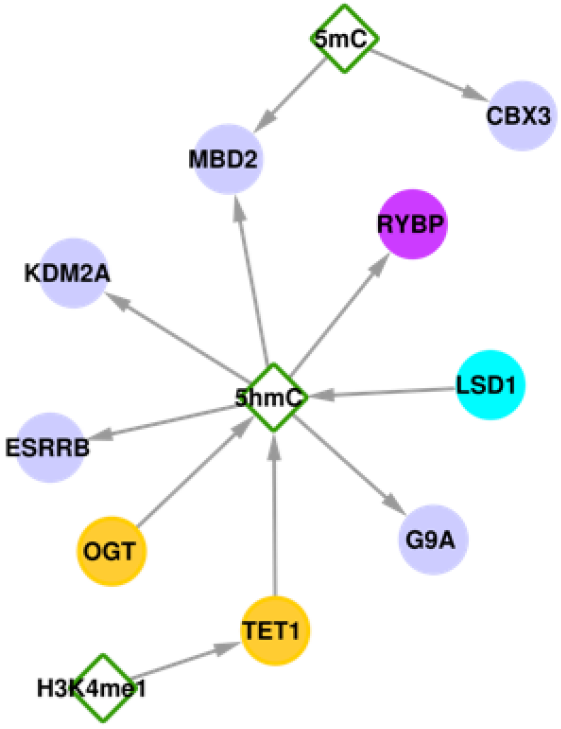
chromnets



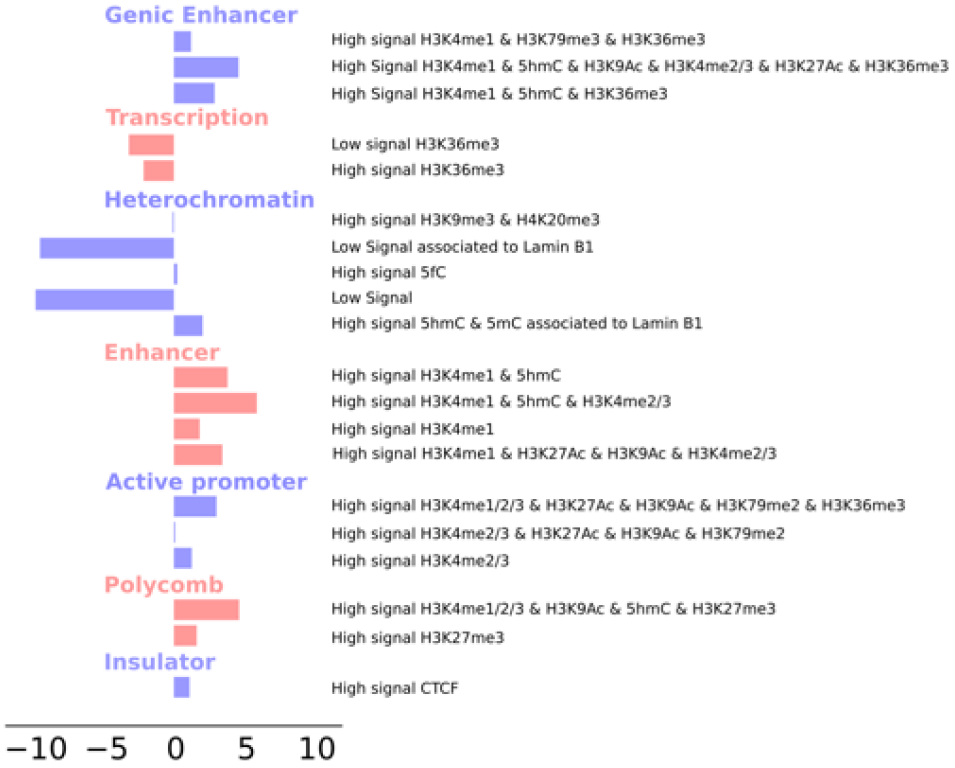
star-like chromnets: communication modules that connect different protein complexes (emitter/signal and signal/receiver interactions)

i.e. two central connectors (5fC and RYBP) connecting Polycomb, Mediator and TET1-SIN3A complexes >>> enriched in active transcription states and regulatory elements.

5hmC nucleates a star like sub-network



overall enrichment of the chromatin states



5hmC indirectly connects to H3K4me1 via TET1, and with 5mC via MBD2T.

Chromnet is enriched in regulatory elements.

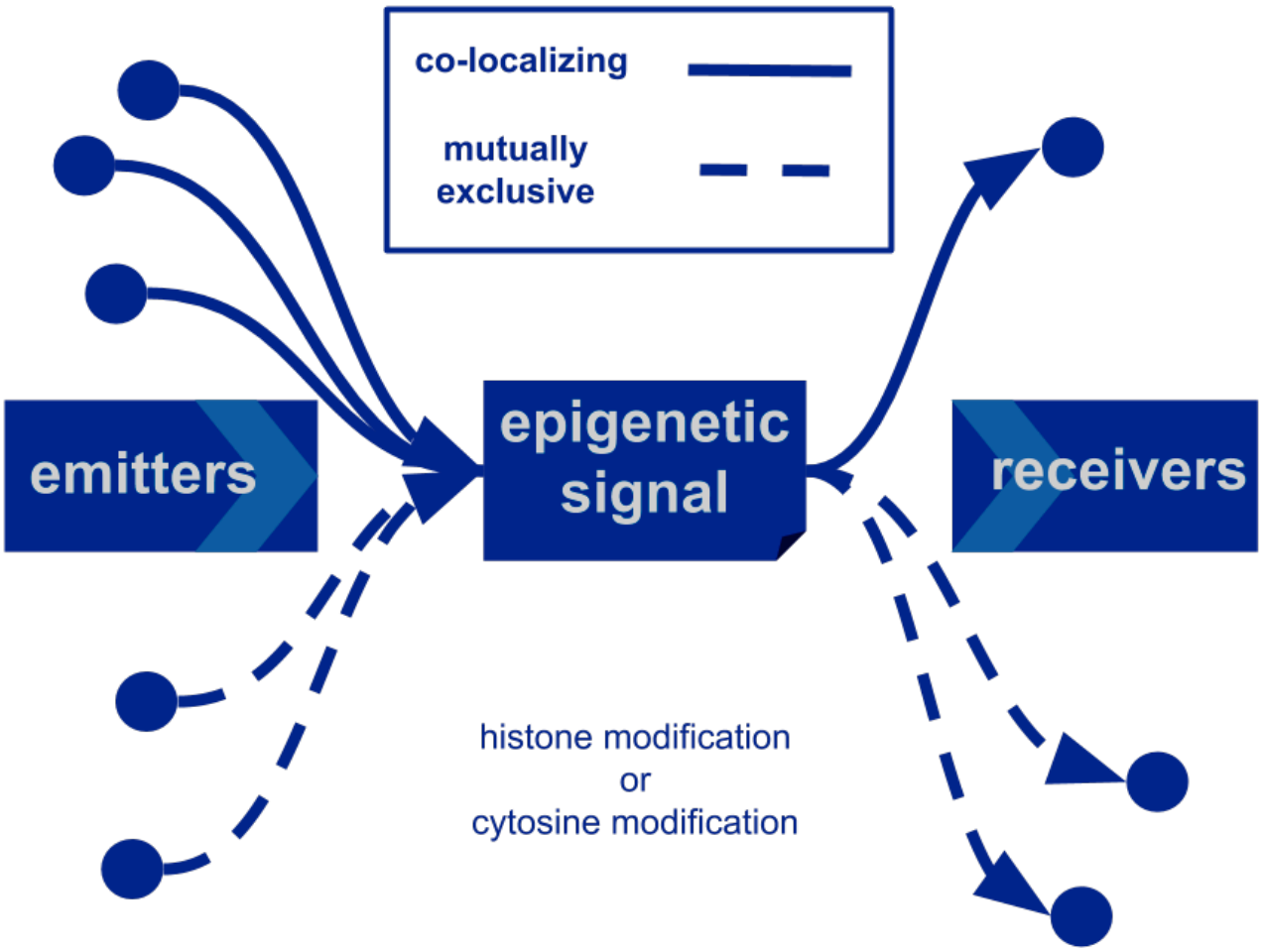
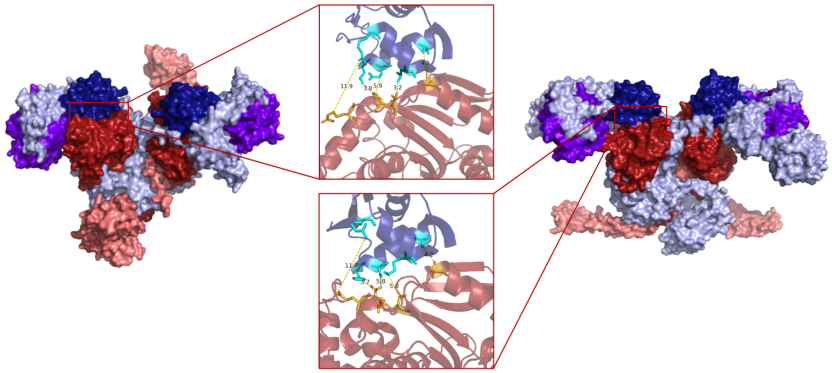

co-localizing
mutually exclusive
emitters
epigenetic signal
receivers
histone modification
or
cytosine modification

Figure 1

Co-evolution and evolutionary interactions



Darwin (1859), *On the origin of species*.

Anderson et al. American Journal of Botany. (2005)

Co-evolution:

- ***Reciprocal evolution between interacting species driven by natural selection (John N. Thompson).***
- ***Co-evolution in about biological objects speaking evolution with each other.***
- ***Evolution of communication requires co-adaptation of the agents and adaptation of the interaction as whole (Maynard Smith, J. & Harper, D.G.C. 2003)***

Mirrortree Method: finding interaction partners at genome level

Proce

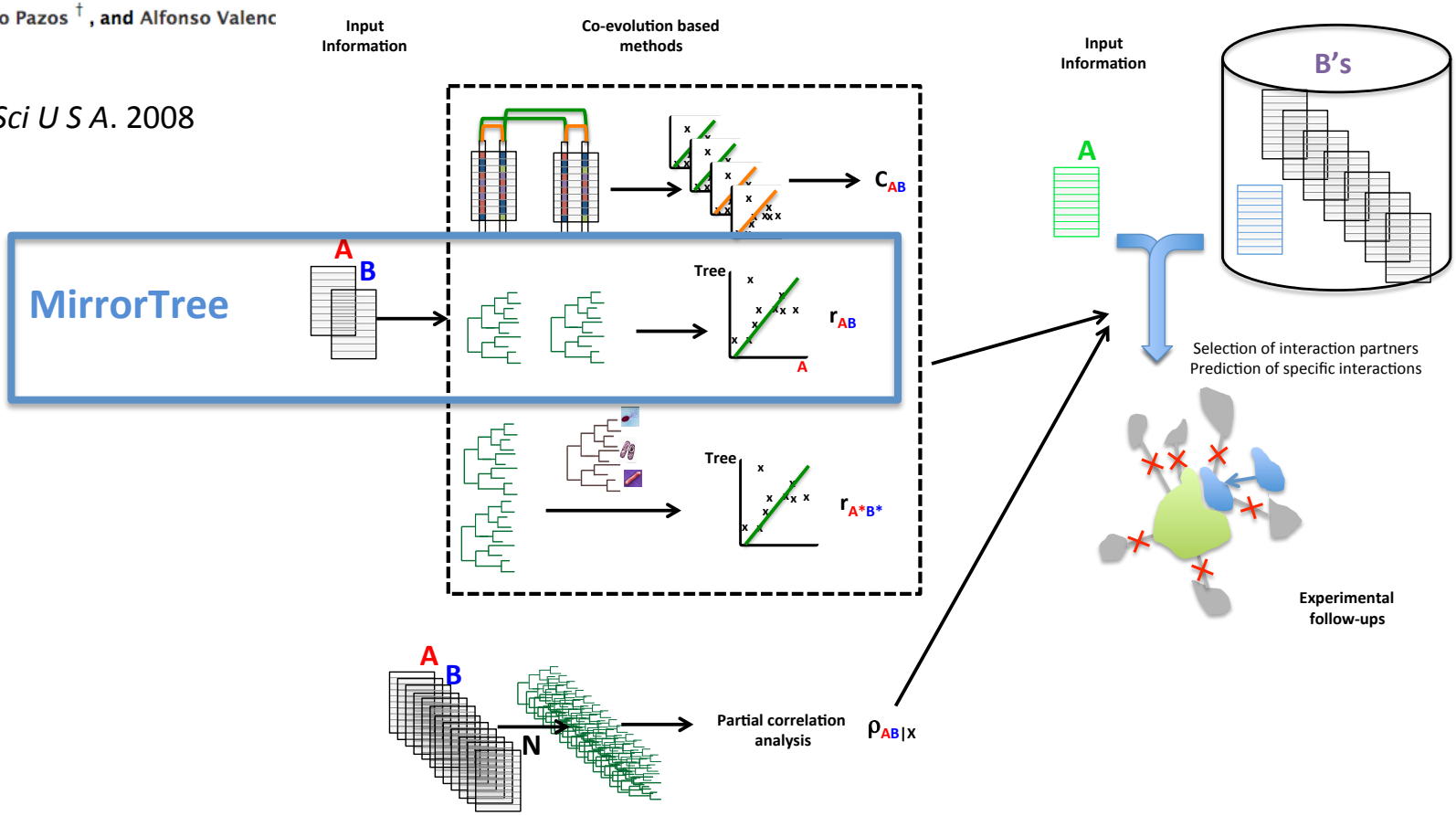
High-confidence prediction of global interactomes based on genome-wide coevolutionary networks

David Juan *, Florencio Pazos †, and Alfonso Valenc

+ Author Affiliations

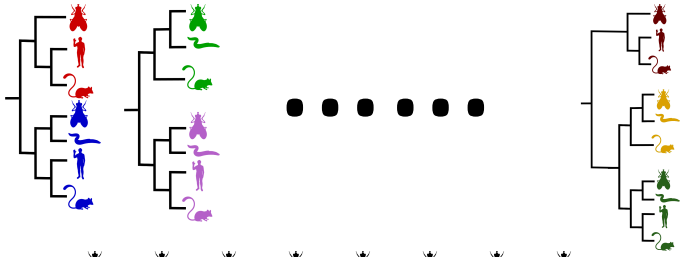
Proc Natl Acad Sci U S A. 2008

David Juan, Florencio Pazos
Nat. Rev Genet 2013



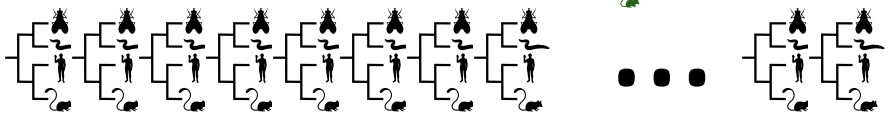
Detecting co-evolving CrPs

13,579 metazoan (88 species) trees from eggNOG (1)



(2,3)

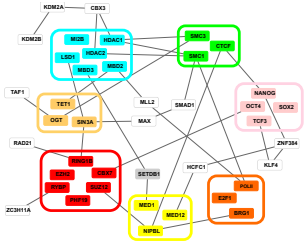
19,267 orthologs-only trees containing mouse proteins

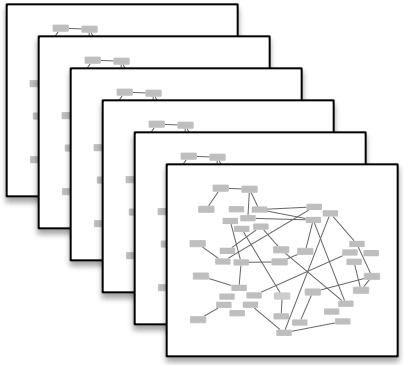


58 CrPs

10,000 x 58 random proteins
(background distribution)

p value < 0.05





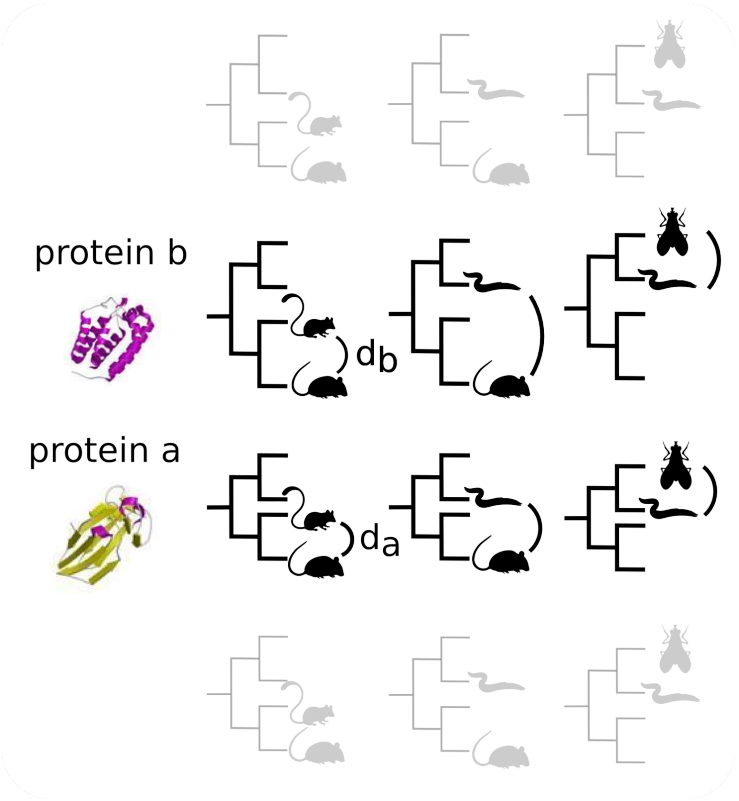
Maximum-entropy distribution in the space of species-species distance bins {d} for fixed single and pair protein frequencies

$$P(d) = \exp\left[\sum_a h_a(d_a) + \sum_{a,b} J_{a,b}(d_a, d_b)\right] \quad \textbf{(4)}$$

Jp,q regulates the interactions between proteins in the model.
A strong positive parameter can be interpreted as the direct symmetrical interaction between the two proteins a, b

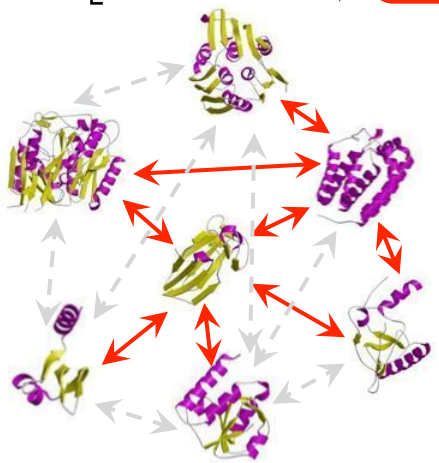
1.-Powell, S. *et al.Nucleic Acids Res.* **42,** D231–239 (2014).
2. Ruan, J. *et al. Nucleic Acids Res.* **36,** D735–D740 (2008).
3.-Juan, D., Rico, D., Marques-Bonet, T., Fernández-Capetillo, O. & Valencia, A.. *Biol. Open* **2,** 1402–1411 (2013).
4.-Schneidman E, *et al.* (2006). Nature 440: 1007–1012 Weak pairwise correlations imply strongly correlated network states in a neural population

Getting a direct co-evolutionary network



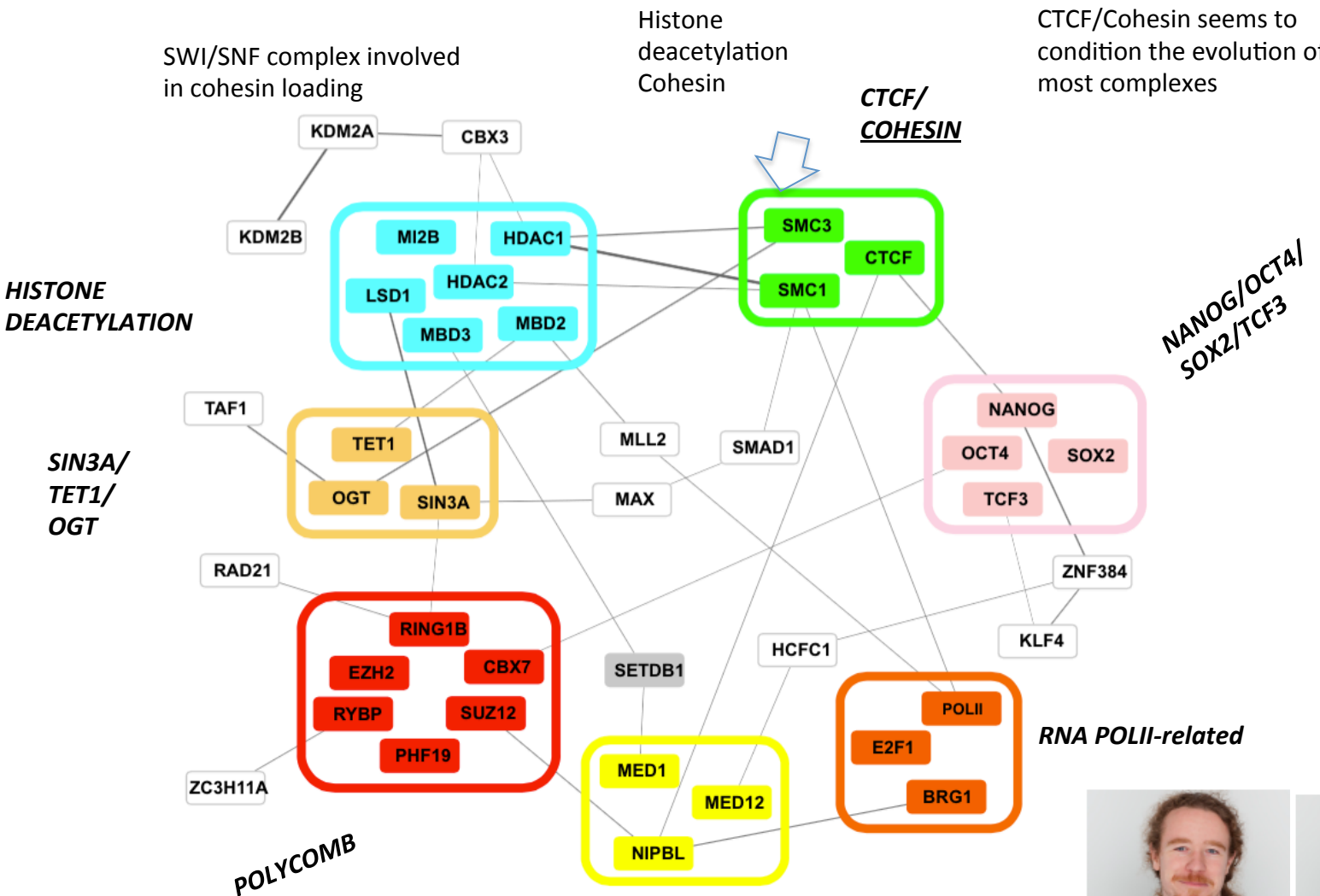
coevolutionary couplings network

$$P(d) = \exp\left[\sum_a h_a(d_a) + \sum_{a,b} J_{a,b}(d_a, d_b)\right]$$



Direct ⟷ indirect ⟷

34 Co-evolution based connections of the 58 CrPs



David Juan

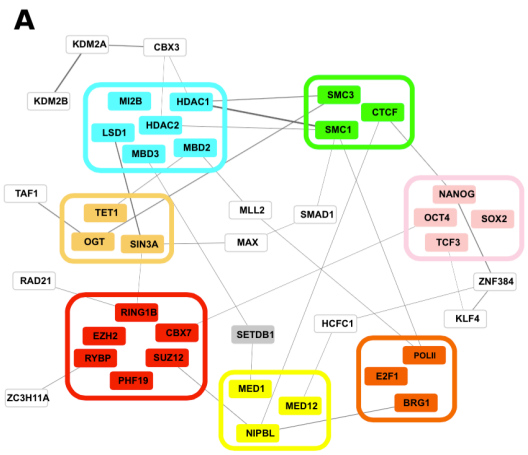
Simone Marsili

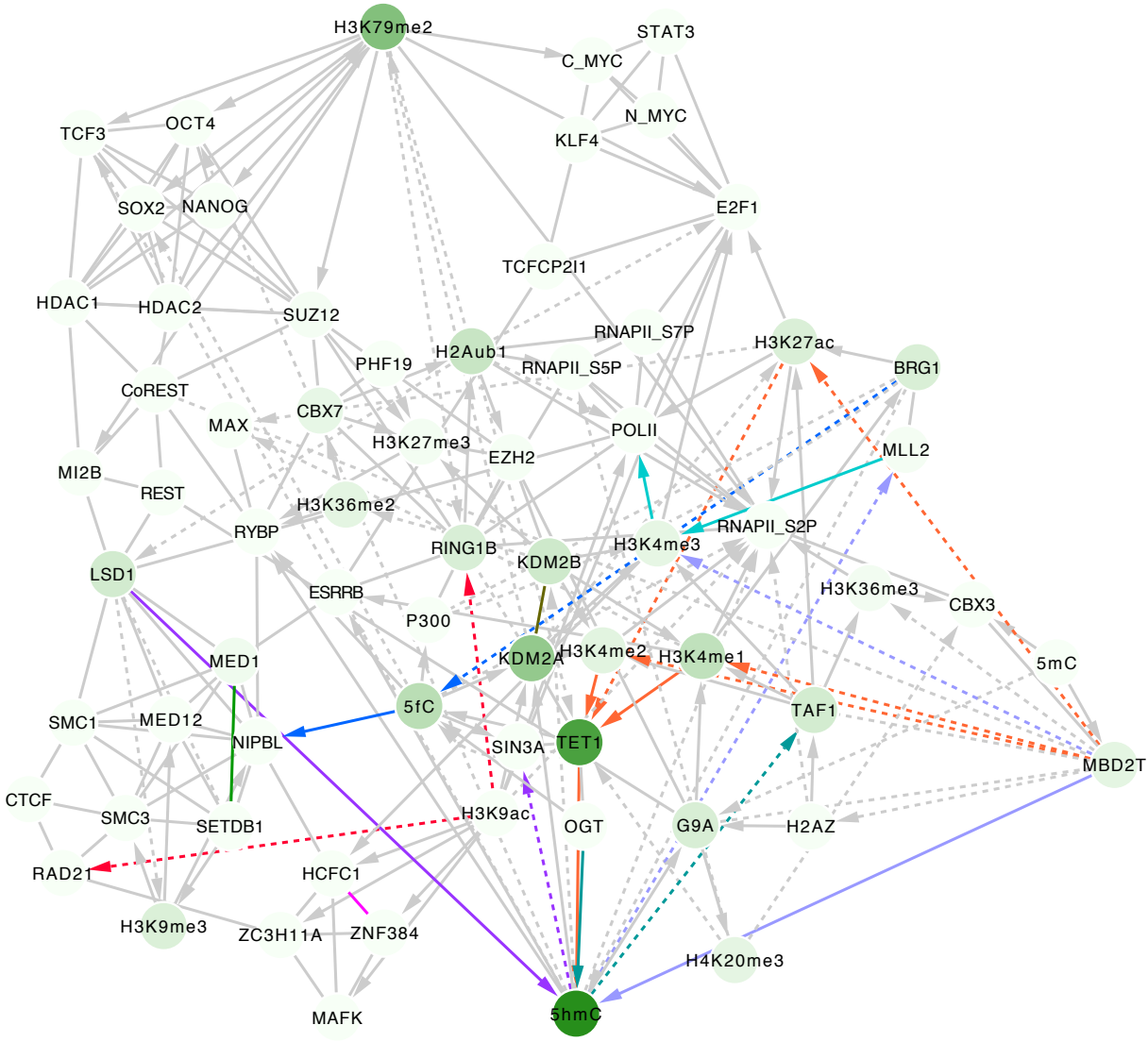
About metazoan co-evolution

(5hmC is key in communication and co-evolution)

32% of co-evolving pairs co-localize in mESCs

P < 1E-5

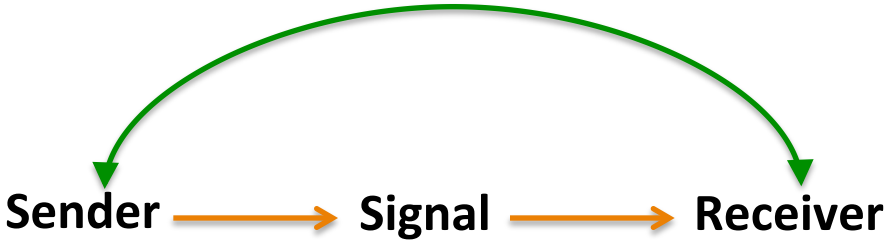




About metazoan co-evolution
(5hmC is key in communication and co-evolution)

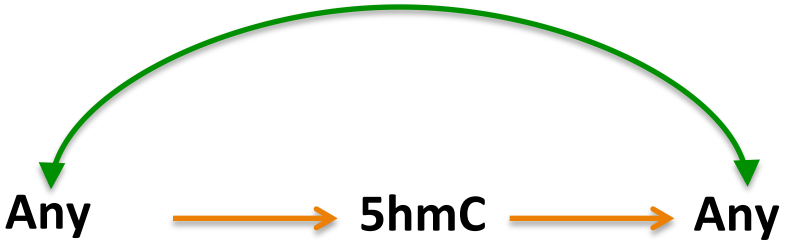
32% of co-evolving pairs co-localize in mESCs P < 1E-5

Co-evolution points to Sender-Receiver epigenetic communication



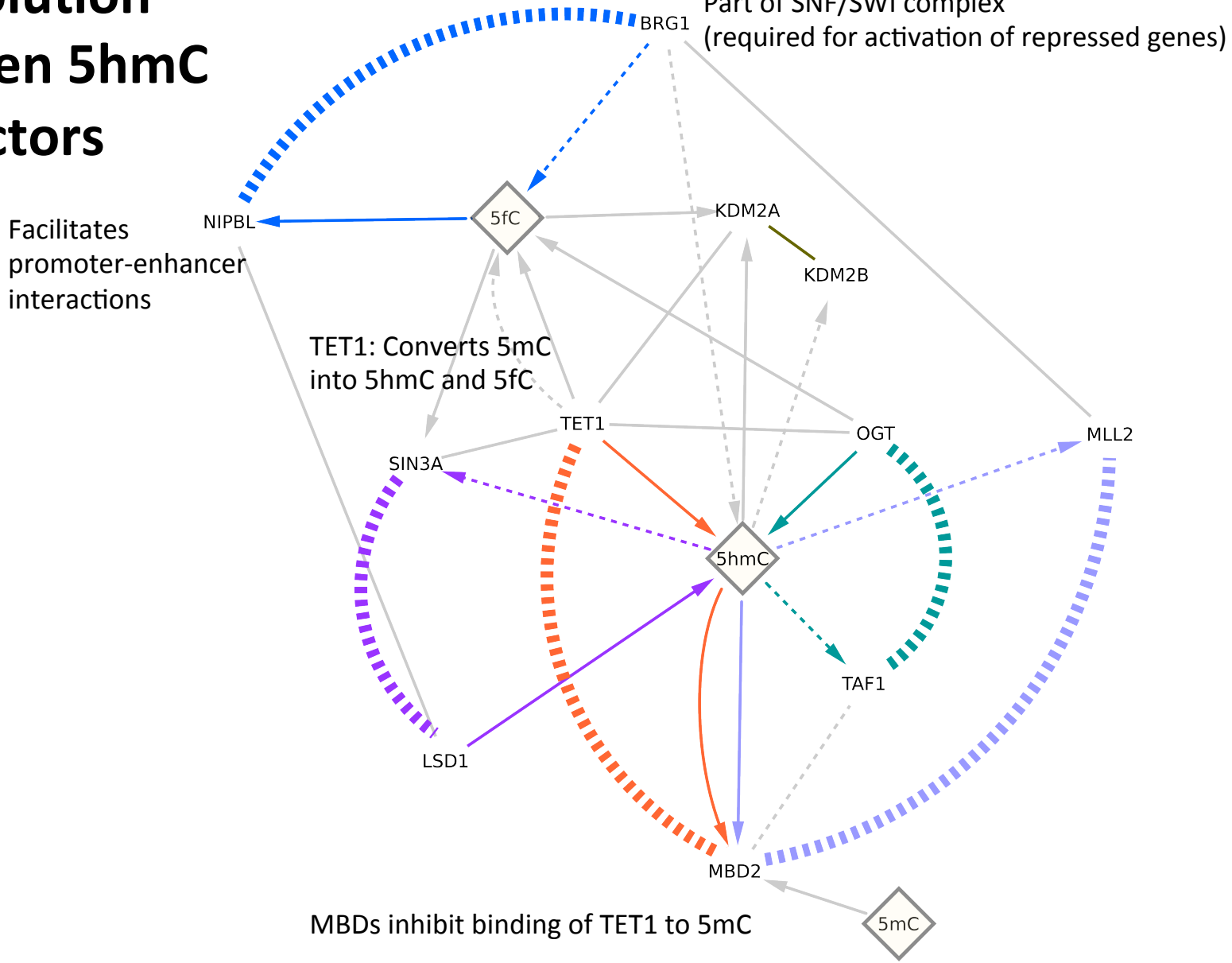
69% (background 22%) P < 1E-6

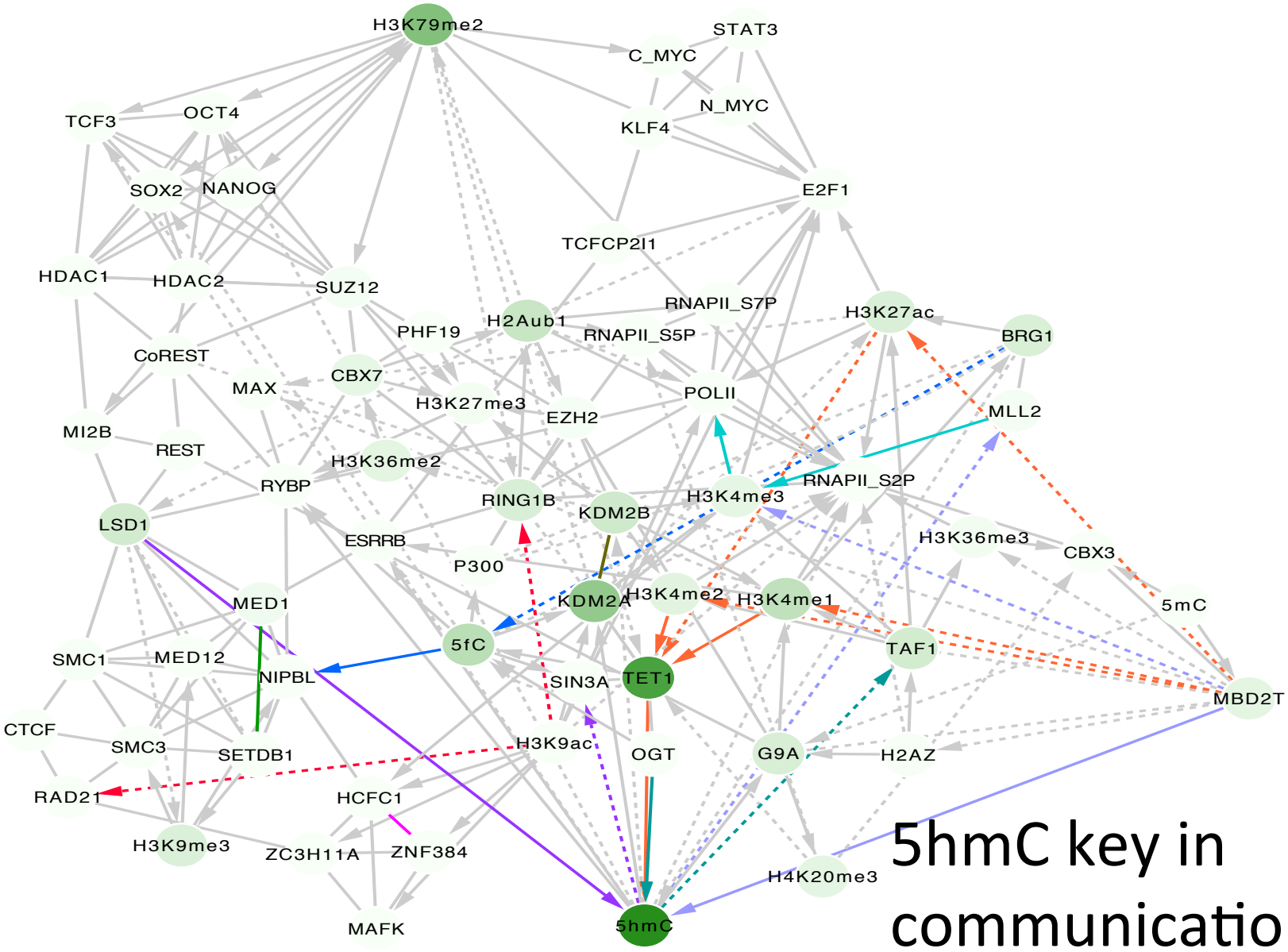
Co-evolution points to **5hmC**-mediated epigenetic communication



25% (background 7%) P < 1E-6

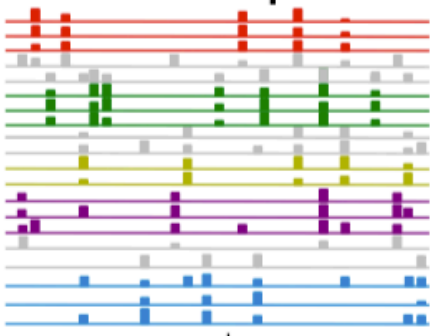
Co-evolution between 5hmC interactors





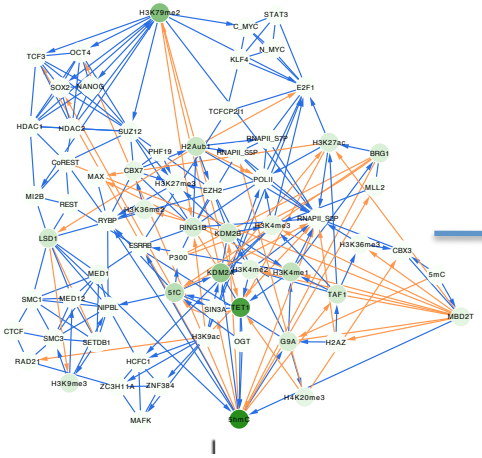
5hmC key in communication & co-evolution in mESC

ChIP-Seq data



Sparse partial correlation
+
Elastic Networks

Genome Wide localization Network



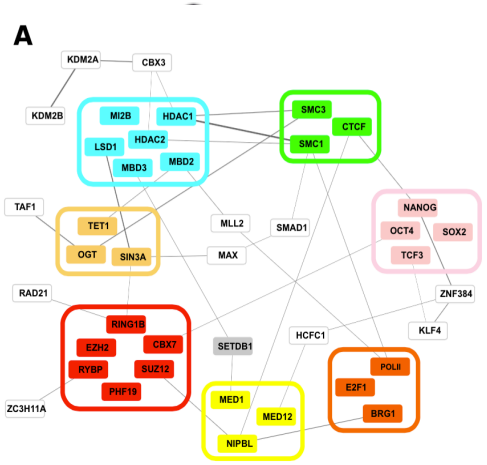
Co-evolution

Evolution of communication requires co-adaptation of the agents and adaptation of the interaction as whole (Maynard Smith, J. & Harper, D.G.C. 2003)

Phylogenetic trees

Pairwise maximum-entropy model based on tree similarities

Co-evolution Network



Sender → Message → Receiver

mESC Epigenetic Network (part 1)

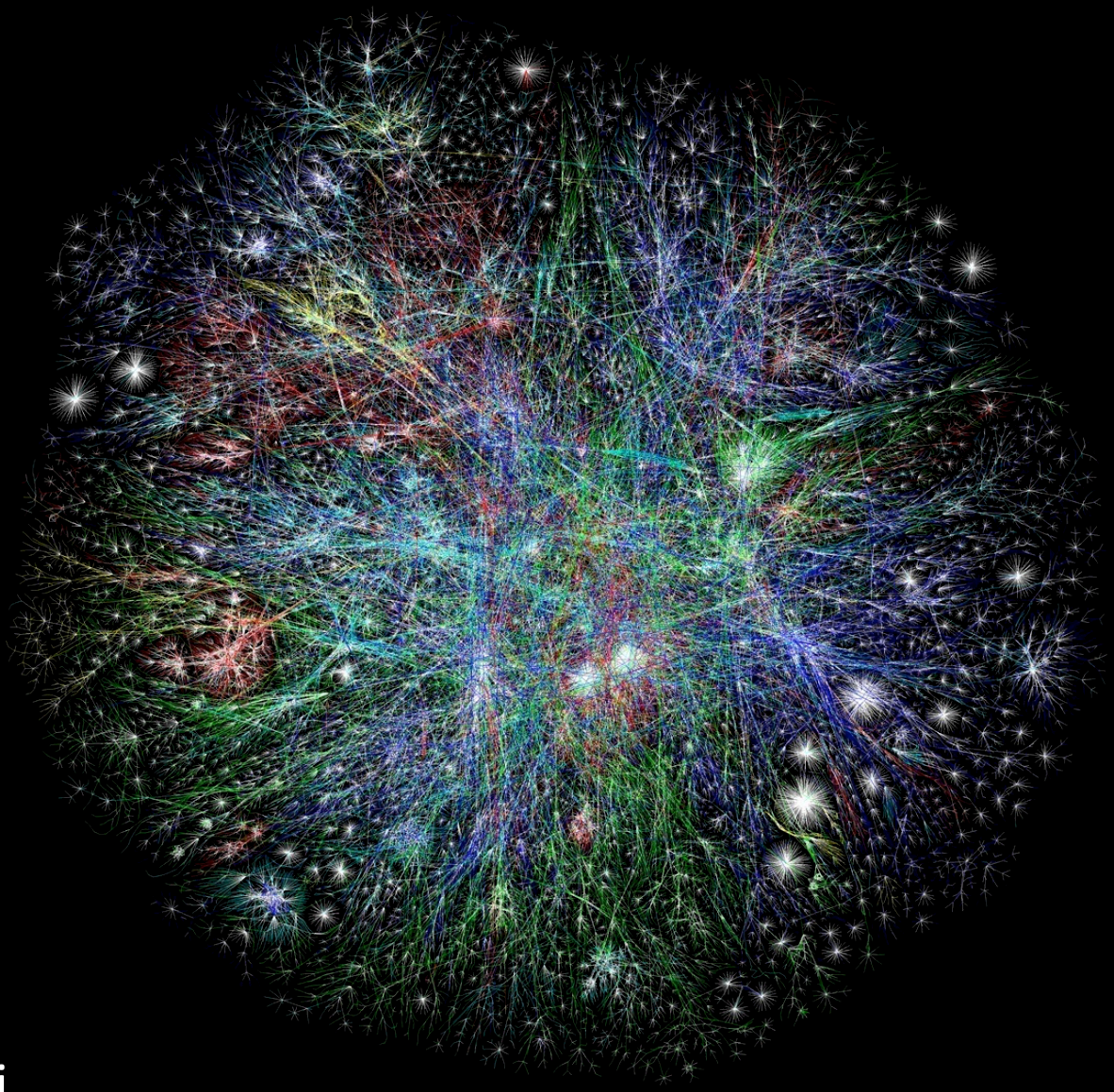
Chip-Seq data – co-localization – biological complexes

Chromatin related proteins biological functions and chormatin states

5hmC, stem cells and development

Co-Evolution of Chromatin related molecular complexes

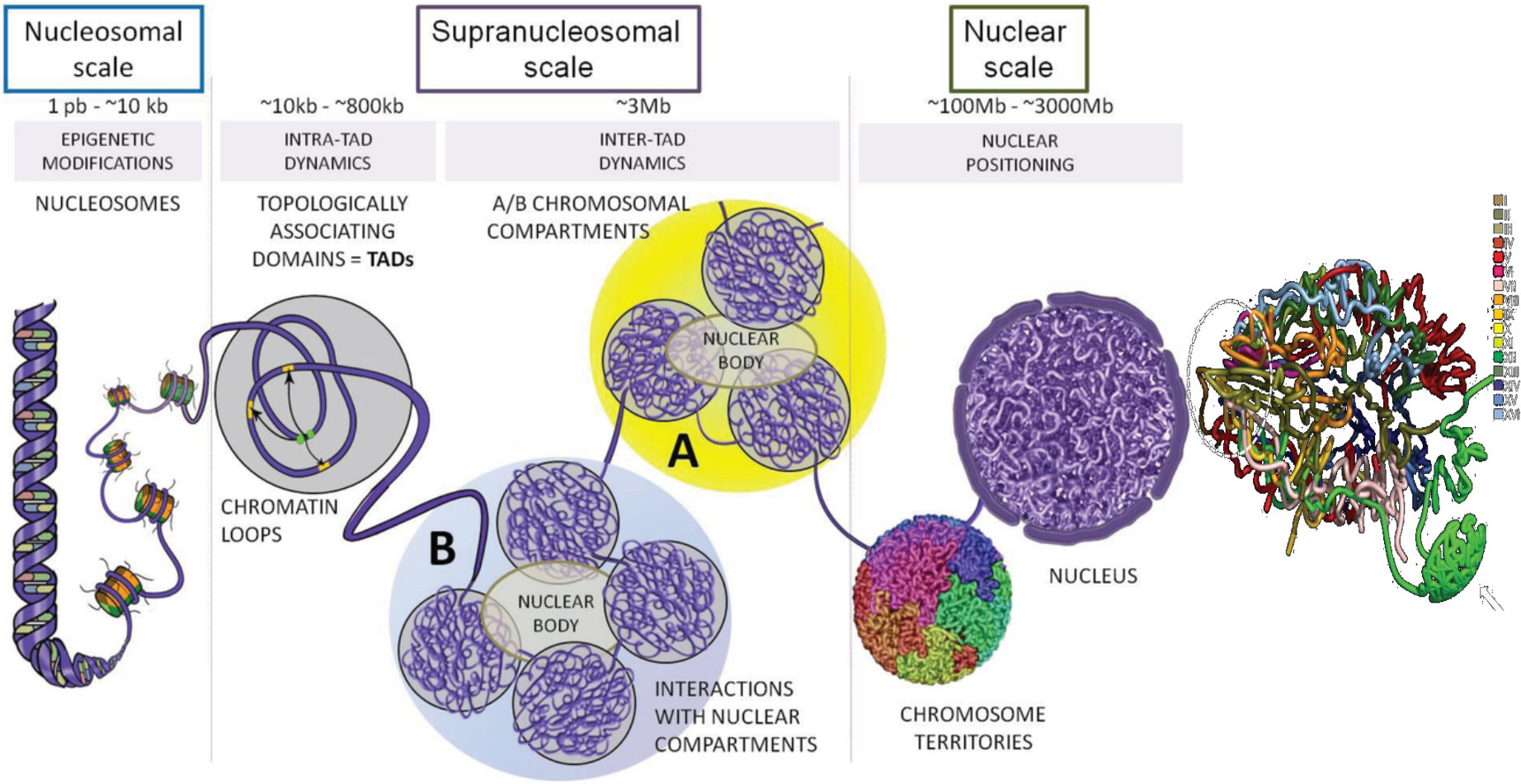
A network approach integrates 3D contacts with epigenomic data (part 2)



Vera Pancaldi
Structural and Computational Biology Group
Spanish National Cancer Research Centre (CNIO)

17 May 2016
EMBL-EBI, Hinxton

The 3D structure of chromatin

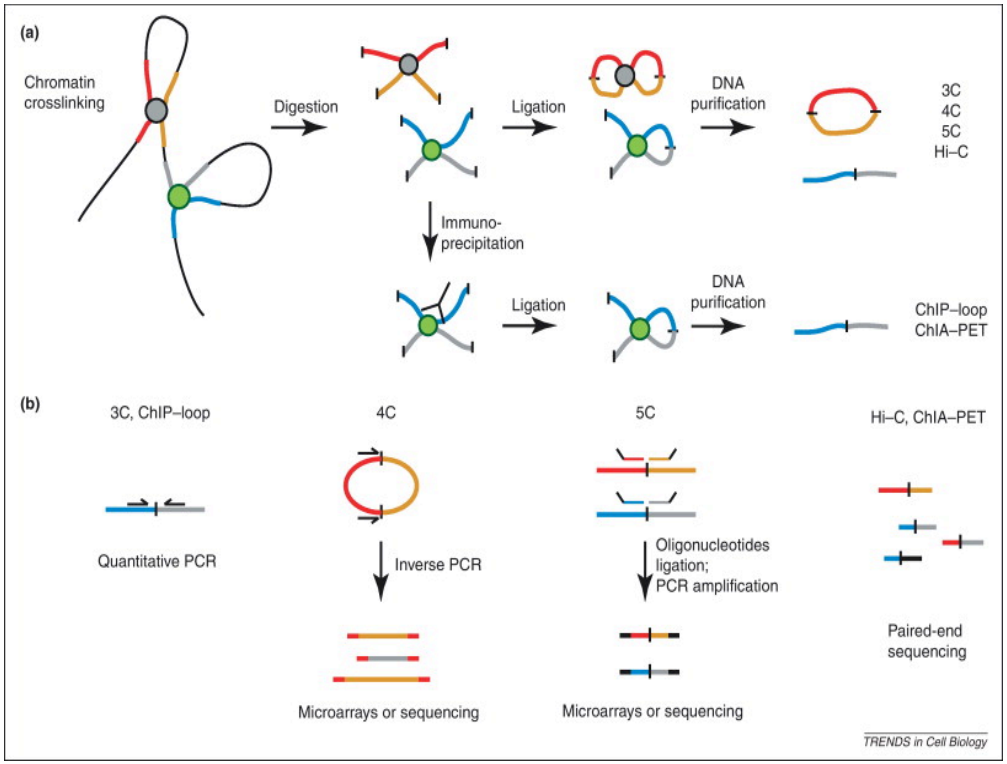


Ea et al. **Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization**. Genes 2015

Z. Duan, ... W Stafford Noble, **A three-dimensional model of the yeast genome**, Nature 2010

Chromatin Interaction Networks (CINs)

Unravelling 3D chromatin contacts
Chromosome Conformation Capture methods

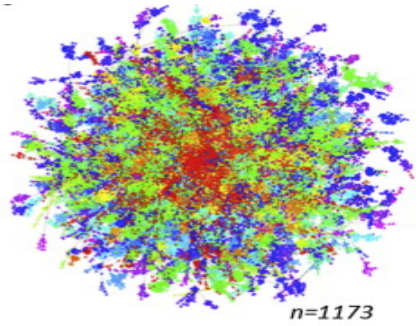


CHiA-PET
Genome-wide
Requires pull-down of specific protein (enriches for interactions mediated by it)

3C, 4C, 5C
Local
no pull-downs

HiC Genome wide
No pull-downs

Construct networks where:
Node = chromatin fragment Edges=contacts in 3D



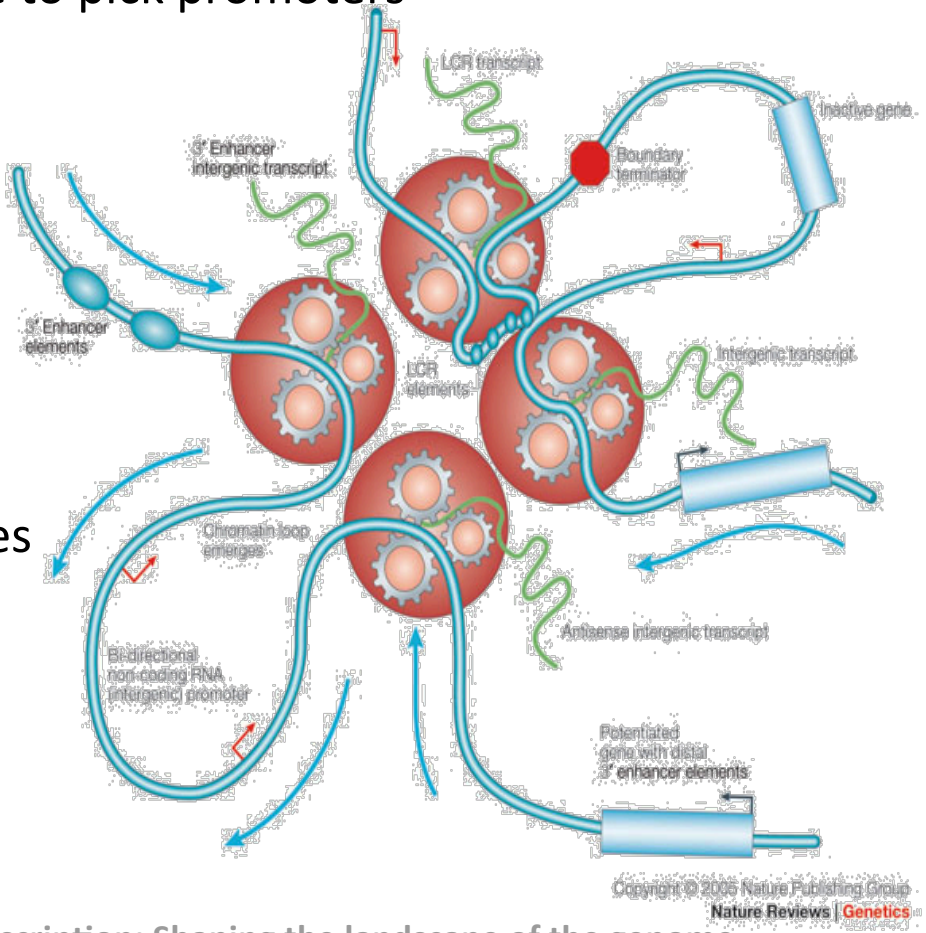
Montavon and Denis Duboule, **Landscapes and archipelagos: spatial organization of gene regulation in vertebrates,** Trends in Bio 2012; KS Sandhu, Y Ruan, **Large-scale functional organization of long-range chromatin interaction networks.** Cell Rep 2012

What about genes? PCHi-C!

Problem so far: HiC contact maps dominated by interactions far from genes.
Need very high coverage to pick promoters

Solution: Promoter-Capture HiC (PCHiC)
Add promoter capture step
Obtain promoter-centred contact maps
(No pull-downs, genome-wide)

It allows to look for transcription factories
(Multiple genes transcribed together)



Lyubomira Chakalova[1], ... Peter Fraser **Replication and transcription: Shaping the landscape of the genome** Nat Rev Gen 2005; Schoenfelder, S. *et al*. **The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements**. Genome Res. 2015.

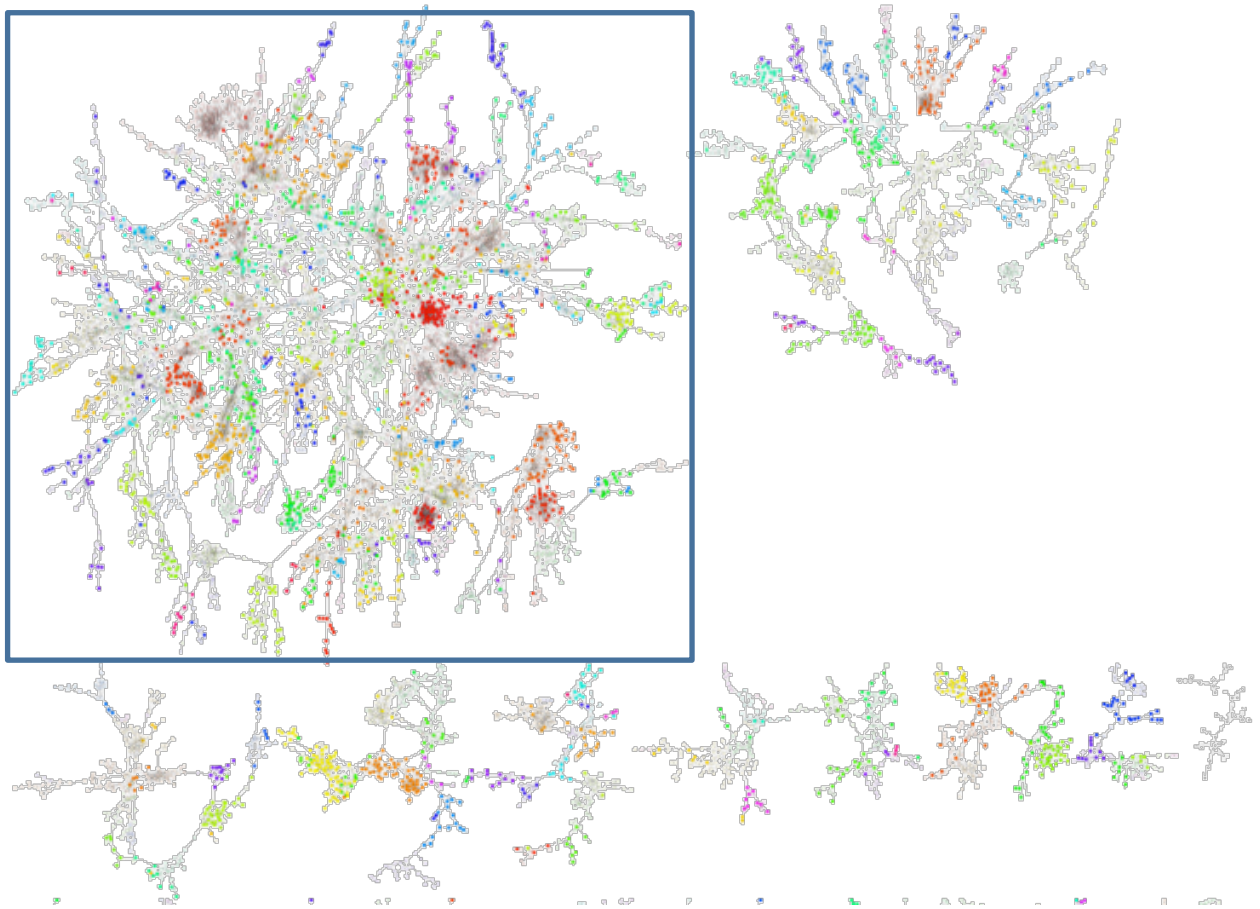
PCHi-C networks in mESCs

Interactions involving at least 1 promoter

Statistics:
Chromatin fragments (4Kb)
55,845 nodes
(19425 promoters)
69,987 edges
1 major connected component

20,523
Edges connecting
Two promoters

Coloured by modules
(highly interconnected
Portions of the network)



Data processing with CHiCAGO
(Mikhail Spivakov)

PCHi-C networks in mESCs

Interactions involving at least 1 promoter

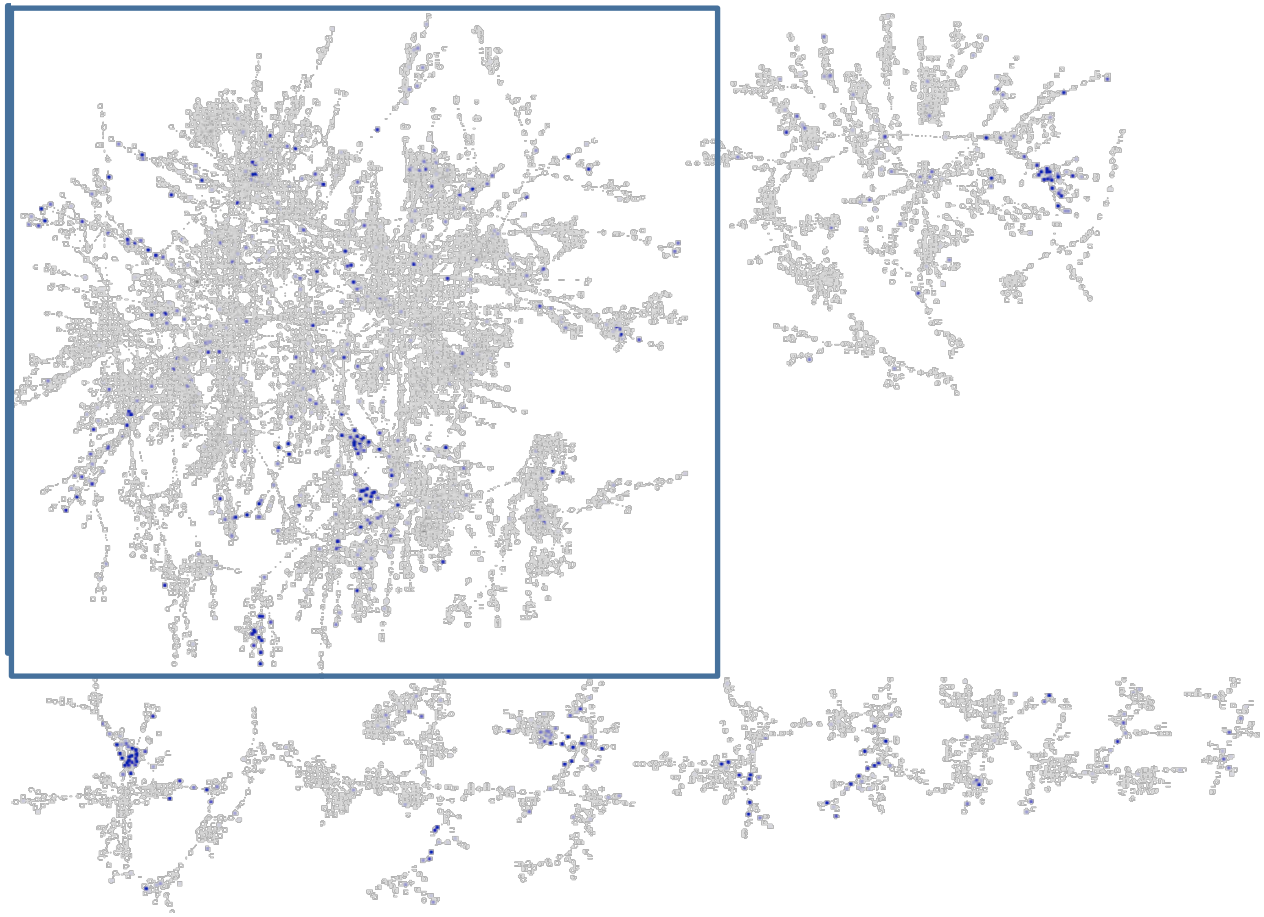
Statistics:
Chromatin fragments (4Kb)
55,845 nodes
(19425 promoters)
69,987 edges
1 major connected component

20,523
Edges connecting
Two promoters

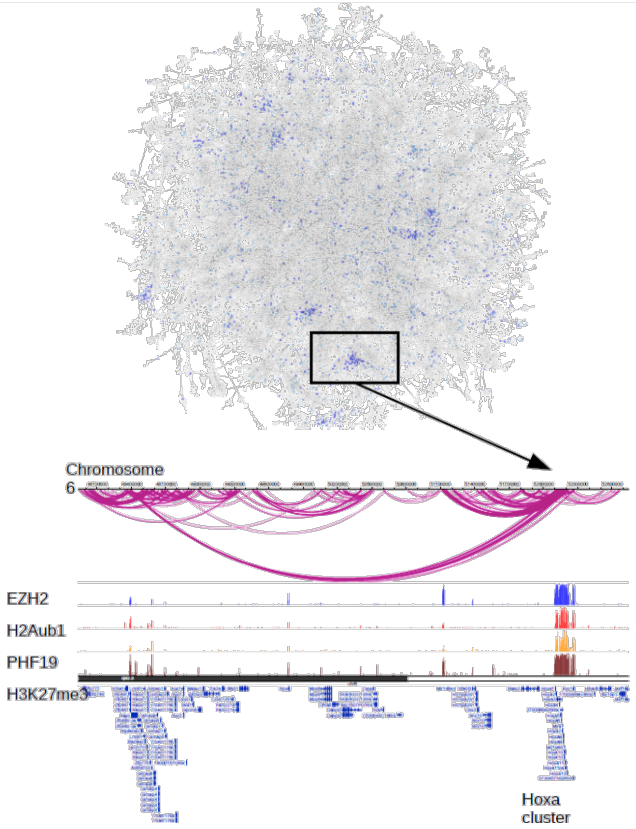
Coloured by modules
(highly interconnected
Portions of the network)

Data processing with CHiCAGO
(Mikhail Spivakov)

EZH2 Peaks in chromatin fragments



Chromatin Assortativity (ChAs)



Approach:

For each chromatin fragment (node):

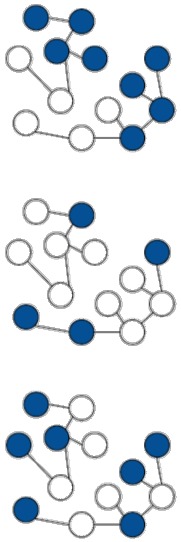
Calculate abundance of epigenetic feature

Eg: Proportion of fragment that has peak of EZH2 ChIP-seq

Average of proportion throughout the network (rare 0.02)

For each epigenetic feature, calculate Assortativity on the network

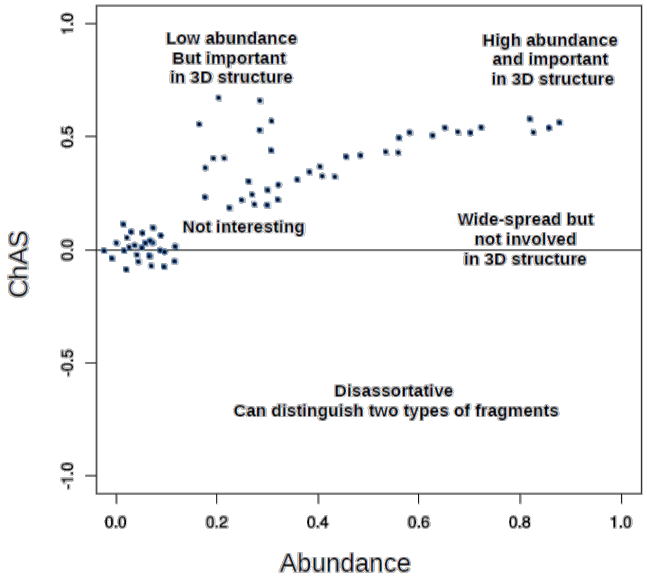
Eg: How much more likely are fragments covered by EZH2 peaks to interact with each other? (high 0.33)



High ChAS > 0

No ChAS = 0

ChAS < 0



Data:

PCHiC networks in mouse Embryonic Stems Cells (mESCs)
(Collaboration with Peter Fraser, Babraham Institute)

78 epigenetic features (3 cytosine modification, 13 histone modifications, chromatin related proteins binding peaks)

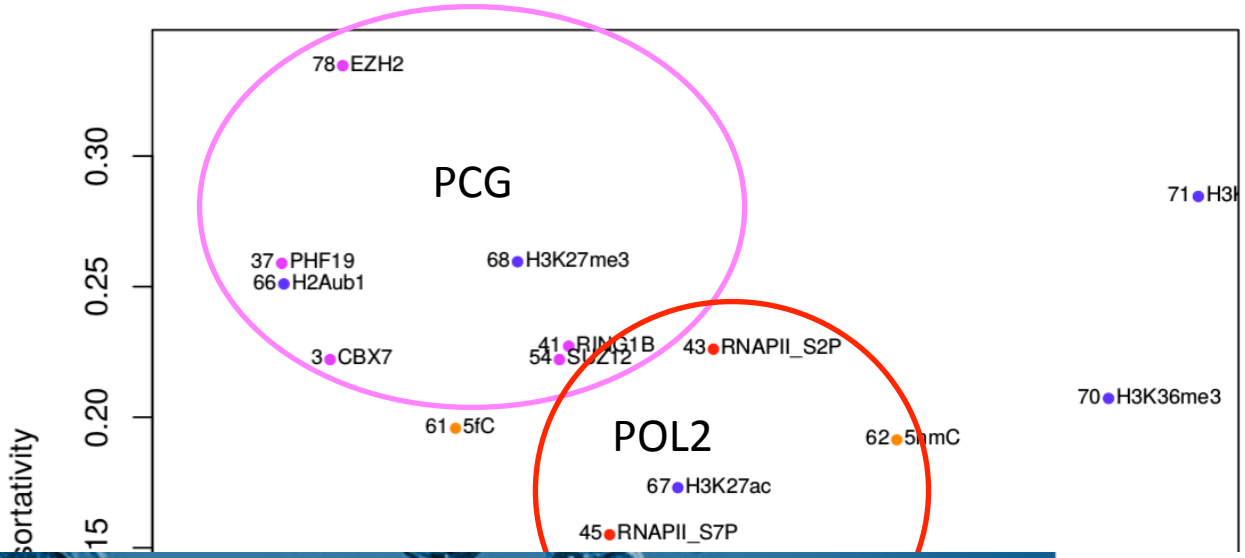
Epigenomic Co-localization and Co-evolution Reveal a Key Role for 5hmC as a Communication Hub in the Chromatin Network of ESCs, Juan et al. Cell Reports 2016

PCG is highly assortative in PCHi-C networks

PolyCombGroup
(PCG)
Proteins and marks
EZH2
RING1b
SUZ12
PHF19

H3K27me3
H2Aub1

RNA Polymerase 2



Polycomb repre
constrains the

Stefan Schoenfelder, Robe
Armstrong, Borbala Mifsuc
Mayra Furlan-Magaril, Ann
Kristina Tabbada, Simon A
Haruhiko Koseki, Peter Fra

Affiliations | Contributions

Nature Genetics (2015) | d
Received 10 December 201

Cell Stem Cell

All Content
Cell Stem Cell

Explore | Online Now | Current Issue | Archive | Journal Information | For Authors

< Previous Article

Volume 17, Issue 6, p748–757, 3 December 2015

Short Article

Switch to Standard View

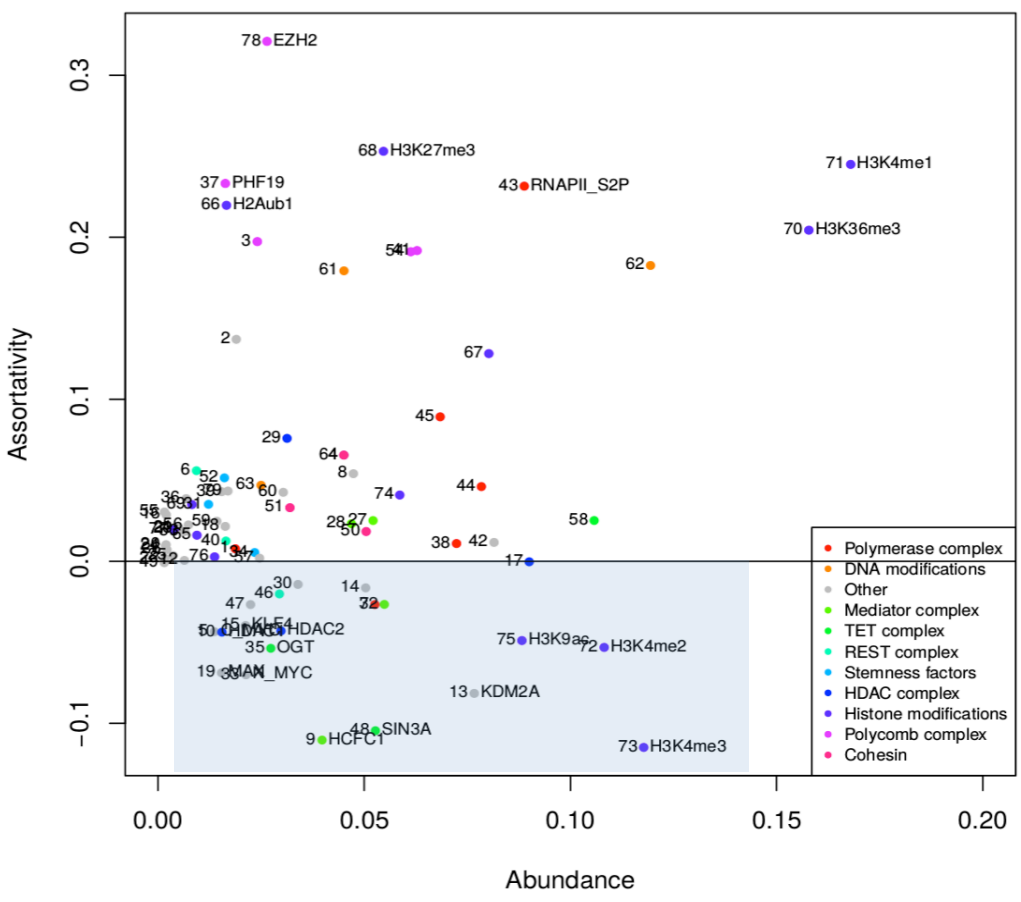
Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency

Onkar Joshi[5], Shuang-Yin Wang[5], Tatyana Kuznetsova, Yaser Atlasi, Tianran Peng, Pierre J. Fabre, Ehsan Habibi, Jani Shaik , Sadia Saeed[6], Lusy Handoko[7], Todd Richmond, Mikhail Spivakov, Daniel Burgess, Hendrik G. Stunnenberg

Chromatin Assortativity in different subnetworks

In P-O subnetwork features that are only on promoter fragments have ChAs <0

HCFC1 (transcription activator complex),
SIN3A (trancriptional repressor complex), KDM2A (H3K26 demethylase),
NMYC, OGT (histone acetyl transferase. complex), H3K4me2



Comparing P-P and P-O

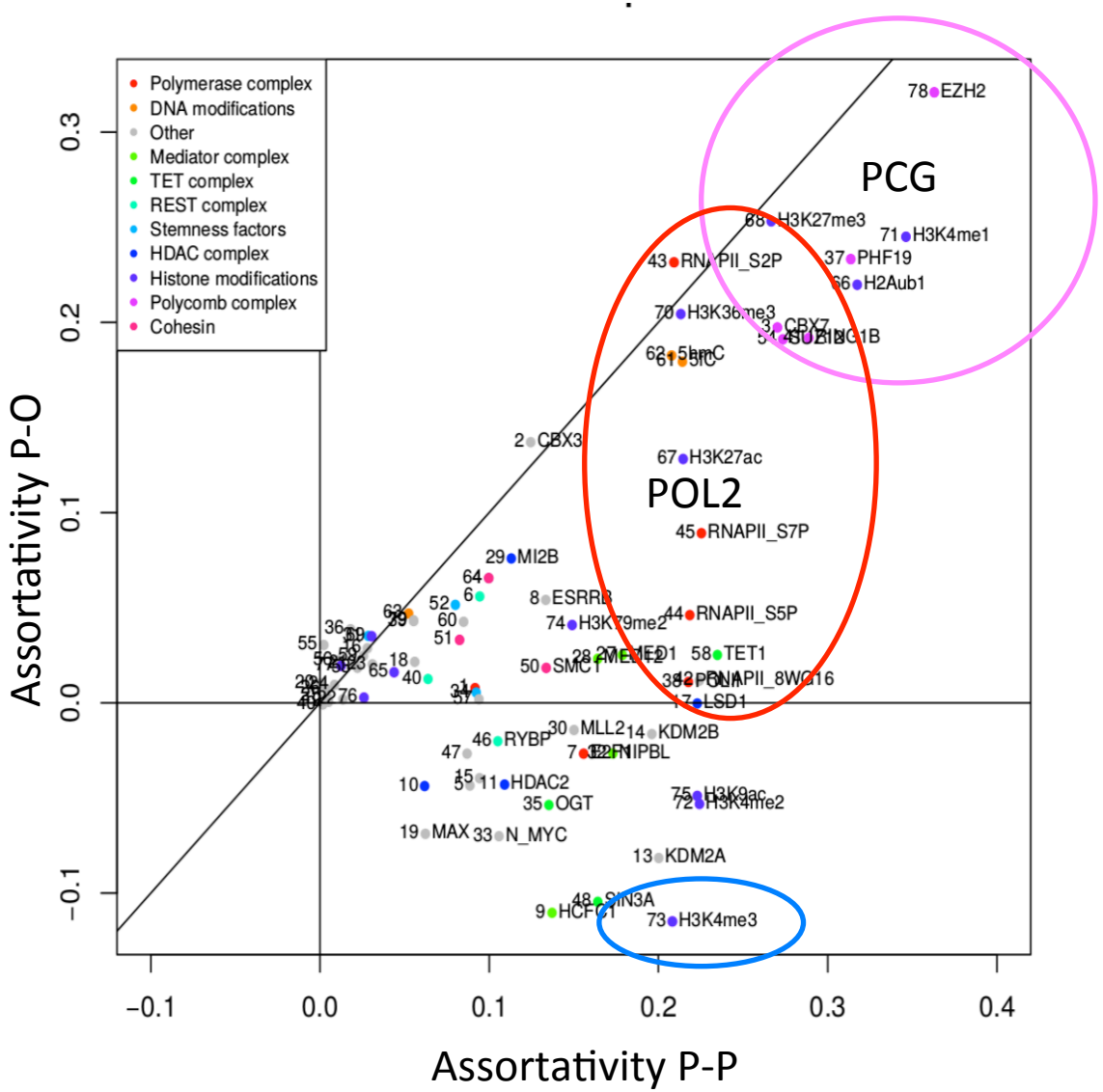
Identify features that have different assortativities in P-P and P-O contacts

PCG on diagonal
Similar ChAs>0 in P-P and P-O
Equal importance

RNAPII:
Variable ChAs in P-O,
ChAs>0 in P-P

H3K4me3: Mark associated to active promoters
ChAs >0 in P-P
ChAs<0 in P-O
(only present in promoters)

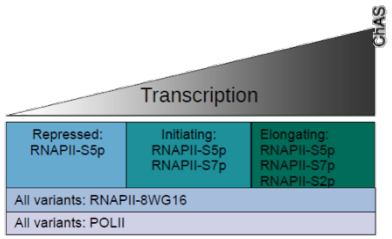
Fragments that have this mark are more likely to interact
Preferential contacts of active gene promoters.

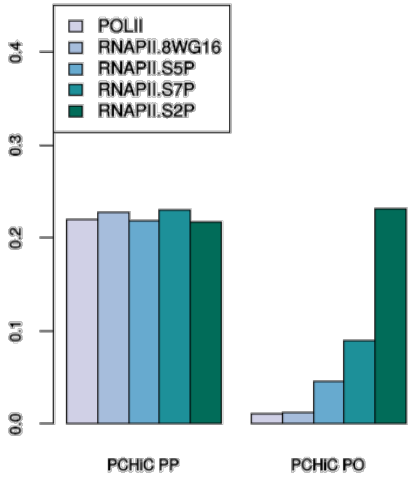


Assortativity of RNA Polymerase 2

5 Different RNAPII features

Binding peaks for different RNAPII variants





ChAs of RNAPII in P-O variable

Non-elongating RNAPII has low ChAs in P-P

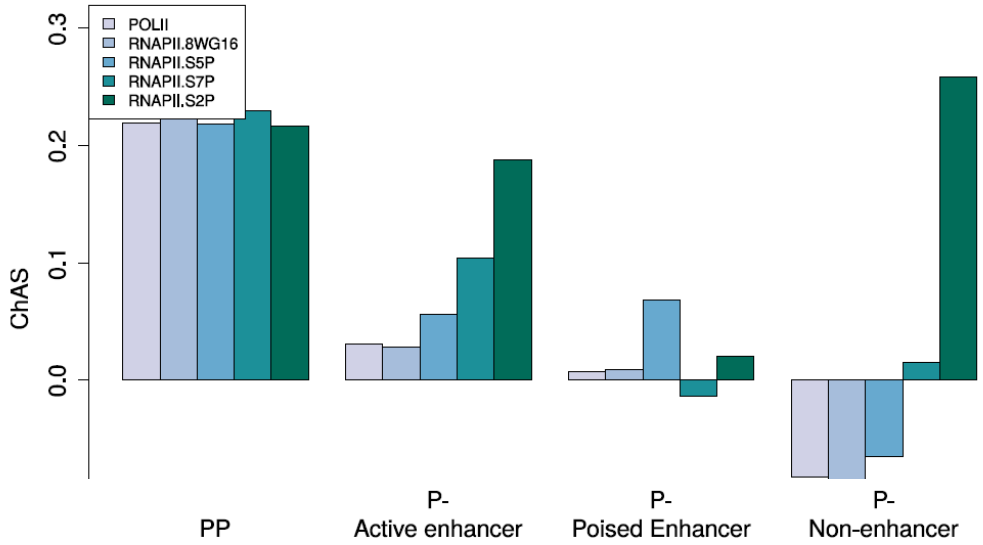
Is expression enhancing mediated by RNAPII S2p?

Assortatiivity of RNAPII forms in
Interactions of promoter and enhancers

Active enhancer
H3K4me1+H3K27ac

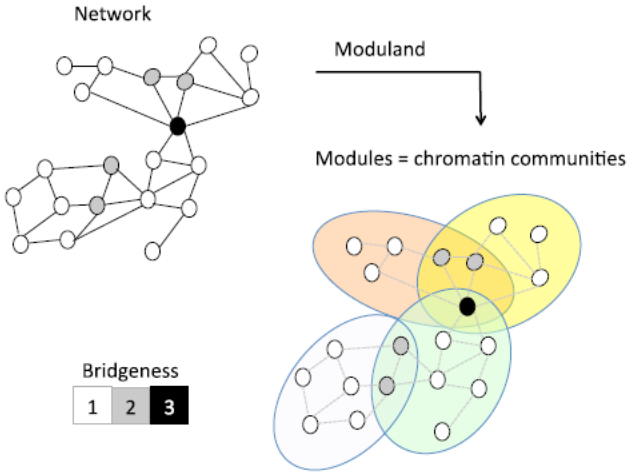
Poised Enhancer
H3K4me1+ H3K27me3

Non-enhancer
No H3k4me1

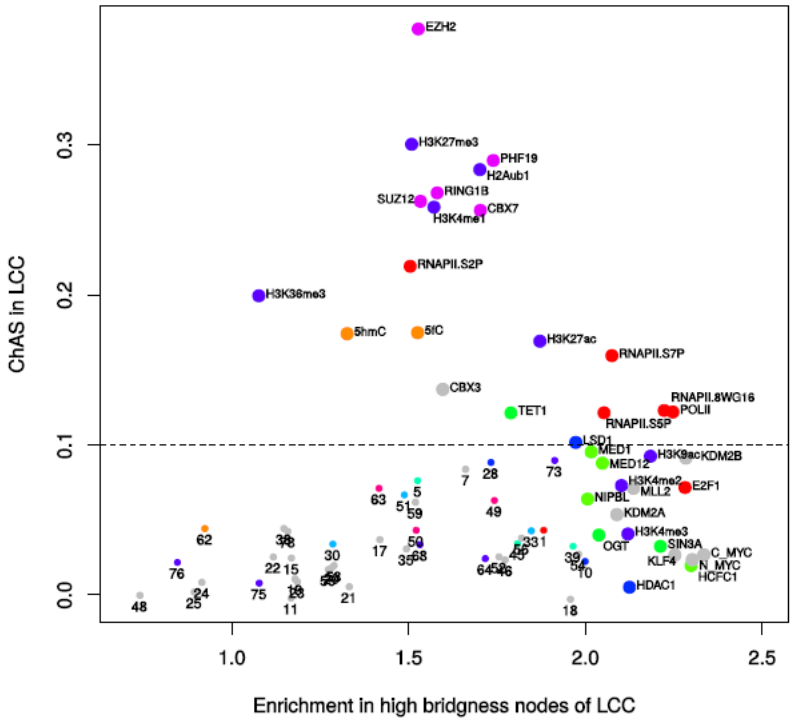


Topological properties of PCG and RNAPII nodes

Apply Moduland to identify overlapping chromatin communities, measure bridgeness

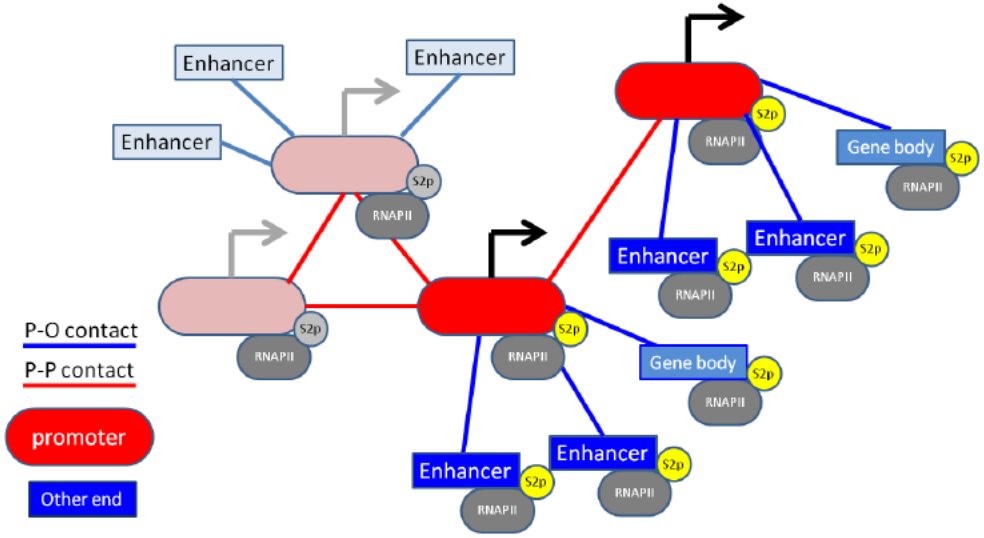


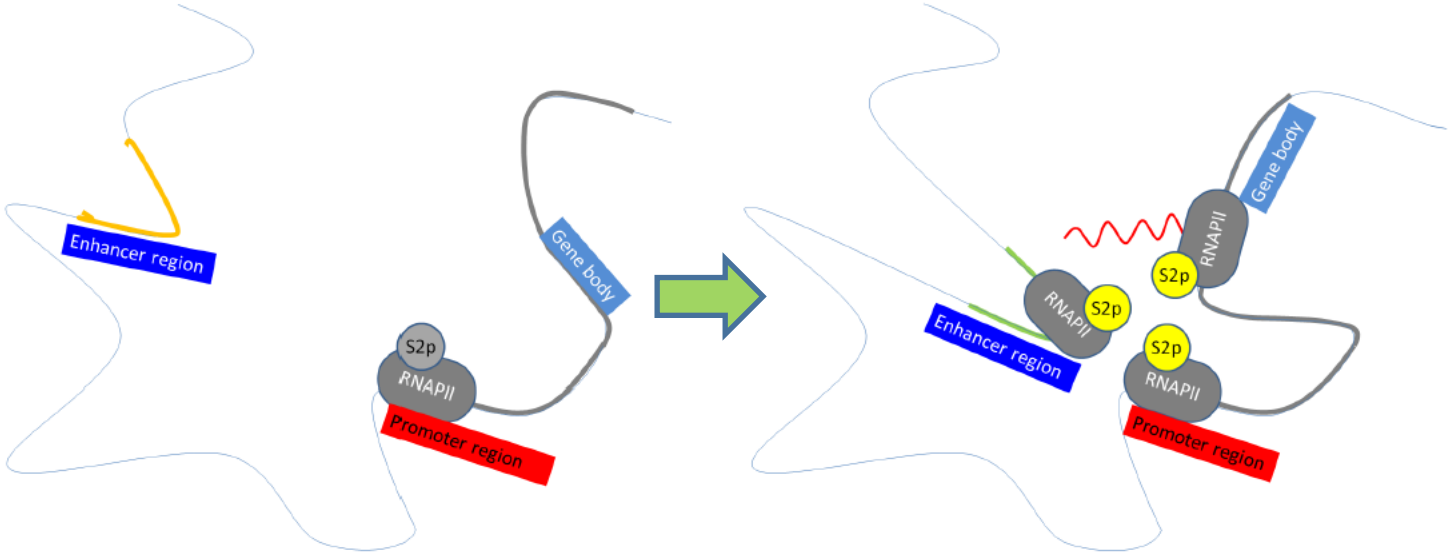
LCC= Large connected component



	Bridgeness	Betweenness centrality	Degree	Clustering Coefficient
PCG	Low	High	High	Very low
RNAPII general	High	Low	Low	Low
RNAPII S2p	Low	Very low	Very low	Medium

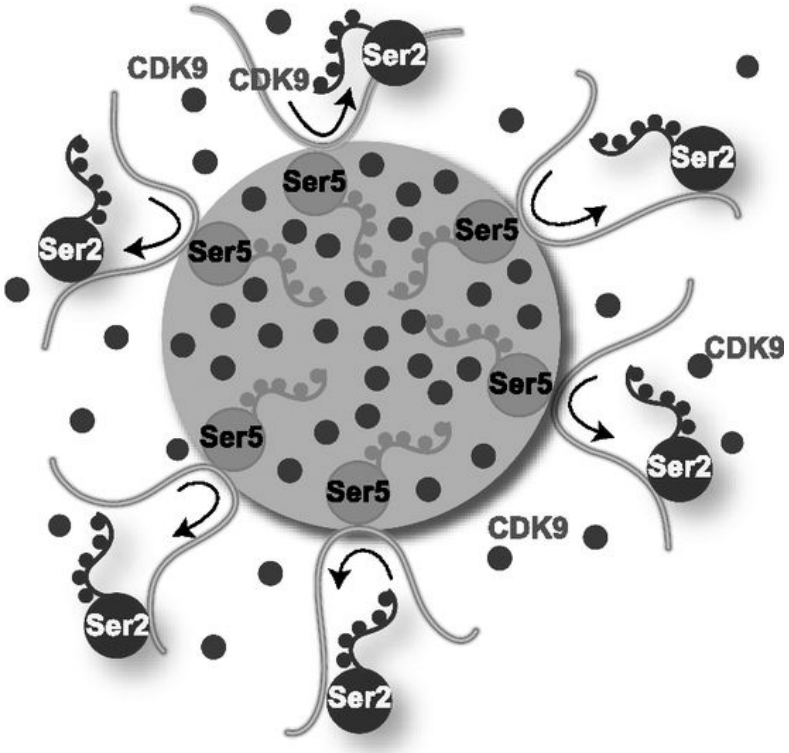
Model proposed:





Experimental results on RNAPII S2p

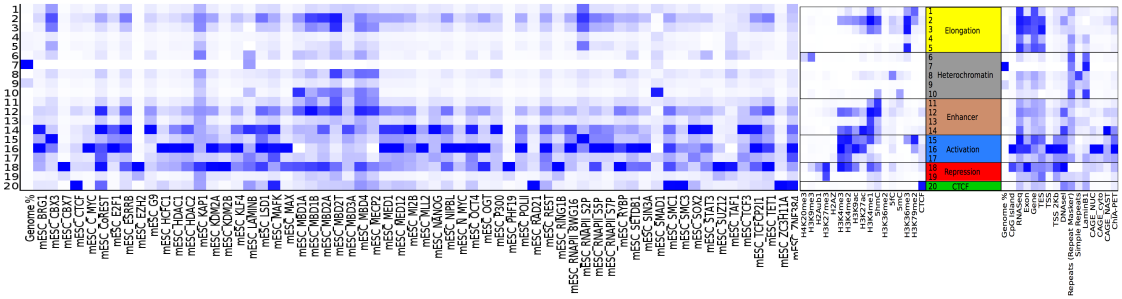
Whereas RNAPII S5P accumulates in transcription factories, RNAPII S2p stays peripheral



A model of transcription; gene promoters are loaded with RNAPII-Ser5P (Ser5 light gray) in factories. Elongating RNAPII-Ser2P (Ser2, dark gray) moves to the adjacent nuclear space when it becomes phosphorylated at Ser2 by CDK9

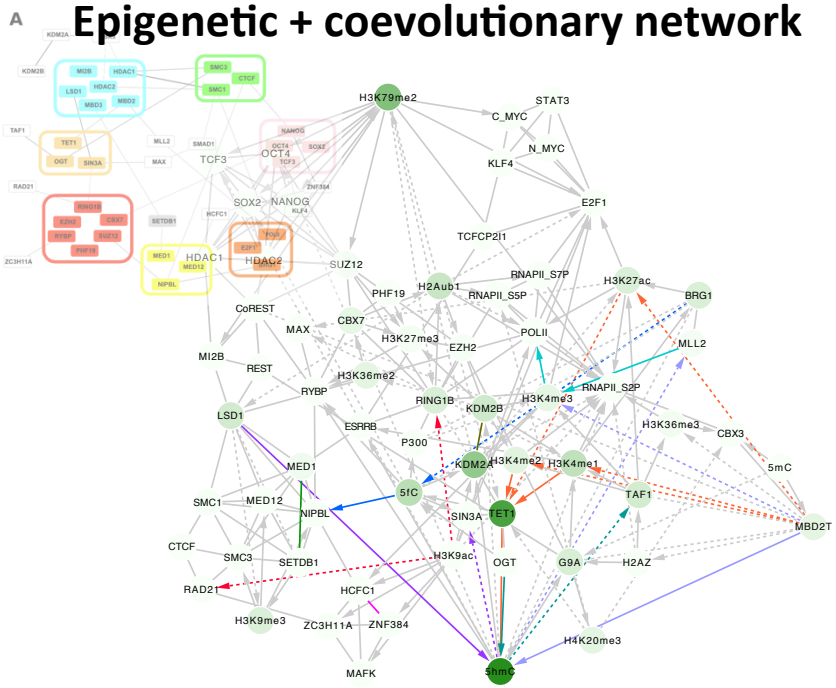
A. Ghamari et al. Genes Dev. 2013;27:767-777

Epigenomic Mouse Stem Cell

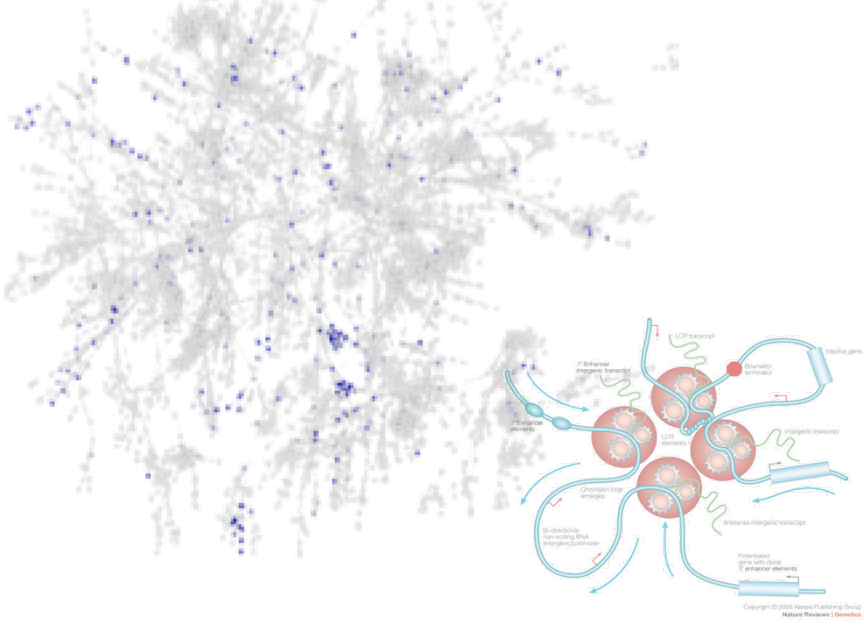


Distribution of CRPs in 20 Heterochromatin states

Part 1
Epigenetic + coevolutionary network



Part 2
Cromatin Capture Network



www.cnio.es

OPENMINDTED

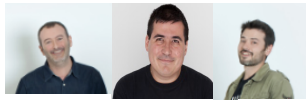
ASSET EU VIIFP

Spanish MINECO "gene sets"

Looking for students, postdocs and bioinformaticians …

Bioinformatics Unit

Translational Bioinformatics



Spanish National Bioinformatics Institute

MPIMG

Juliane Perner | Martin Vingron | Ho-Ryun Chung

Babraham

Biola-Maria Javierre | Peter Fraser | Mikhail Spivakov

Enrique Carrillo | David Juan | Simone Marsili | Daniel Rico | Vera Pancaldi

ISCB fellows

Alfons_Valencia