



Omics Data Integration in Biomedical Research and Precision Medicine

Jornadas SEFORI, CITIC-UGR


Oct 2018

Pedro Carmona-Saez

Genyo Bioinformatics Unit



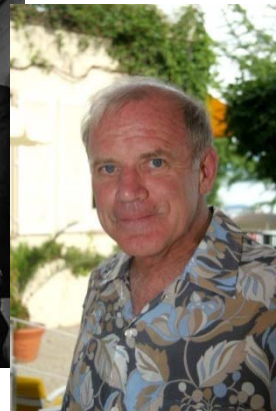
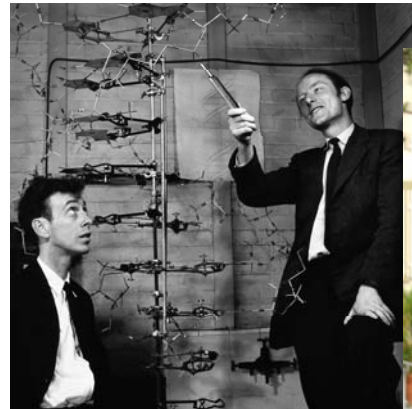
Universidad de Granada



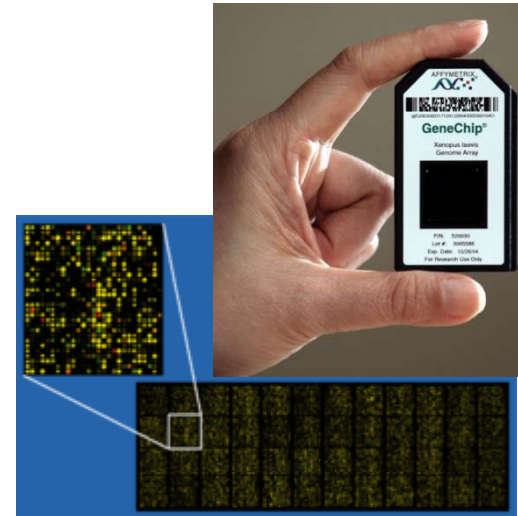
Biomedical research guided by the evolution of technology



1960-80



Biomedical research guided by the evolution of technology



1960-80

Late 90s



REPORT

Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray

Mark Schena⁽¹⁾, Dari Shalon⁽¹⁾, Ronald W. Davis⁽²⁾, Patrick O. Brown⁽³⁾

+ Author Affiliations

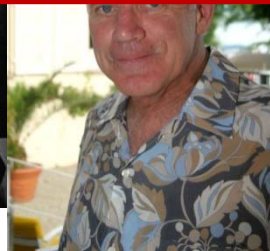
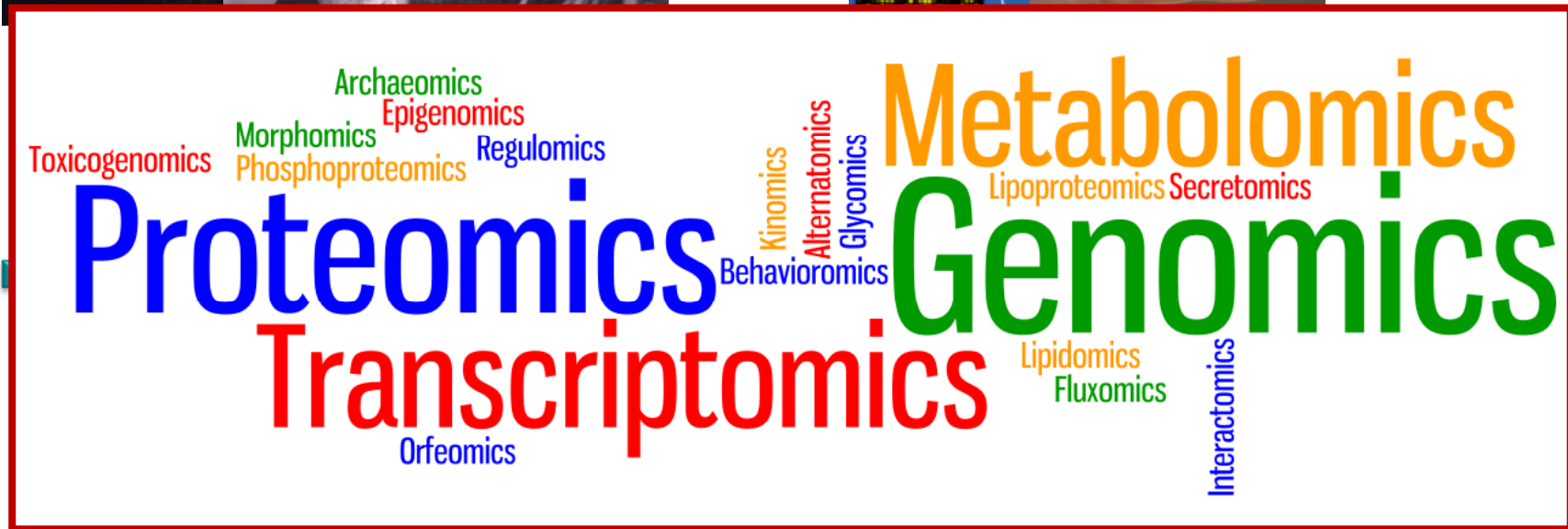
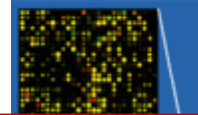
⌋⁽¹⁾ These authors contributed equally to this work.

⌋⁽²⁾ Present address: Synteni, Palo Alto, CA 94303, USA.

⌋⁽³⁾ To whom correspondence should be addressed. E-mail: pbrown@cmgm.stanford.edu

Science 20 Oct 1995:
Vol. 270, Issue 5235, pp. 467-470
DOI: 10.1126/science.270.5235.467

Biomedical research guided by the evolution of technology

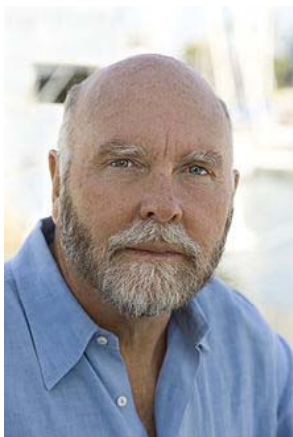


Biomedical research guided by the evolution of technology

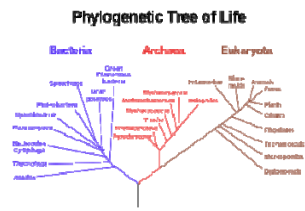
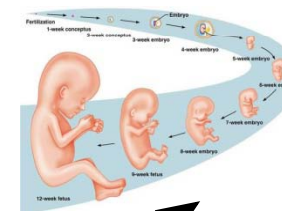
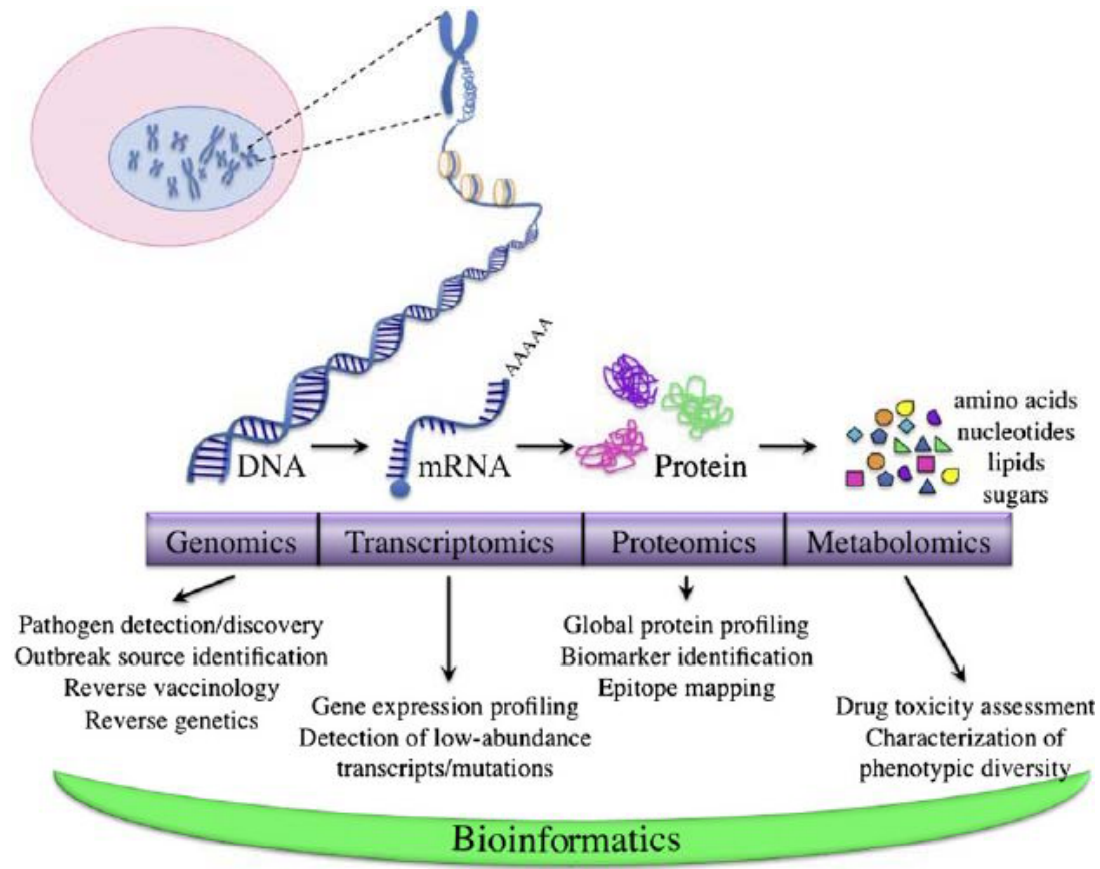


2001

2005

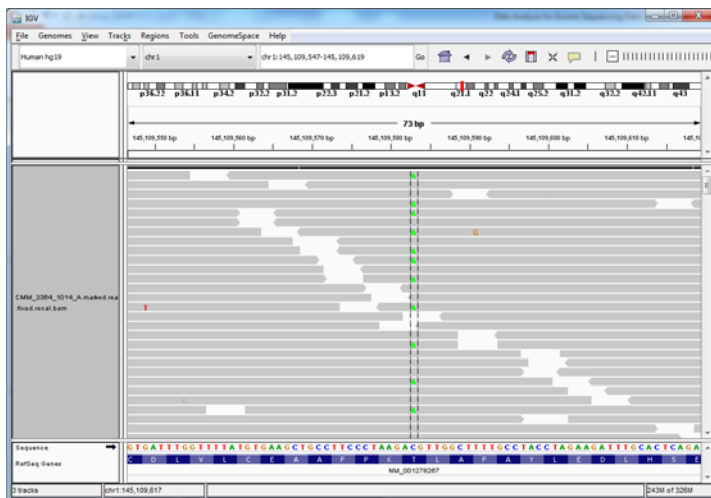
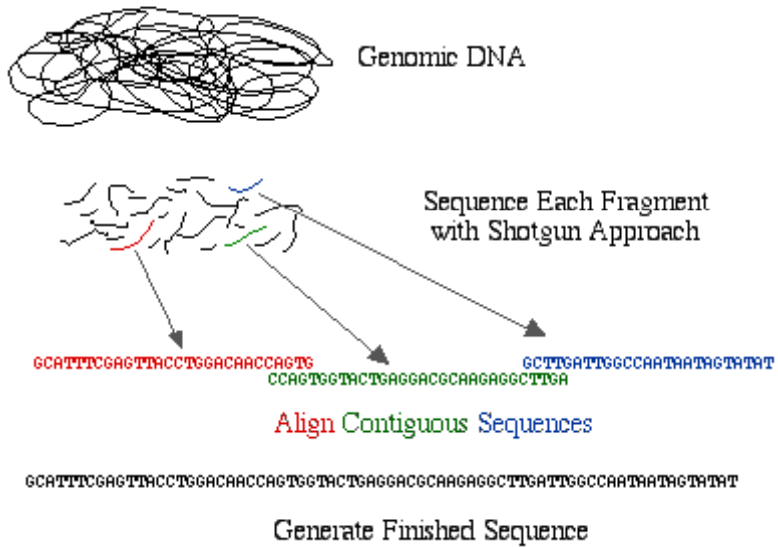


The code of life

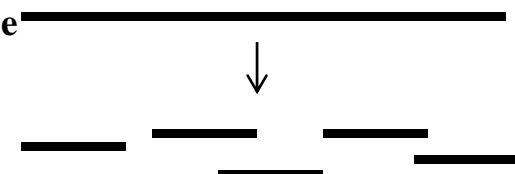


Next Generation Sequencing

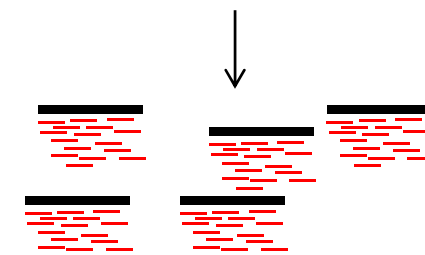
Whole Genome Shotgun Sequencing Method



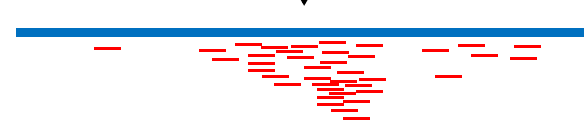
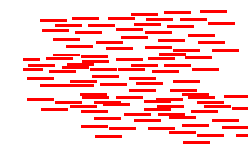
DNA sample



Sequencing



MILLIONS of short sequences



Reference genome

Next Generation Sequencing

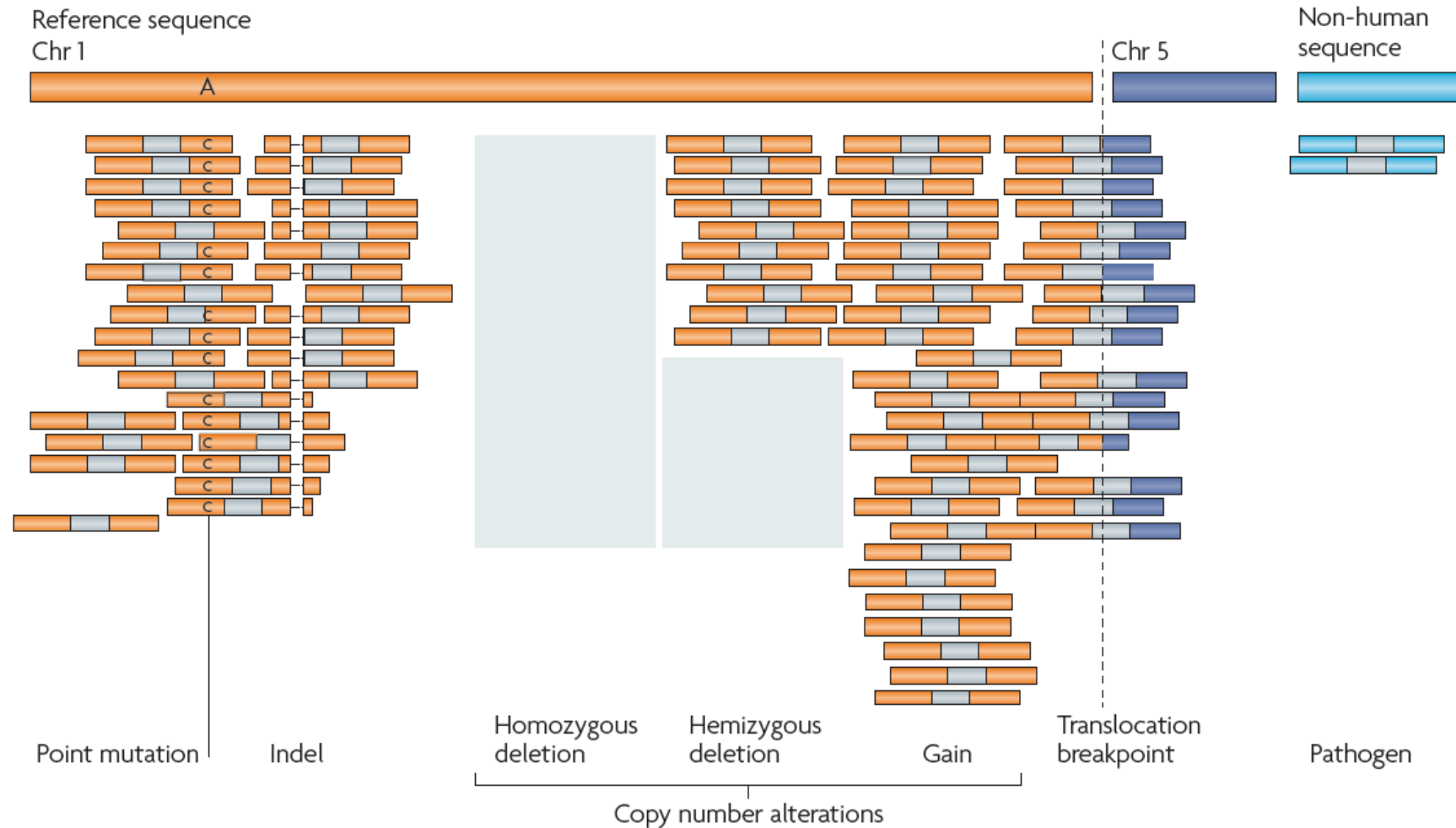


Figure 3 | **Types of genome alterations that can be detected by second-generation sequencing.** Sequenced

Pricing trend

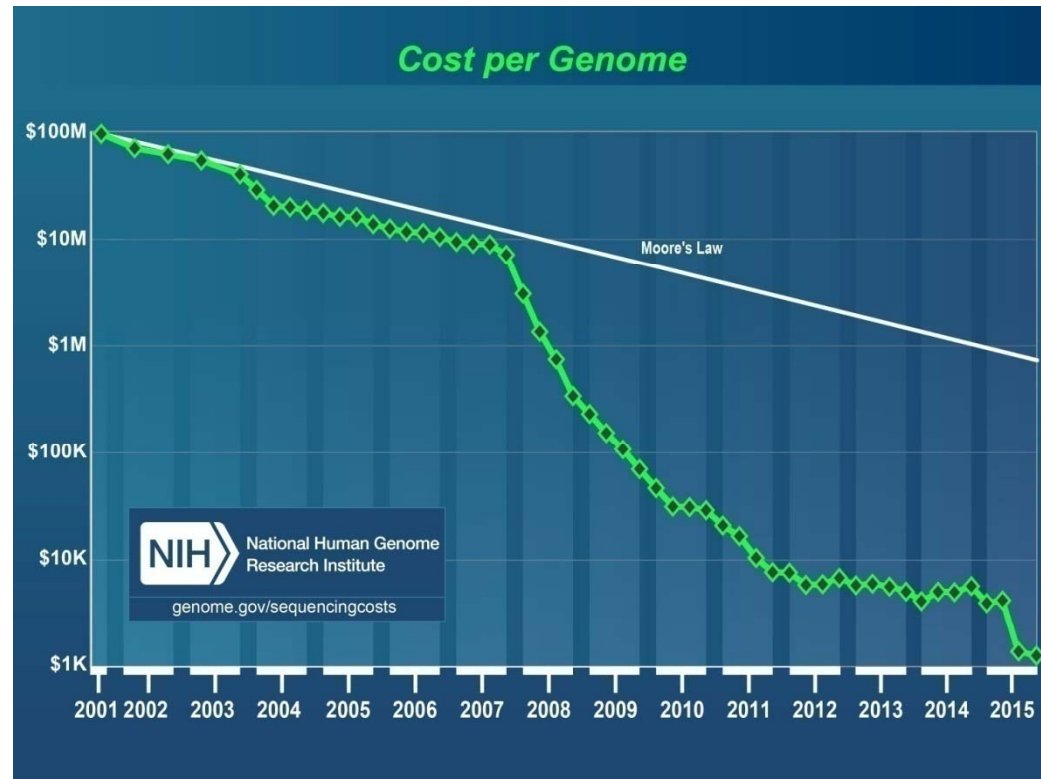
(2001) The Human genome

project:

- 13 years
- 23 labs
- \$500 Million.

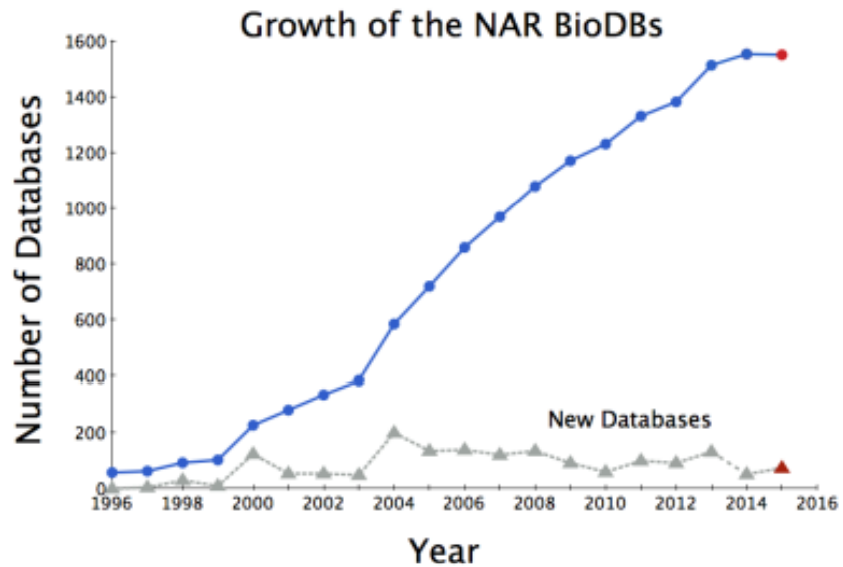
• A Human genome today:

- 1 day
- 1 machine
- \$1,000

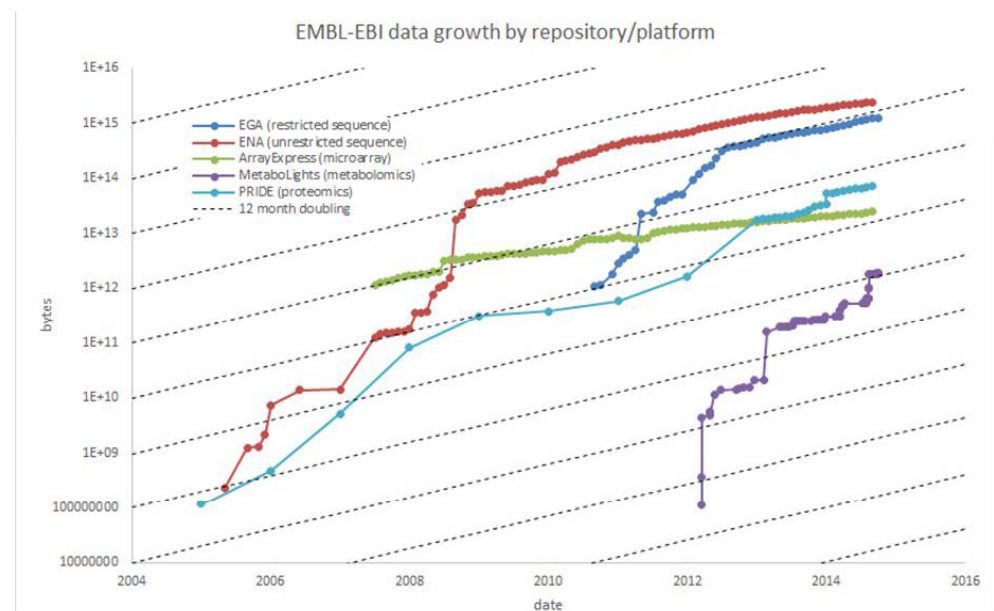


Increasing Information and Resources

Exponential growth of Biological Databases and information



Growth of Omics Data (EBI Repositories)



Big Data in Biology

PERSPECTIVE

Big Data: Astronomical or Genomical?

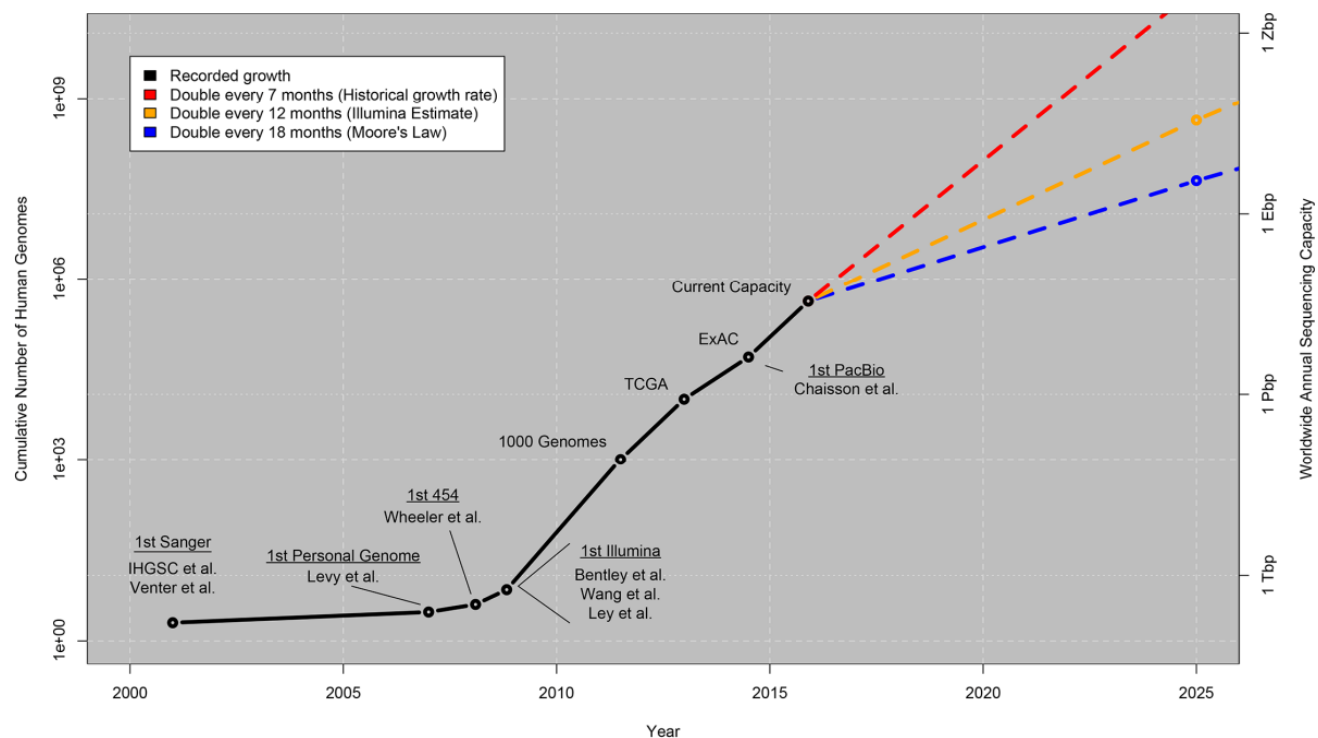
Zachary D. Stephens¹, Skylar Y. Lee¹, Faraz Faghri², Roy H. Campbell², Chengxiang Zhai³, Miles J. Efron⁴, Ravishankar Iyer¹, Michael C. Schatz^{5*}, Saurabh Sinha^{3**}, Gene E. Robinson^{6*}

PLOS Biology 2015

Abstract

Genomics is a Big Data science and is going to get much bigger, very soon, but it is not known whether the needs of genomics will exceed other Big Data domains. Projecting to the year 2025, we compared genomics with three other major generators of Big Data: astronomy, YouTube, and Twitter. Our estimates show that genomics is a “four-headed beast”—it is either on par with or the most demanding of the domains analyzed here in terms of data acquisition, storage, distribution, and analysis. We discuss aspects of new technologies that will need to be developed to rise up and meet the computational challenges that genomics poses for the near future. Now is the time for concerted, community-wide planning for the “genomical” challenges of the next decade.

Growth of DNA Sequencing



Biological research has changed

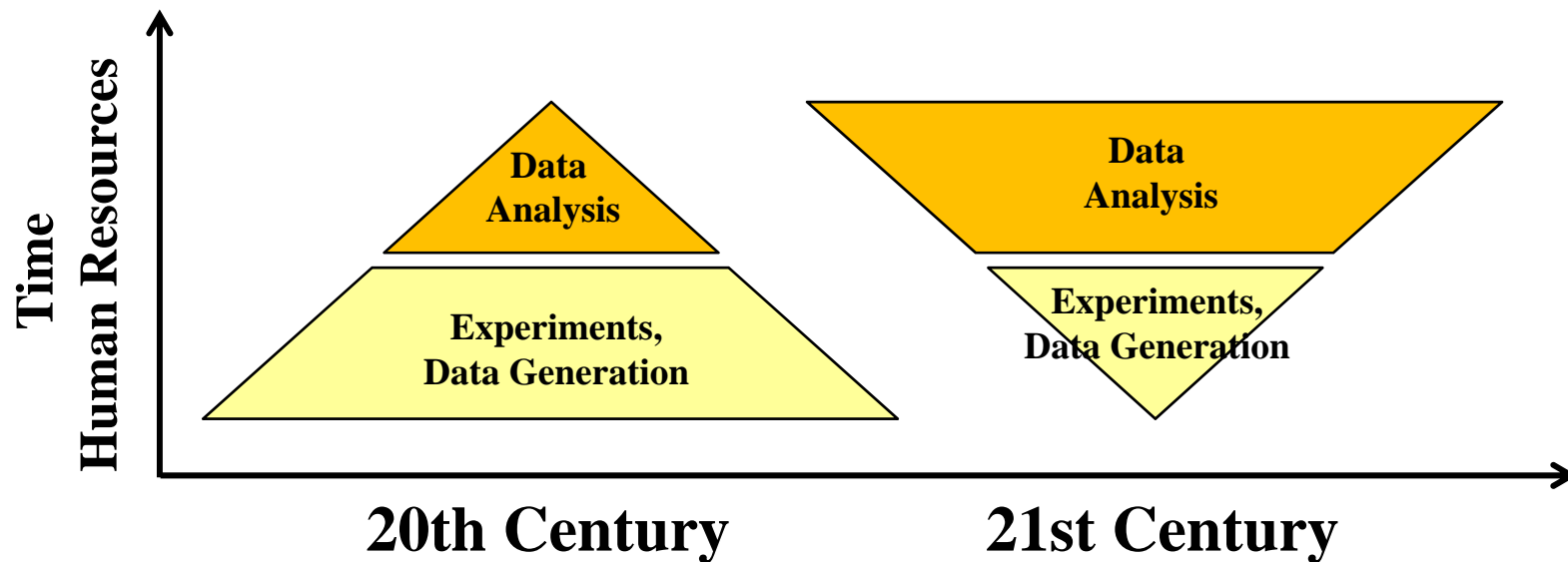
Mardis *Genome Medicine* 2010, 2:84
<http://genomemedicine.com/content/2/11/84>



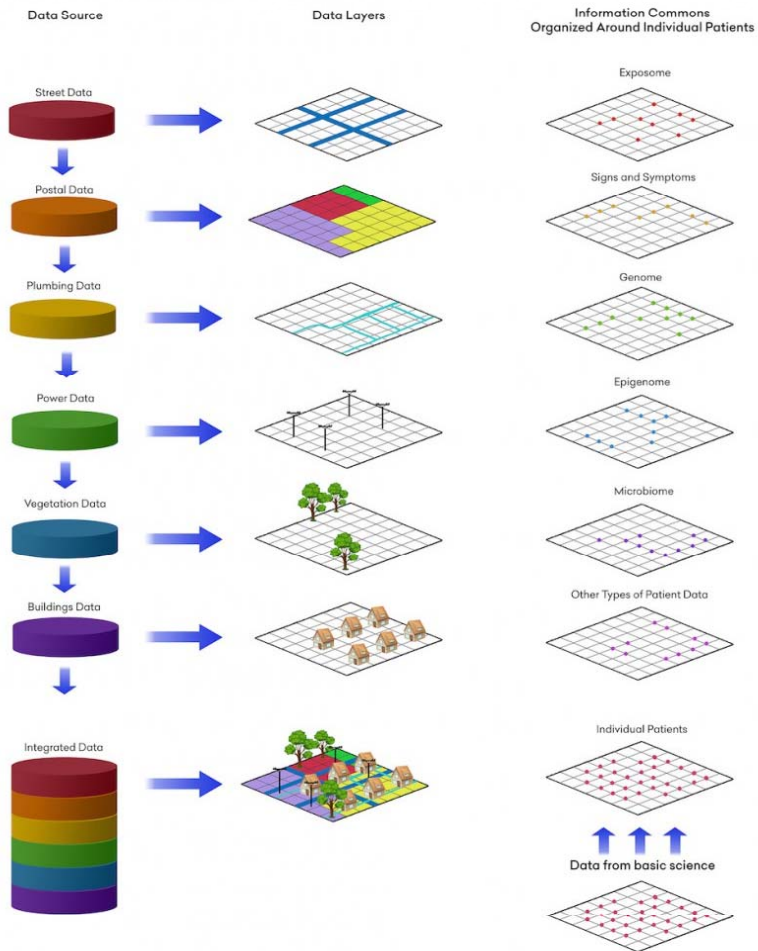
MUSINGS

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

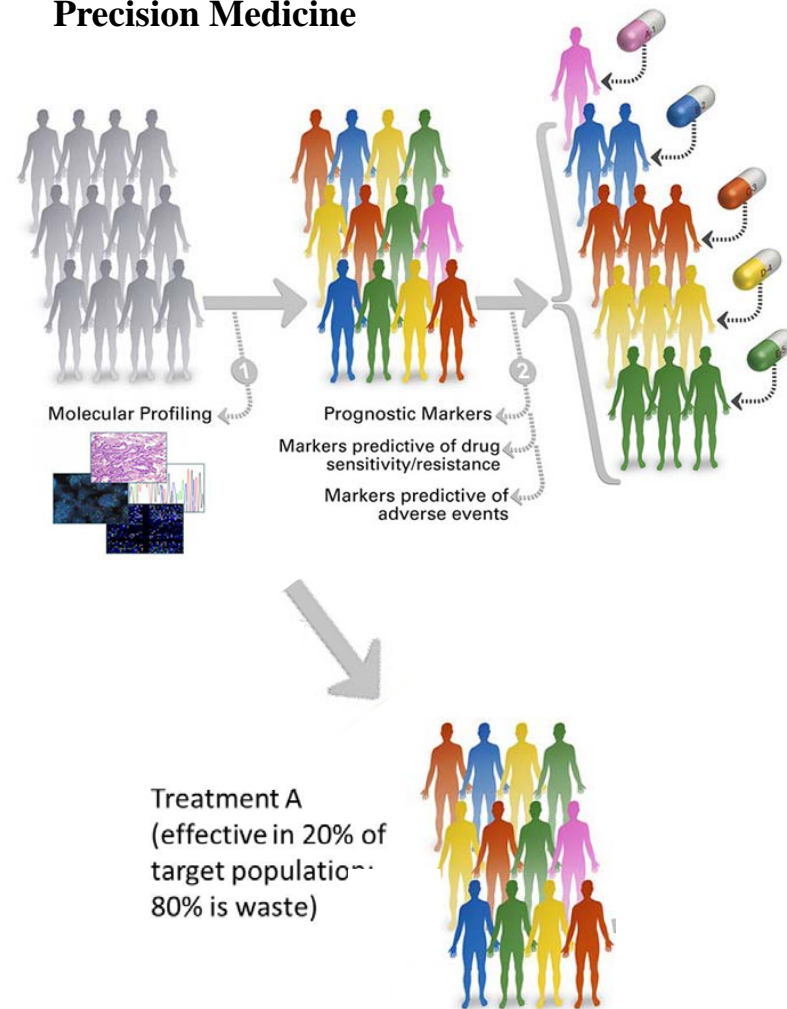
Elaine R Mardis*



Precision Medicine



Precision Medicine



Traditional Approach

What are the Challenges?

Storage

- Fast and efficient storage systems to query large collections of ‘omics data.

Distribution and data access

- Cloud-computing systems such as EasyGenomics (BGI) or “Embassy” as part of ELIXIR project

Analysis

- Algorithms/methods to **extract knowledge and information from the data**
- Integration of biological domain expertise, large-scale machine learning systems and efficient computing infrastructure

A COMPLEX TASK

Biological systems are complex. Genes may carry out different functions in different cell types / tissues, even antagonist functions.

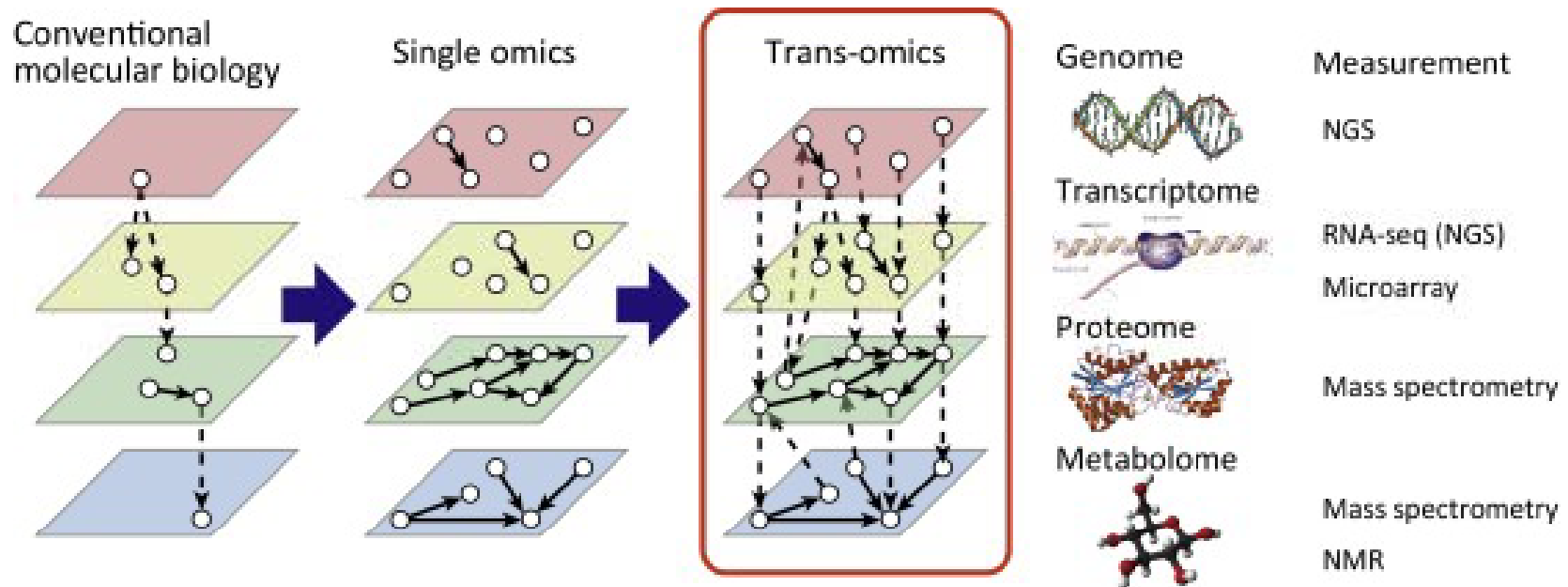
- **Signal-to-noise ratio and confounding.** High throughput biological data have low signal-to-noise ratio and many variables that make it difficult to distinguish signals from random patterns.

- **High-dimensionality**, large number of variables and small number of samples

- **Biological systems are dynamic**

Integrating Omics Data

- **Integrating different studies/datasets with the same type of data.** Integrate data with the thousands of other published datasets and look for similar patterns.
- **Integrating multiple data types.** Integrate different omics data (genomics, transcriptomics, metabolomics, ...)

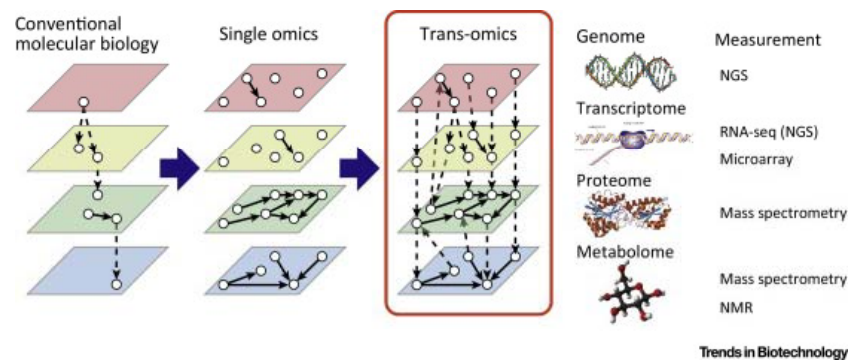


Multi-omics integration for Biomarker Discovery

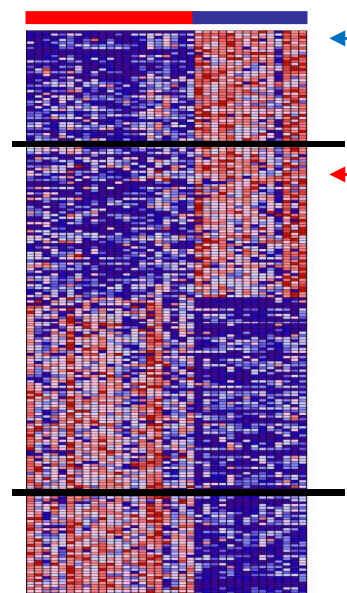
An increasing number of projects include measurements of the same samples from multiple omics techniques.

Integrating different OMICS layers will potentially provide:

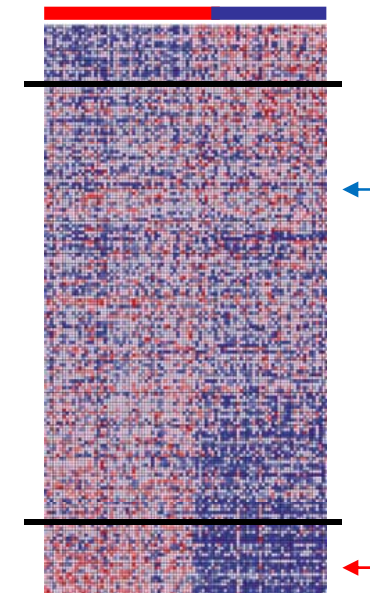
- More comprehensive view of biological processes
- More robust patterns
- Better predictions and classifications



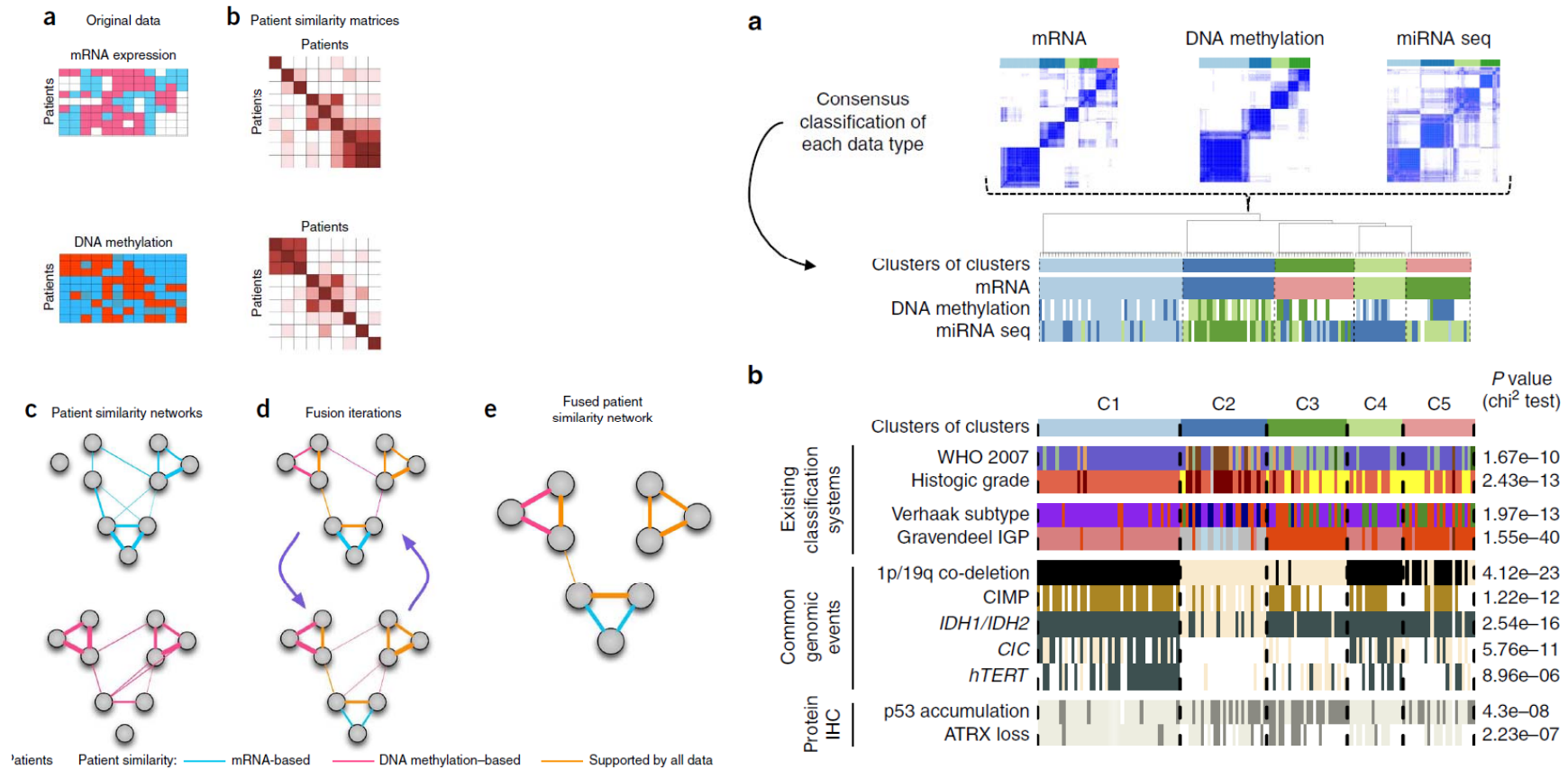
Gene Expression



Methylation



Multi-omics integration for Classification and Clustering



Wang et al. Nature Methods 2014

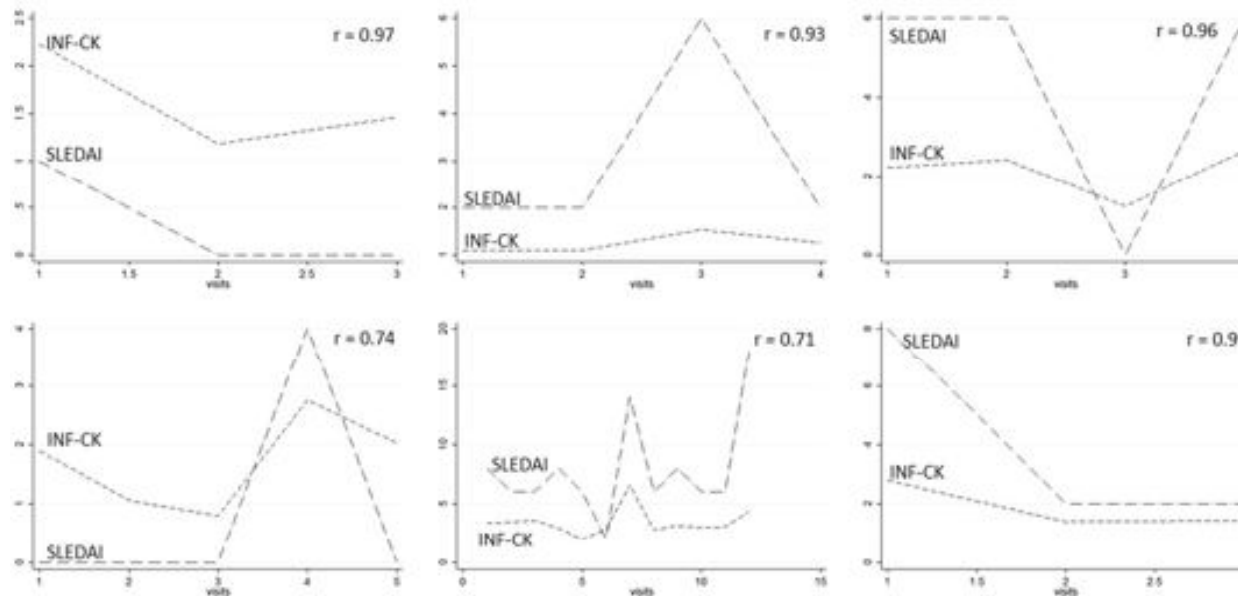
Kamoun et al. Nature Comm 2016

Cluster of clusters approach merges consensus clustering probability matrix to infer groups of sample with similar patterns across data types

Systemic lupus erythematosus

SLE is an autoimmune diseases are characterized by immune responses to self antigens that result in tissue damage

The disease course is unpredictable, with periods of remission and flares that lead to cumulative damage over time

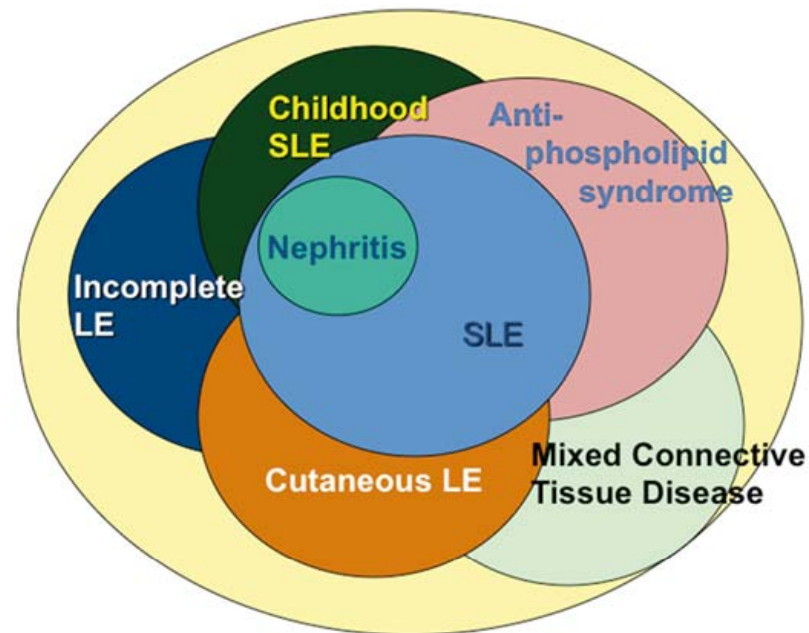


Systemic lupus erythematosus

Diagnosis is primarily clinical and remains challenging because of the heterogeneity of SLE

Only one drug (belimumab) has been approved for use in SLE in the past 60 years

Unlike in cancer, genetic traits do not seem to be generally useful as diagnostic biomarkers at present and there is an urgent need for better biomarkers and better treatments



Systemic lupus erythematosus

Published in final edited form as:

Cell. 2016 April 21; 165(3): 551–565. doi:10.1016/j.cell.2016.03.008.

Personalized Immunomonitoring Uncovers Molecular Networks That Stratify Lupus Patients

Romain Banchereau^{1,7}, Seunghee Hong^{1,7}, Brandi Cantarel¹, Nicole Baldwin¹, Jeanine Baisch¹, Michelle Edens¹, Alma-Martina Cepika¹, Peter Acs¹, Jacob Turner¹, Esperanza Anguiano¹, Parvathi Vinod¹, Shaheen Kahn², Gerlinde Obermoser¹, Derek Blankenship¹, Edward Wakeland², Lorien Nassi^{2,3}, Alisa Gotte^{2,3,4}, Marilyn Punaro^{2,3}, Yong-Jun Liu^{1,5}, Jacques Banchereau⁶, Jose Rossello-Urgell¹, Tracey Wright^{2,3}, and Virginia Pascual^{1,3,*}

¹Baylor Institute for Immunology Research, Dallas, TX 75204, USA

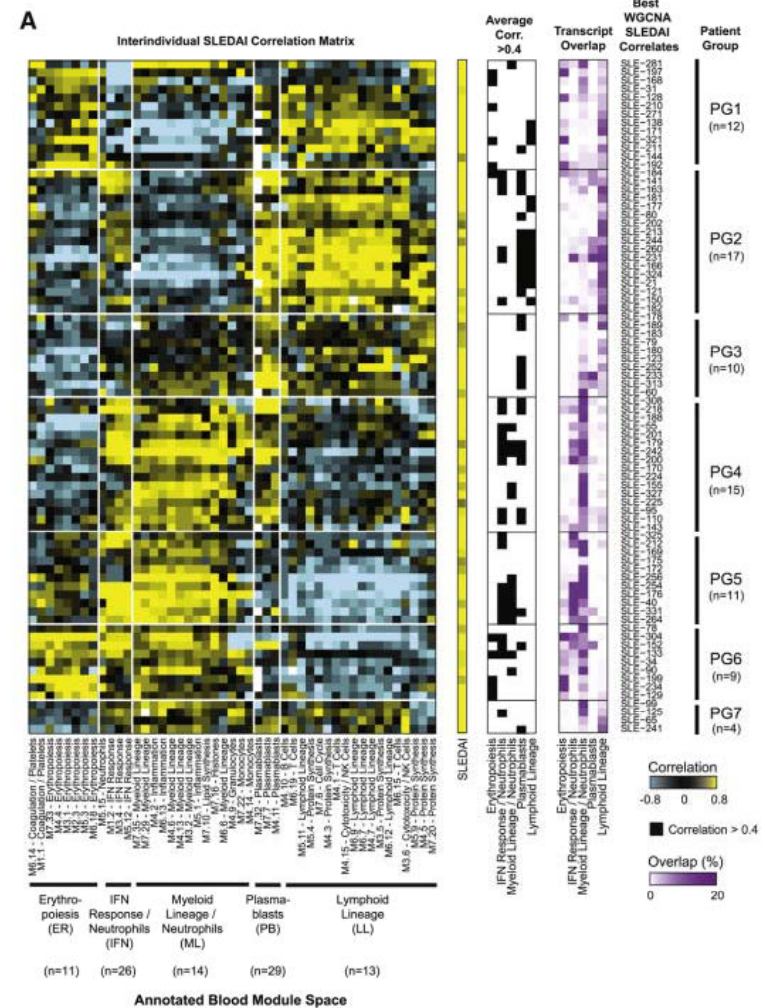
²UT Southwestern Medical Center, Dallas, TX 75235, USA

³Texas Scottish Rite Hospital for Children, Dallas, TX 75219, USA

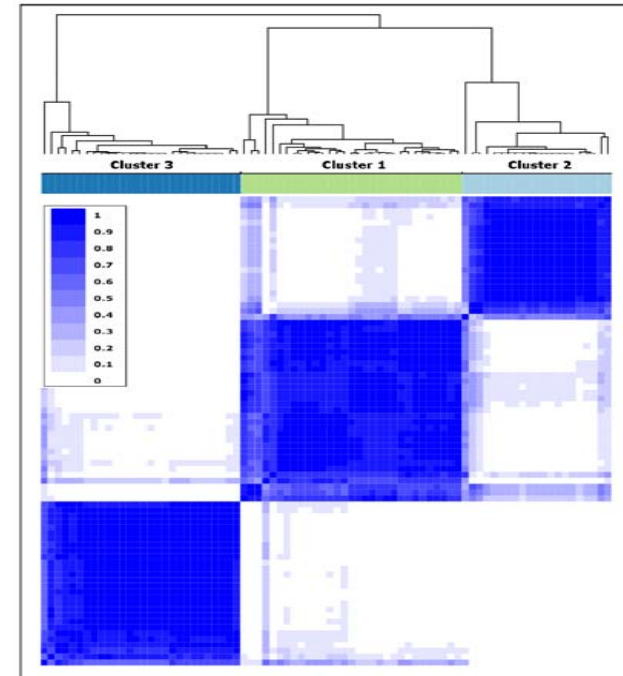
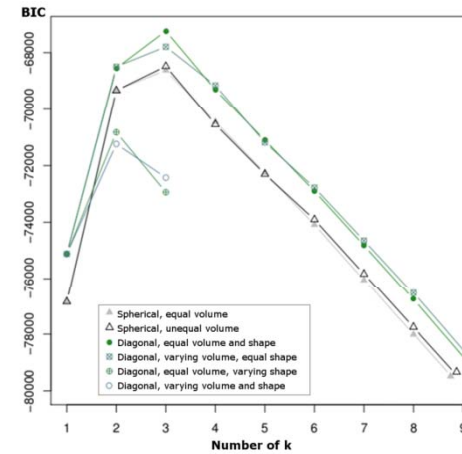
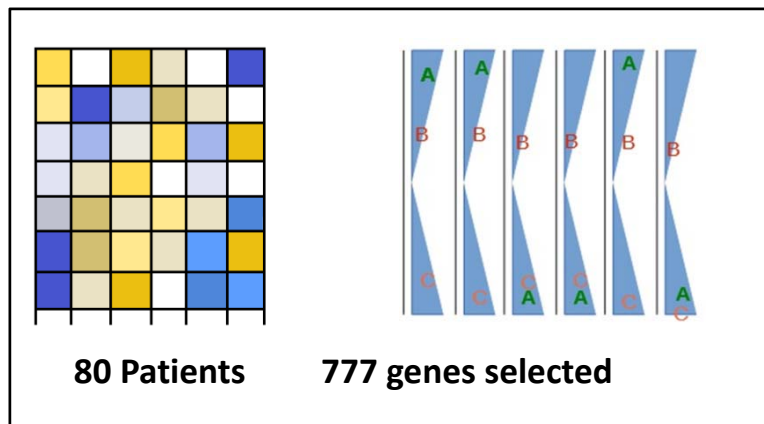
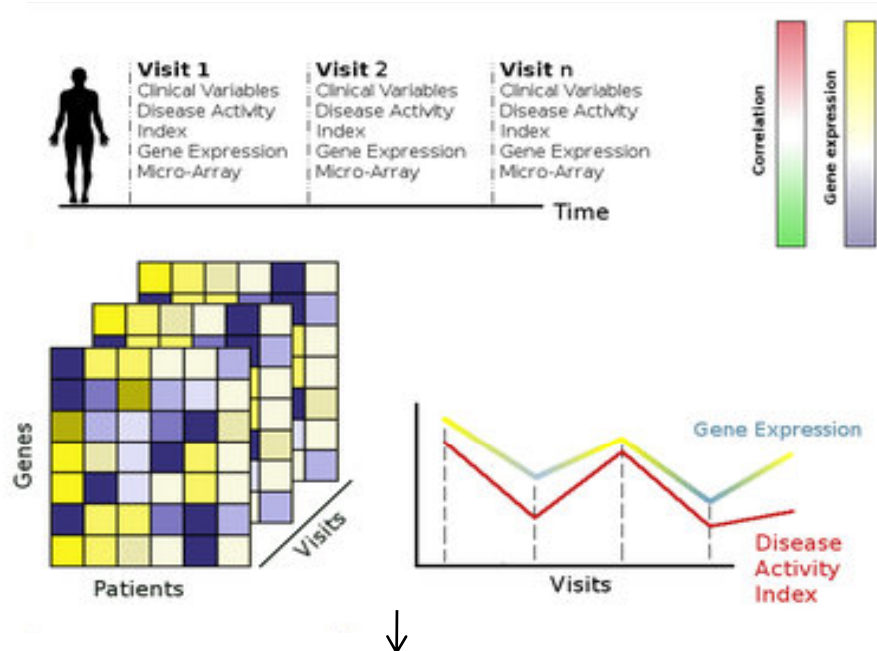
⁴Vanderbilt University School of Medicine, Nashville, TN 37232, USA

⁵MedImmune, Gaithersburg, MD 20878, USA

⁶The Jackson Laboratory for Genomic Medicine, Farmington, CT 06030, USA



Systemic lupus erythematosus



Systemic lupus erythematosus

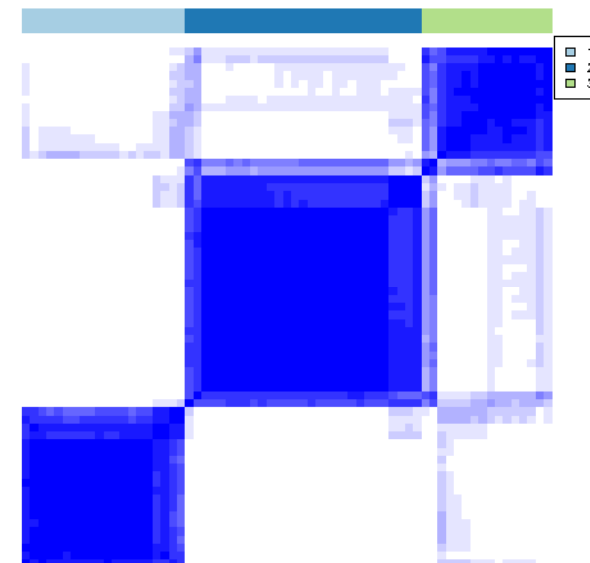
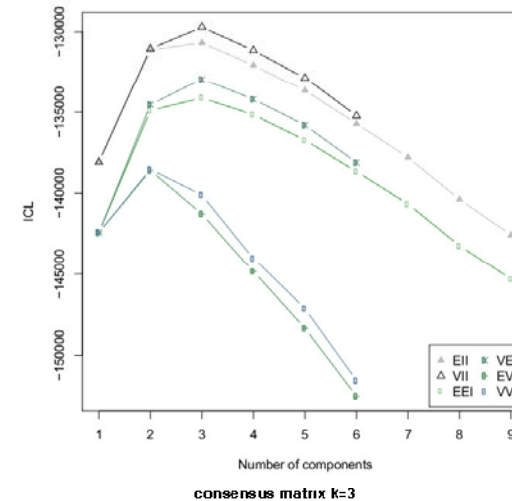
Two independent Datasets

Dataset 1. Pediatric Patients

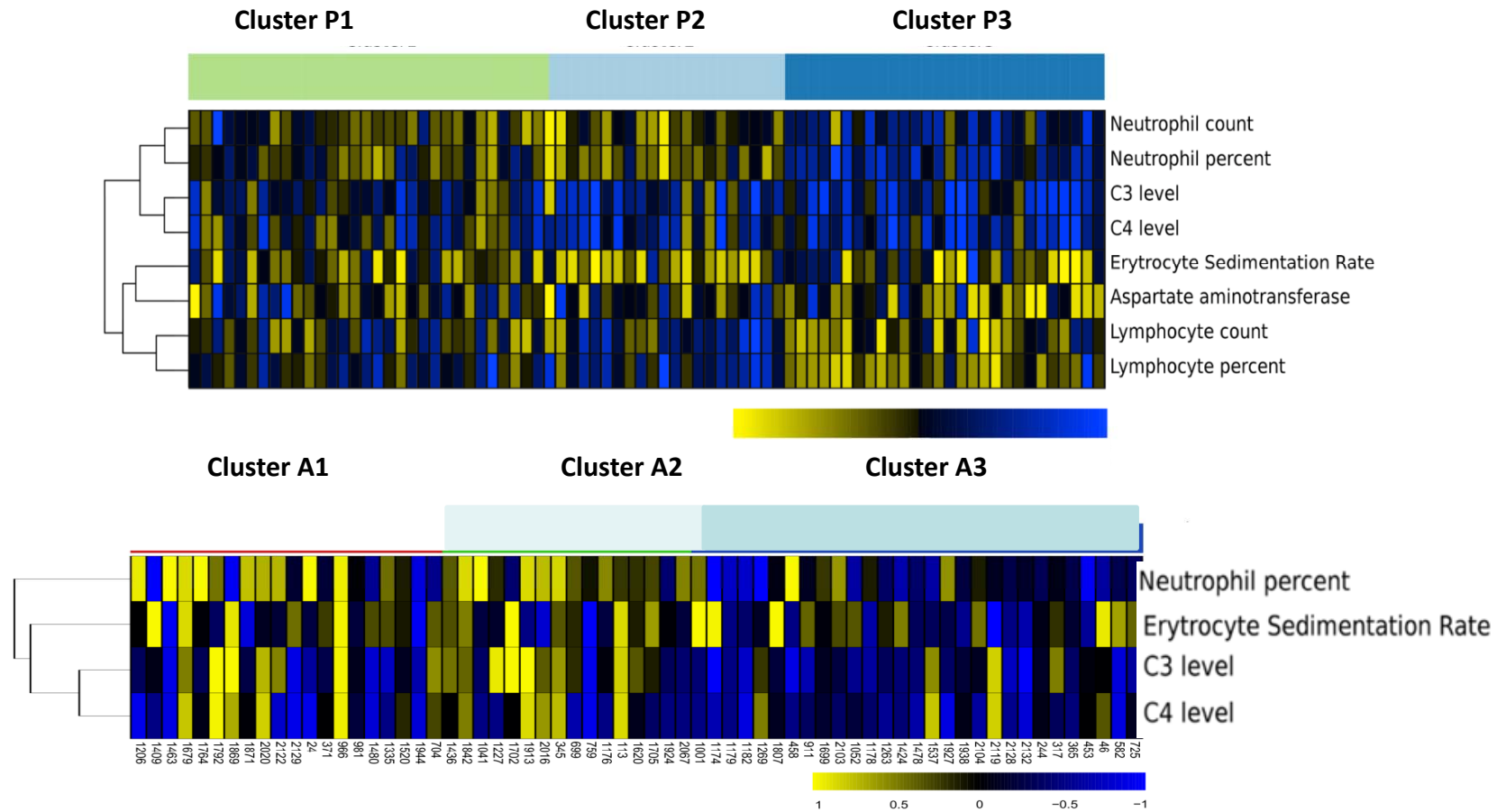
- Banchereau et al. Cell. 2016
- Longitudinal study with **158 patients and 46 Healthy Samples**
- **997 samples in total**
- Illumina GE Arrays

Dataset 2. Adult Patients

- Hopkins lupus cohort
- Longitudinal study with 301 patients **and 20 Healthy Samples**
- **747 samples in total**
- Affymetrix GE Arrays



Systemic lupus erythematosus



The percentage of neutrophils decreased with disease activity in cluster A3 and increased in clusters A2 and A1.

Toro-Domínguez D, et al. *Longitudinal Stratification of Gene Expression Reveals Three SLE Groups of Disease Activity Progression. Arthritis Rheumatol.* 2018

TF regulons

Most SLE loci occur in likely gene regulatory Regions (Maurano et al. Science 2012)

nature
genetics

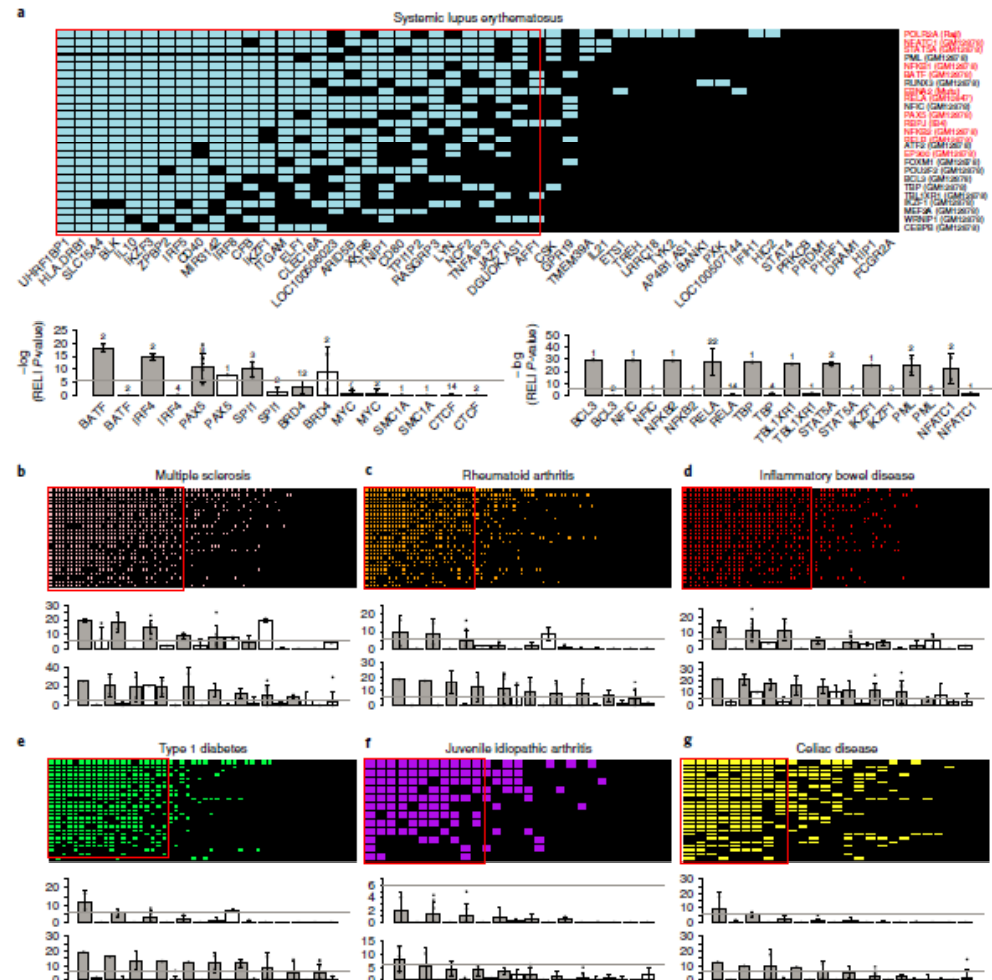
ARTICLES

<https://doi.org/10.1038/s41588-018-0102-3>

Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity

John B. Harley^{1,2,3,4,5,9*}, Xiaoting Chen^{1,9}, Mario Pujato^{1,9}, Daniel Miller¹, Avery Maddox¹, Carmy Forney¹, Albert F. Magnusen¹, Arthur Lynch¹, Kashish Chetal⁶, Masashi Yukawa⁷, Artem Barski^{1,4,7,8}, Nathan Salomonis^{4,6}, Kenneth M. Kaufman^{1,2,4,5}, Leah C. Kottyan^{1,4*} and Matthew T. Weirauch^{1,3,4,6*}

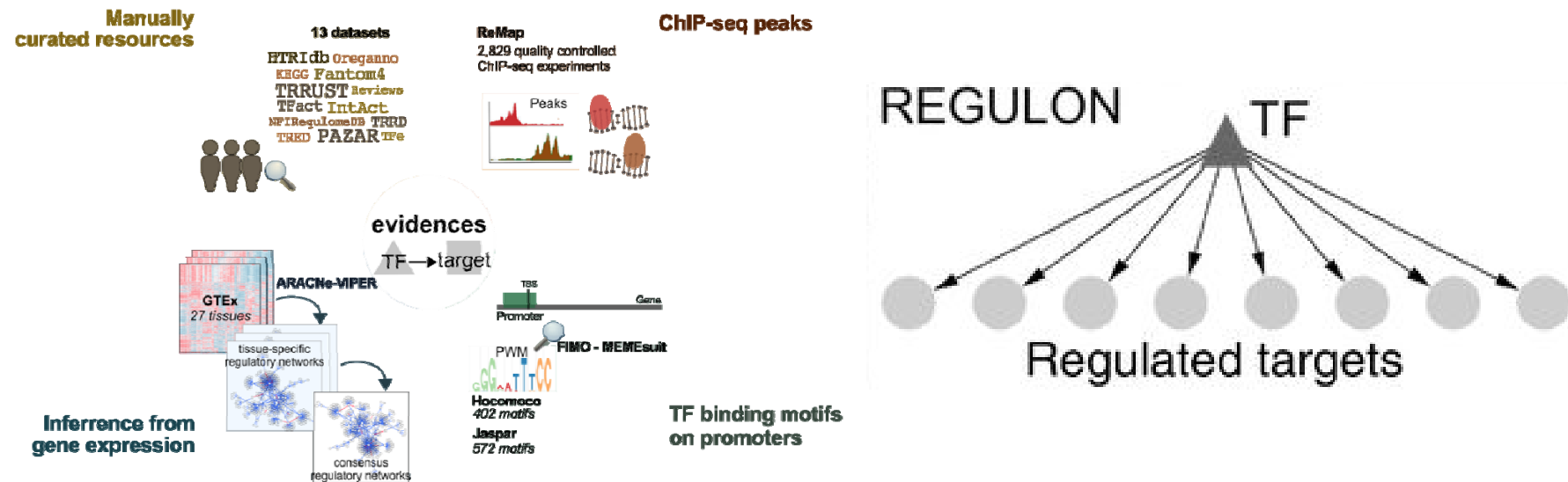
Harley et al. Nat Genet. 2018 May;50(5):699-707



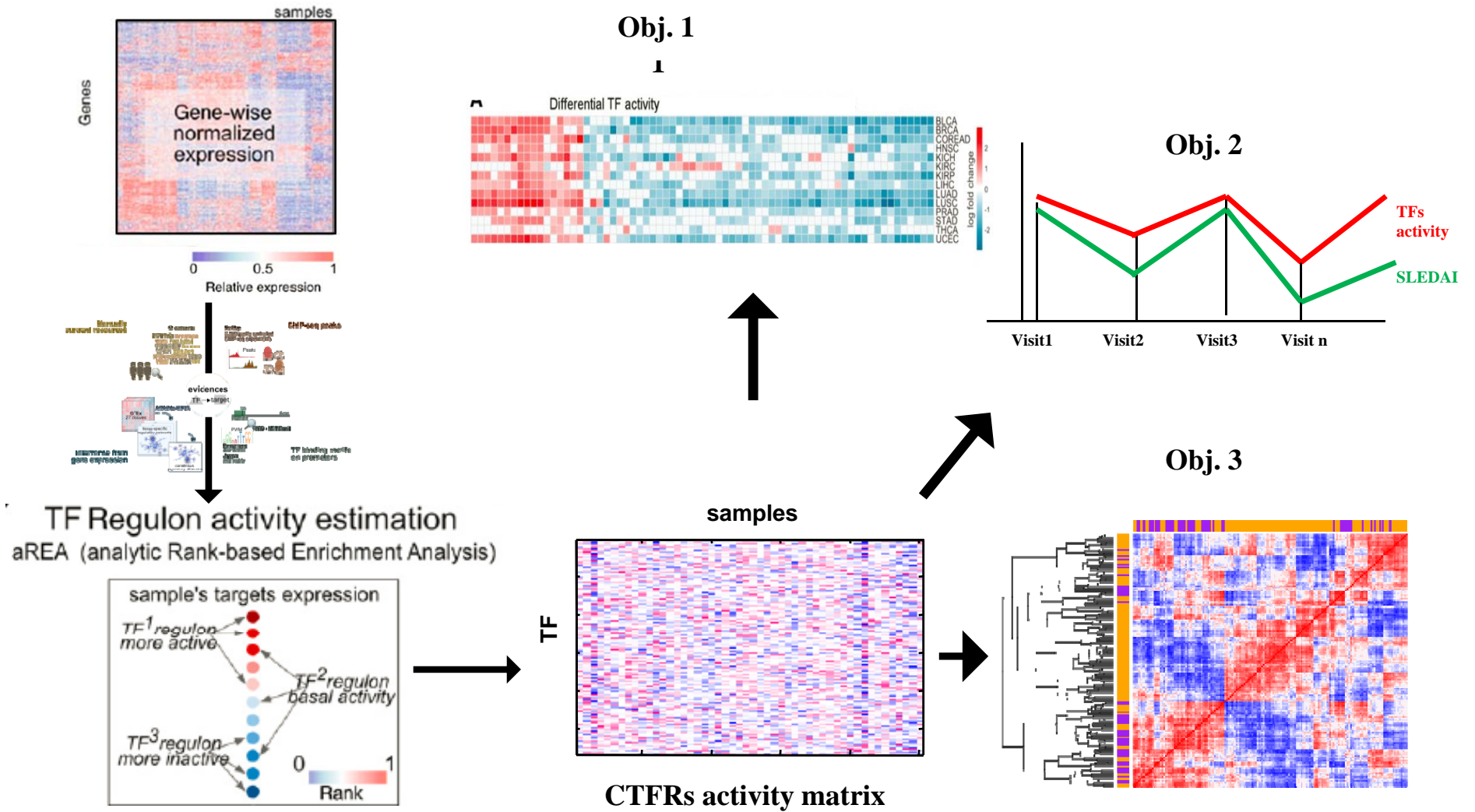
TF regulons

Activities of TF activities can be estimated computationally from the gene expression levels of their direct targets (the so-called TF regulon)

The assumption behind is that the level of protein activity of a TF is reflected on the transcript levels of its targeted genes



TF regulons



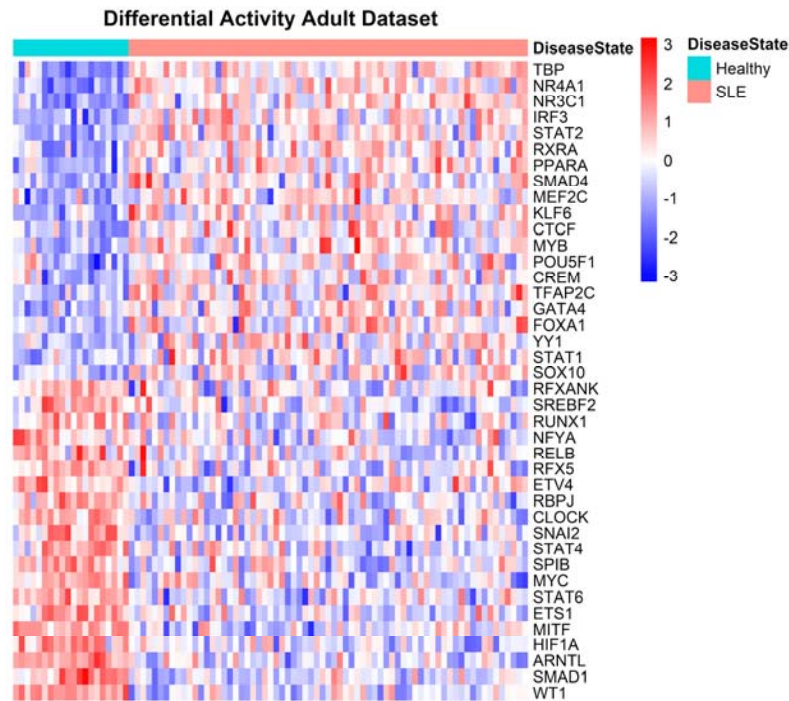
Adapted from Garcia-Alonso et al

Systemic lupus erythematosus

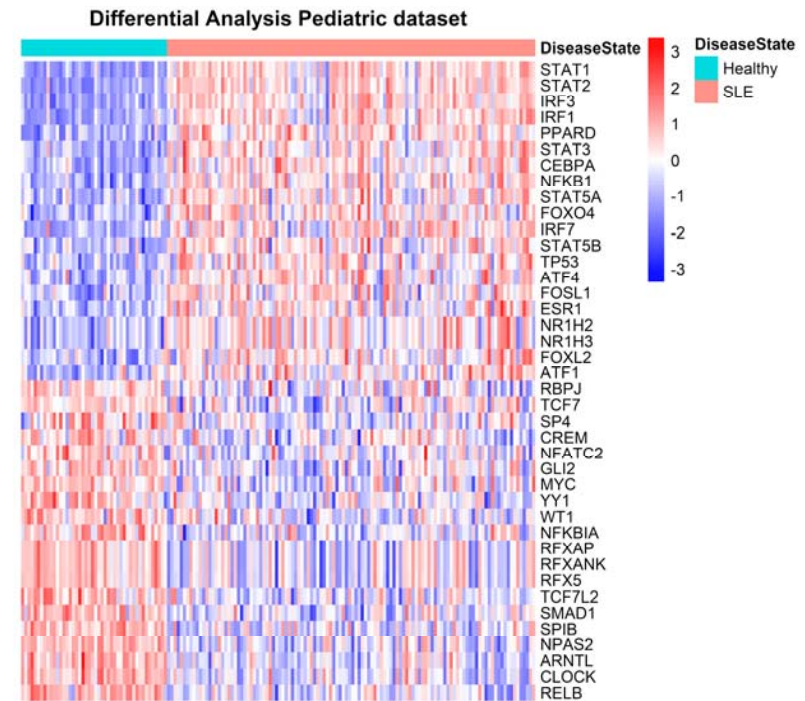
Differential Activity Analysis among lupus and healthy samples where lupus samples had a high value of SLEDAI

■ High Activity
■ Low Activity

adjPval < 0.05

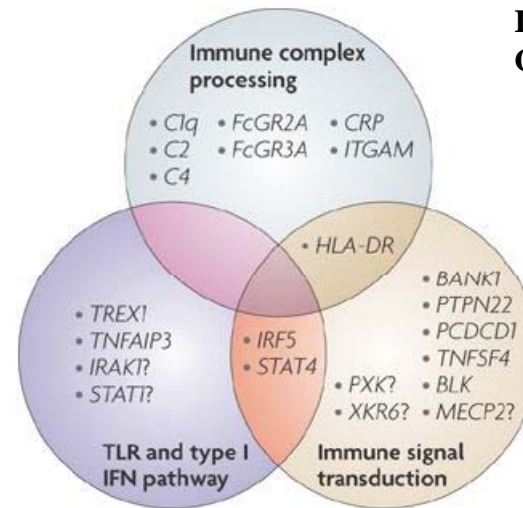
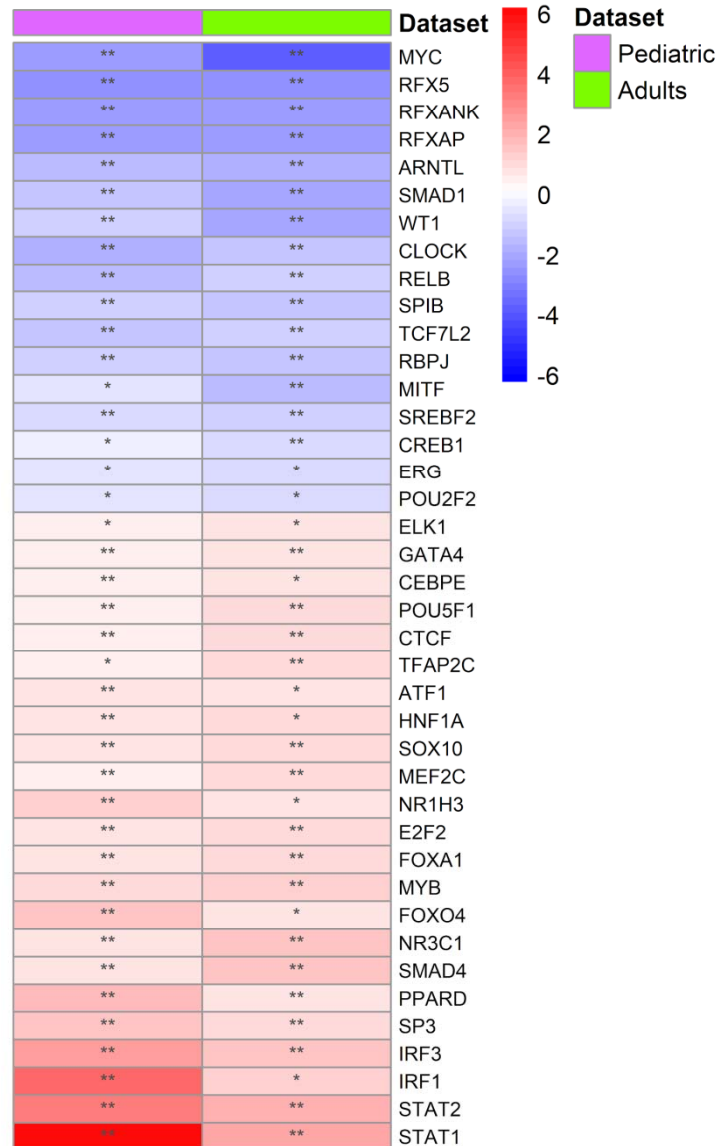


Up Activity	33
Down Activity	39



Up Activity	81
Down Activity	30

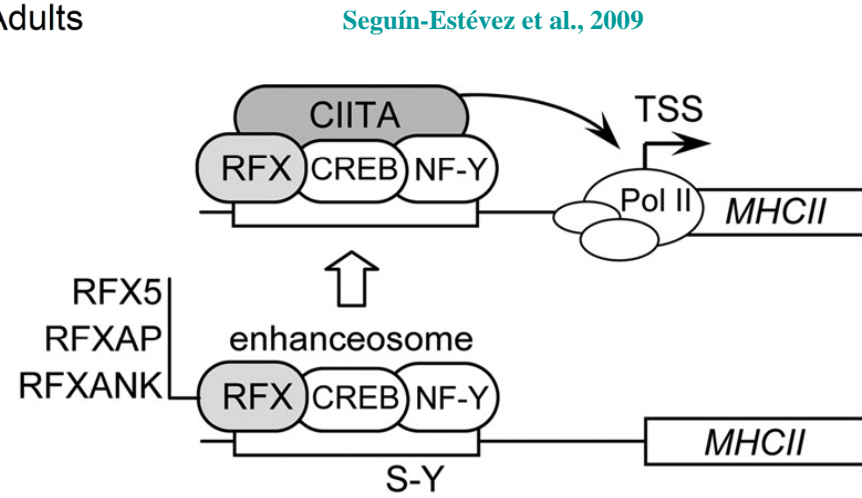
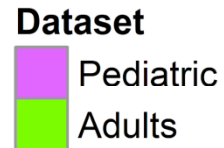
Systemic lupus erythematosus



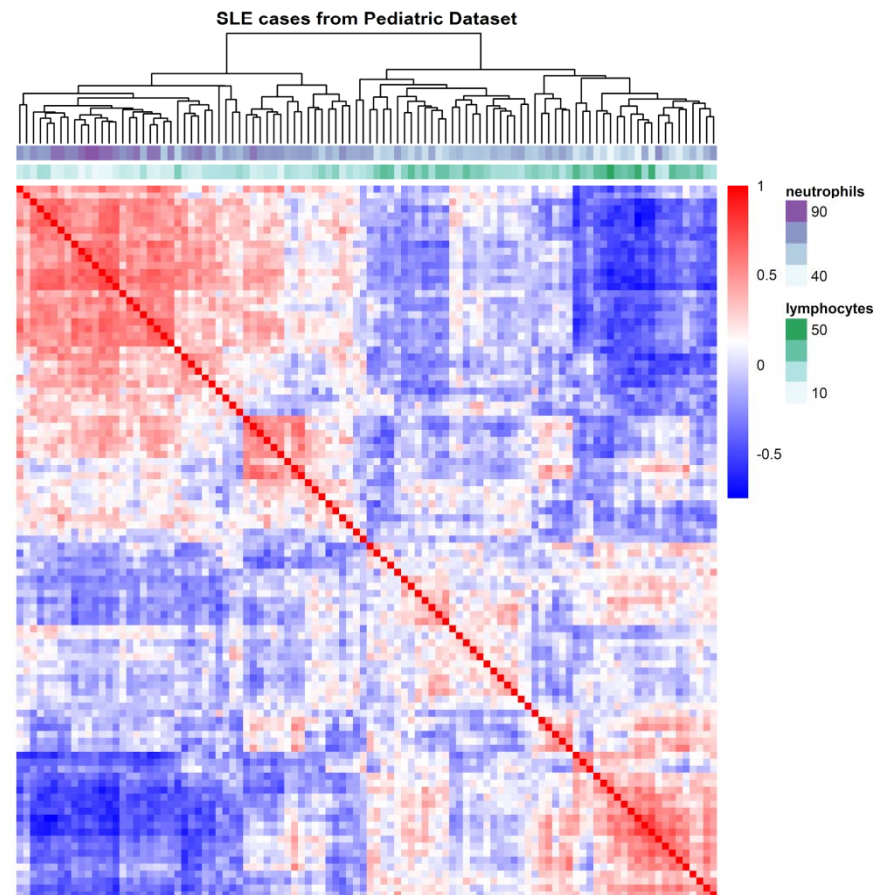
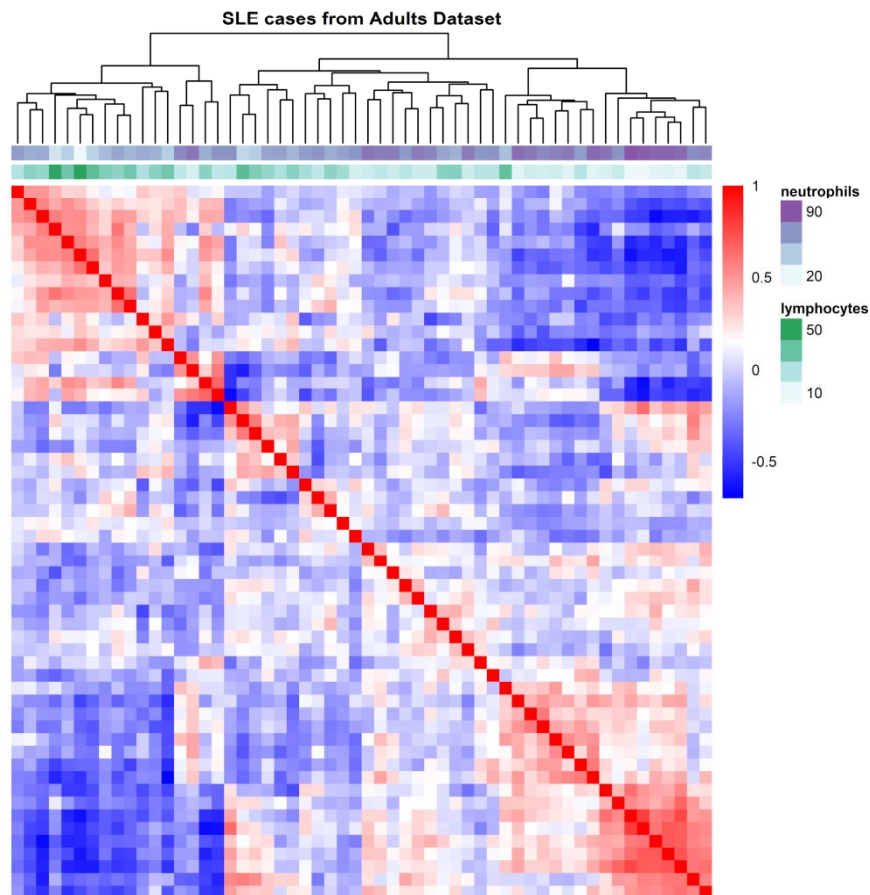
Isaac. Nature Reviews Genetics 2009

Systemic lupus erythematosus

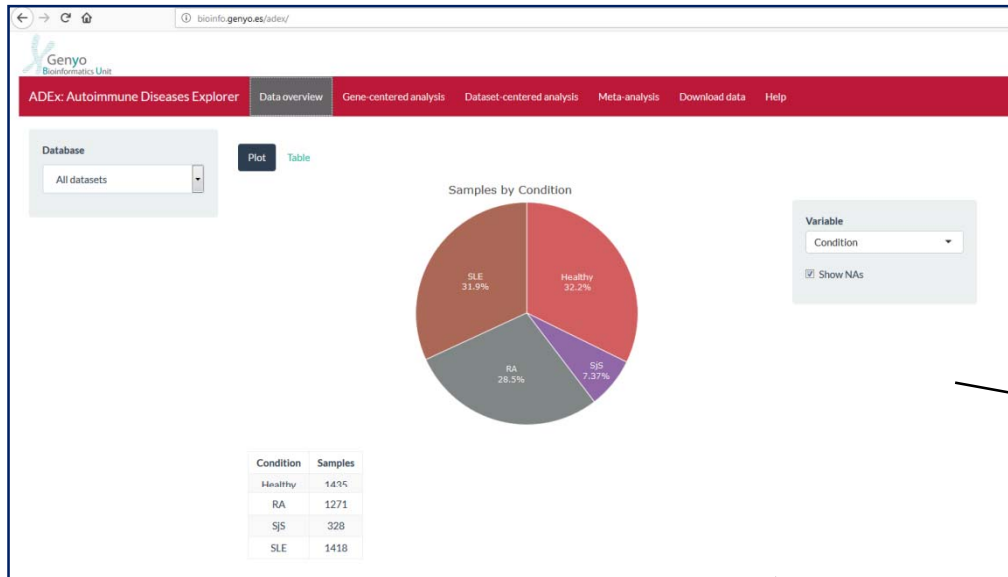
Dataset		Dataset
**	**	MYC
**	**	RFX5
**	**	RFXANK
**	**	RFXAP
**	**	ARNTL
**	**	SMAD1
**	**	WT1
**	**	CLOCK
**	**	RELB
**	**	SPIB
**	**	TCF7L2
**	**	RBPJ
*	**	MITF
**	**	SREBF2
*	**	CREB1
*	*	ERG
*	*	POU2F2
*	*	...



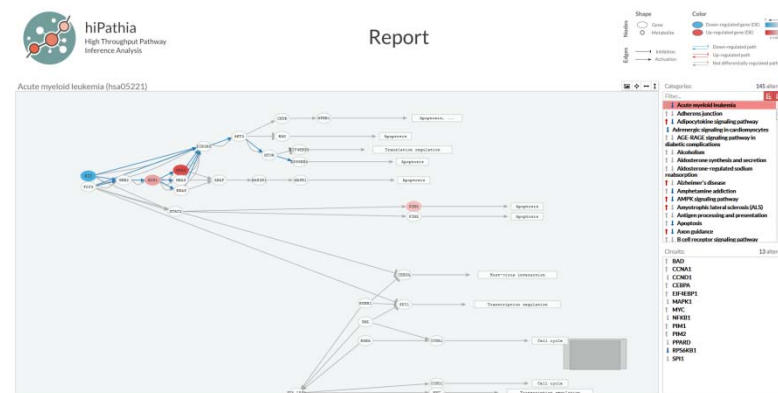
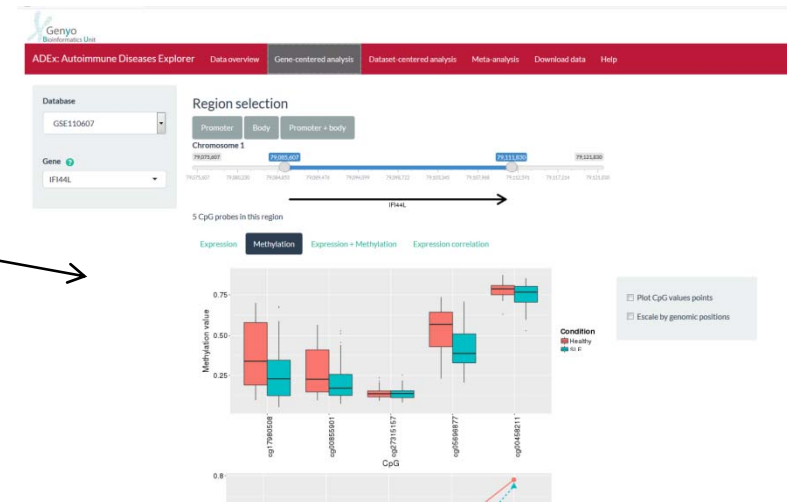
Systemic lupus erythematosus



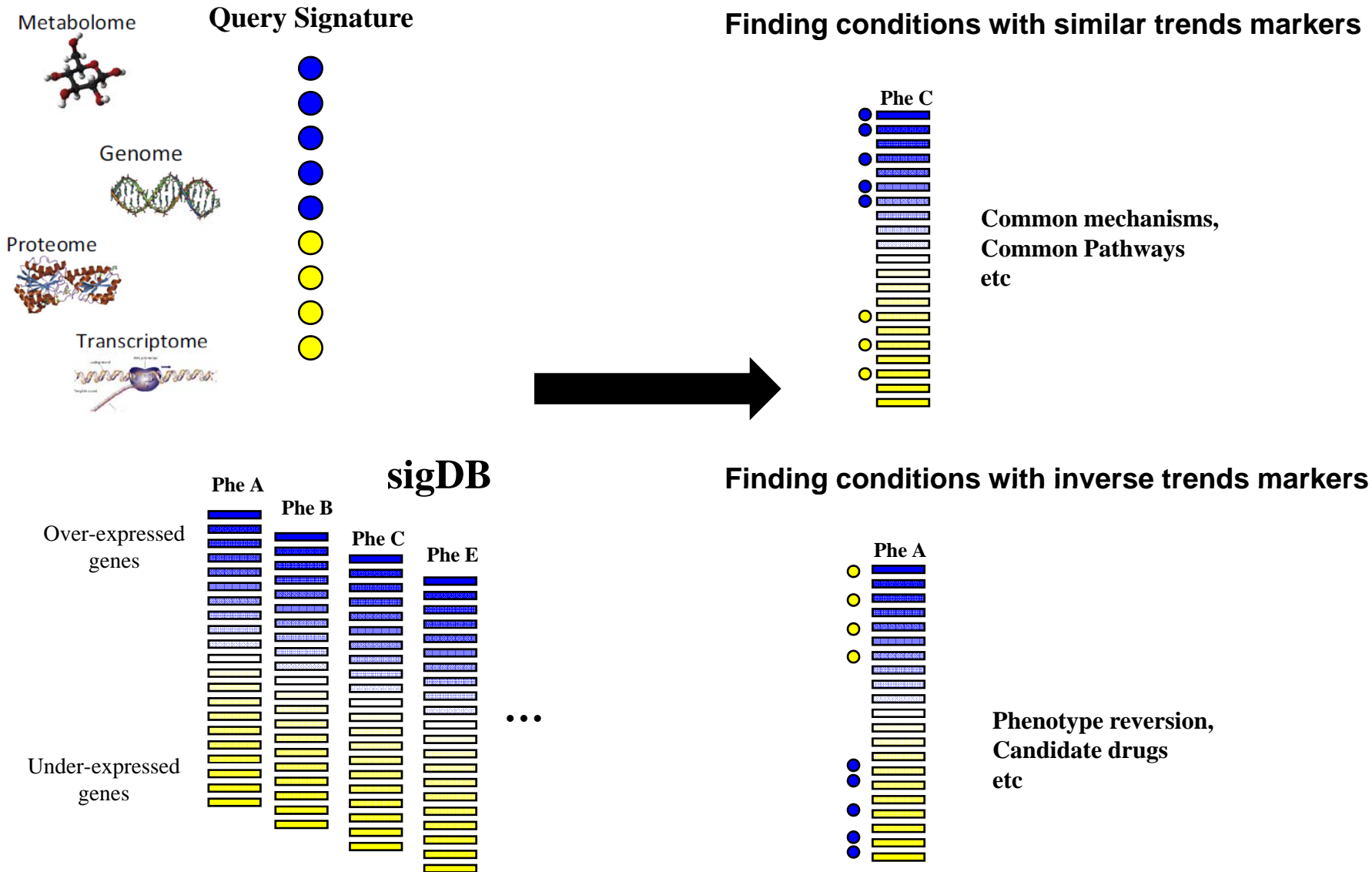
ADEX: Autoimmune Diseases Explorer



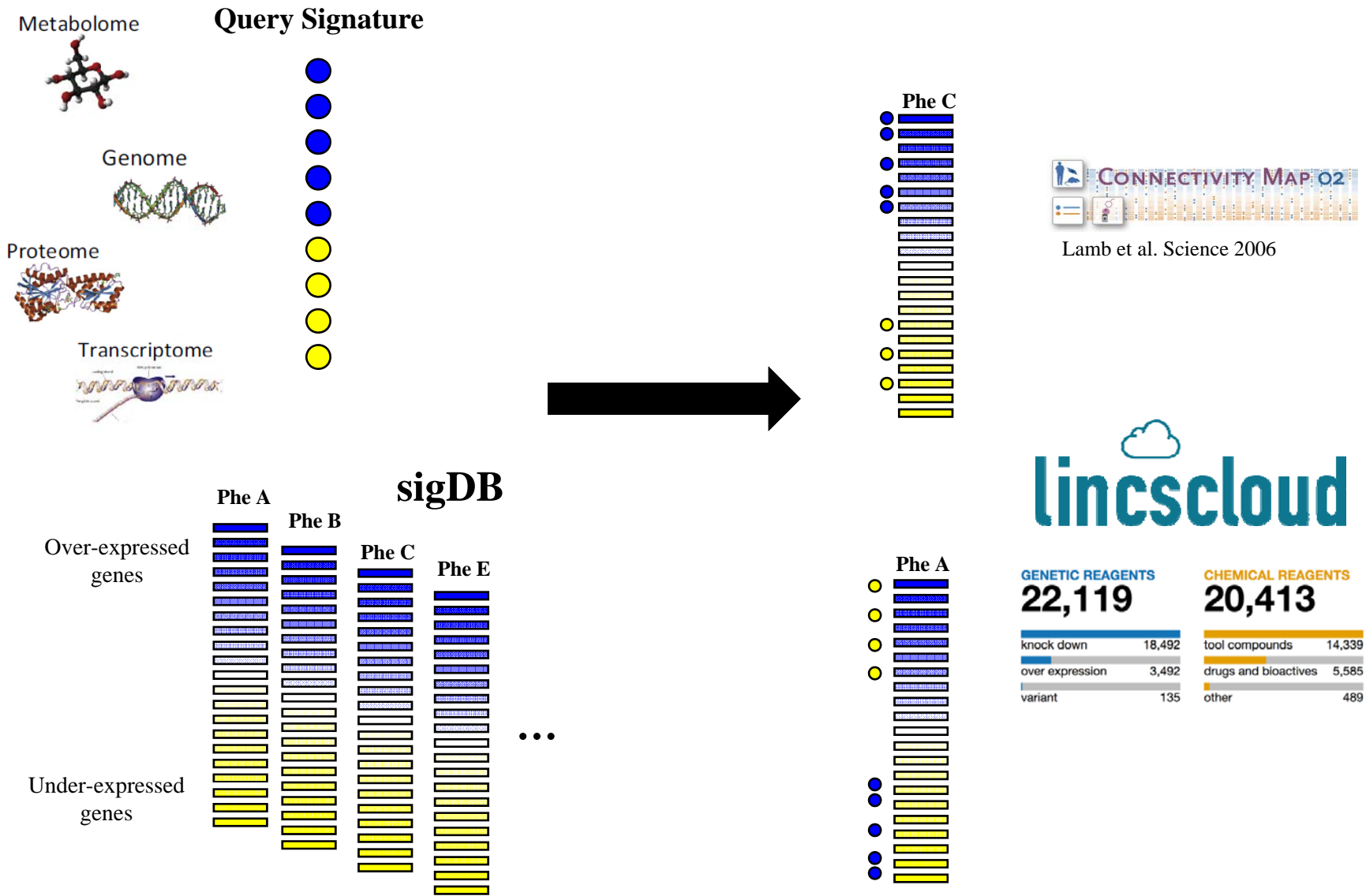
<http://bioinfo.genyo.es/adex/>



Integrating datasets: Connecting Phenotypes

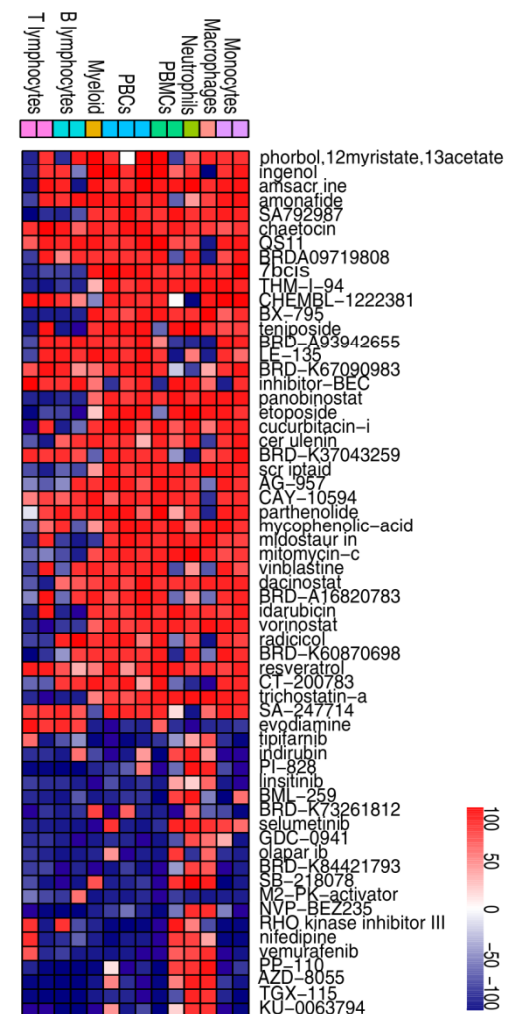
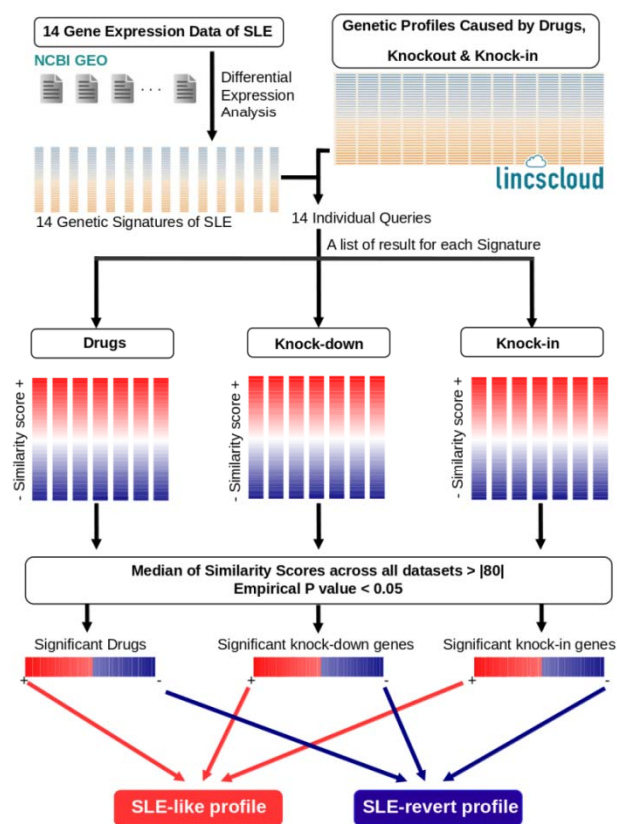


Integrating datasets: Connecting Phenotypes



Drug repurposing in SLE

We queried Lincscldb to connect compounds and SLE signatures. We obtained a list of drugs, knock-in and knockout genes with significant similarity scores with respect to the SLE signatures



Toro-Domínguez et al. Support for phosphoinositol 3 kinase and mTOR inhibitors as treatment for lupus using in-silico drug-repurposing analysis. *Arthritis Res Ther.* 2017

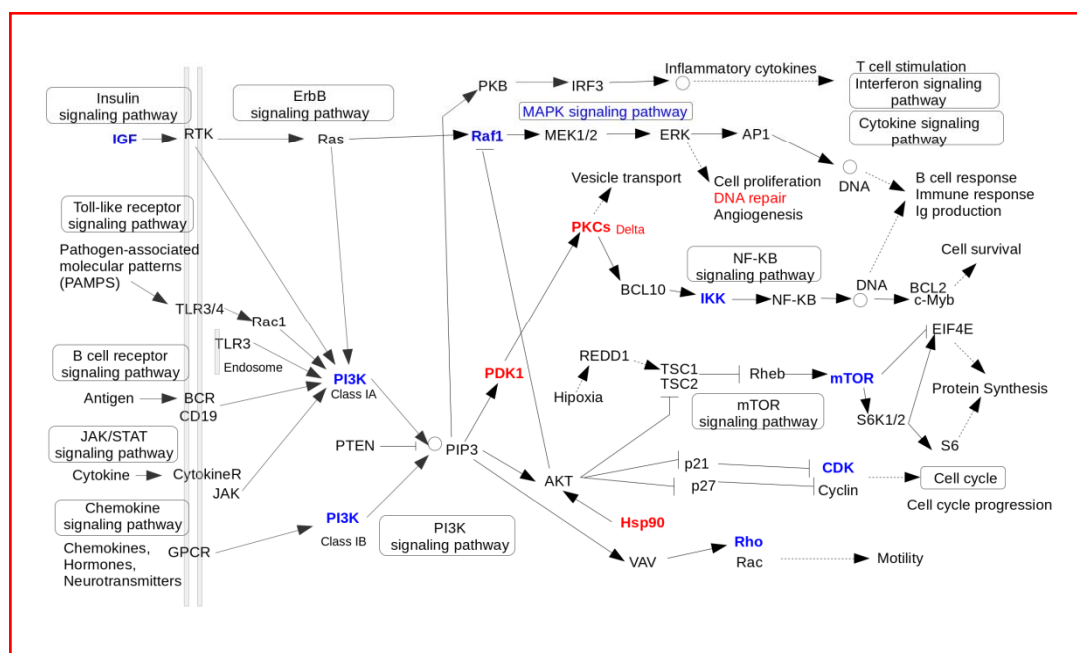
Drug repurposing in SLE

-	PI3K	Inhibitor	PI828 - GDC0941 - NVP-BEZ235 PP110 - TGX115	4.915E-06
-	mTOR	Inhibitor	NVP-BEZ235 - AZD8055 - TGX115 Ku0063794	1.792E-05
-	CDK	Inhibitor	BML259 - Indirubin	1.463E-02
-	IKBalfa	Inhibitor	Evodiamine	
-	Farnesyltransferase	Inhibitor	Tipifarnib	
-	IGF1R	Inhibitor	Linsitinib	
-	MAP2K1	Inhibitor	Selumetinib	
-	CHK1	Inhibitor	SB218078	
-	Piruvate kinase	Inhibitor	M2PK Activator	
-	Rho kinase	Inhibitor	Rho kinase inhibitor III	
-	Voltage dependent calcium channel	Inhibitor	Nifedipine	
-	Braf	Inhibitor	Vemurafenib	

Table 3: Significant knock-down and knock-in genes obtained.

Score	Type of experiment	Genes
+	Knock-in	<i>Ifinb1 - Ifng - Cd40 - Bcl10 - Klf6 - Lyn - Tyrap</i>
+	Knock-down	<i>Clen3 - Ppp1r14b - Lmb2 - Tbx2 - Pmm2 - Myc - Atp6v1f - Max - Pcpd - Puf60 - Phb2 - Akr1a1 - Btg1 - Abhd2 - Tfdp1 - Pax8 - Fosl2 - Nt5e - Rrm1 - Nr2f6 - Ramp1 - Ryk - Cish - Ppp2r1a - Cd14 - Ufdll1 - Htral - Slc35a1 - Twf2 - Nnt - Homer2 - Hs2st1 - Znf768 - Ggt1 - Dffb - Hspa2 - Prkdc - Arpc5 - Nfkbia - Slc39a8 - Thap11 - Gstp1 - Erv1 - Geat - Kiaa0907 - Dlx3 - Elk1 - Ptas4 - Meox2 - Gper - Nras - Tceb3c - Kif2c - Polr2f - Ctbp2 - Chaf1b - Cep55 - Hook2 - Znf8 - Ndufb7 - Nisch - Hoxc10 - Aap12a - Yes1 - Psmd5 - Jag1 - Mah2 - Polr2i - Ddfl - Hras - Hdac10 - Slc25a14 - Med7 - Hmger - Pdxp - Fdx1 - Nipbl - Prkag3 - Ppia - Eif2ak3 - B4galt1 - Uck2 - Jun - Med4 - Ybx1 - Bub1b - Crep - Med1 - Hdac11 - Sbnol</i>
-	Knock-down	<i>Mitf - Etf2 - Pip4k2b - Vrk2 - Spen - Nsdhl - Znf586 - Gnpd1 - Six4 - Parn - Dusp14 - Iqgap1 - Lrrk2 - Gpr123 - Sfl - Fez2 - Ipmk - Sat1 - Elf4 - Rptor - Eif4e - Axl3 - Kars - Csnk1a1 - Sptlc2 - Men1 - Snx17 - Vegfc - Ppp3ca - Bnip3 - Erbb3 - Erol1 - Copb2 - Serpine1 - Ak4 - Hla_a (Pik3ca - <u>Pik3c2a</u>) - Igf2r - Iypla1 - Str4 - Atm - Esp11 - Igf1r - St3gal5 - Mtor - Grn - Hsp90aa1 - Prpf4b - Tm9sf3</i>

PI3K molecular signaling pathway



Toro-Domínguez et al. Support for phosphoinositol 3 kinase and mTOR inhibitors as treatment for lupus using in-silico drug-repurposing analysis. *Arthritis Res Ther.* 2017

OPEN ACCESS

ESSAY

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • <https://doi.org/10.1371/journal.pmed.0020124>

- Article
- Authors
- Metrics
- Comments
- Media Coverage

Abstract

Modeling the Framework for False Positive Findings
Bias
Testing by Several Independent Teams
Corollaries
Most Research Findings Are False for Most Research Designs and for Most Fields
Claimed Research Findings May Often Be Simply Accurate Measures of the Prevailing Bias

Abstract

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

68,836 Save	3,556 Citation
2,605,883 View	10,479 Share

Download PDF 

Print Share

 Check for updates

Related PLOS Articles

- Why Current Publication Practices May Distort Science
- Why Most Published Research Findings Are False: Author's Reply to Goodman and Greenland
- Why Most Published Research Findings Are False: Problems in the Analysis
- Most Published Research



John P. A. Ioannidis

Professor of Medicine, [Stanford University](#) (previously at U Ioannina, Greece)

Verified email at stanford.edu - [Homepage](#)

[Evidence-based medicine](#) [research methods](#) [meta-analysis](#) [clinical epidemiology](#) [genetic epidemiology](#)

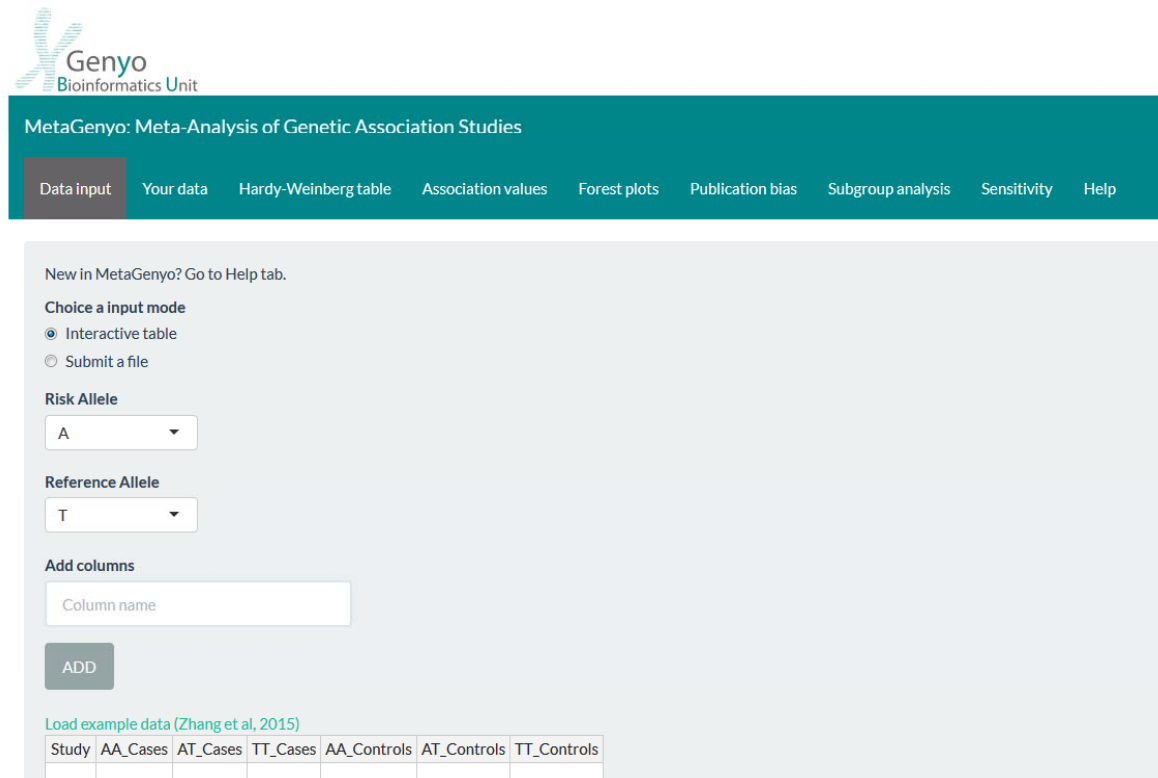
Cited by

[VIEW ALL](#)

	All	Since 2013
Citations	181686	132586
h-index	173	131
i10-index	794	674

MetaGenyo: Meta-Analysis of Genetic Association Studies

<http://bioinfo.genyo.es/metagenyo/>



Genyo
Bioinformatics Unit

MetaGenyo: Meta-Analysis of Genetic Association Studies

Data input | Your data | Hardy-Weinberg table | Association values | Forest plots | Publication bias | Subgroup analysis | Sensitivity | Help

New in MetaGenyo? Go to Help tab.

Choice a input mode

Interactive table

Submit a file

Risk Allele

A

Reference Allele

T

Add columns

Column name

ADD

[Load example data \(Zhang et al, 2015\)](#)

Study	AA_Cases	AT_Cases	TT_Cases	AA_Controls	AT_Controls	TT_Controls

Martorell-Marugan J, Toro-Dominguez D, Alarcon-Riquelme ME, Carmona-Saez P. *MetaGenyo: A web tool for meta-analysis of genetic association studies*. **BMC Bioinformatics**. 2017

MetaGenyo: Meta-Analysis of Genetic Association Studies

<http://bioinfo.genyo.es/metagenyo/>

MetaGenyo: Meta-Analysis of Genetic Association Studies

Data input | Your data | Hardy-Weinberg table | Association values | **Forest plots** | Publication bias | Subgroup analysis | Sensitivity | Help

Select comparison

- Allele contrast (A vs. T)
- Recessive model (AA vs. AT+TT)
- Dominant model (AA+AT vs. TT)
- Overdominant model (AT vs. AA+TT)
- AA vs. TT
- AA vs. AT
- AT vs. TT

Select model

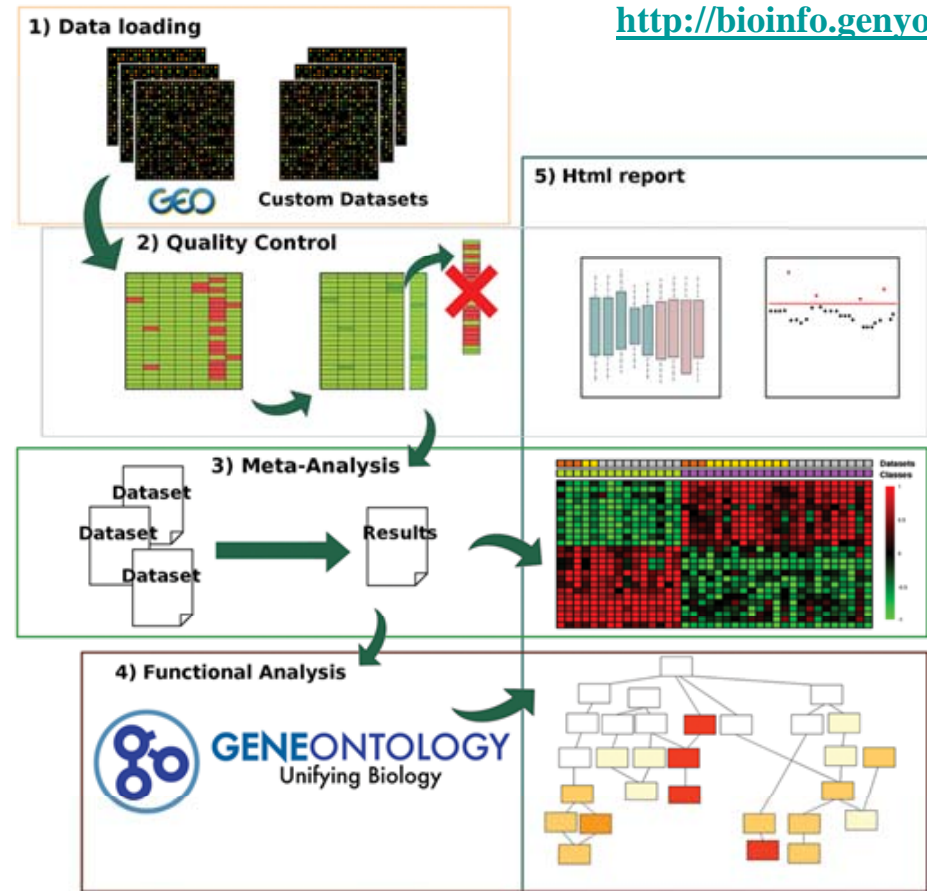
- Both
- Fixed Effect Model
- Random Effect Model

Study	Experimental Events	Control Total	Experimental Total	Control Total	Odds Ratio	OR	95%-CI	W(fixed)	W(random)
Bo	180	416	148	380		1.20	[0.90; 1.59]	3.5%	4.0%
Canedo	275	666	627	1386		0.85	[0.71; 1.03]	7.9%	5.1%
Crusius	209	472	1074	2278		0.89	[0.73; 1.09]	7.0%	4.9%
de Oliveira	192	414	224	480		0.99	[0.76; 1.29]	4.0%	4.2%
Felipe	88	208	189	392		0.79	[0.56; 1.11]	2.4%	3.5%
Garcia-Gonzalez	79	156	153	378		1.51	[1.04; 2.20]	2.0%	3.2%
Kamali-Sarvestani	24	38	118	306		-2.73	[1.36; 5.49]	0.6%	1.4%
Kamangar	84	224	159	414		0.98	[0.69; 1.35]	2.5%	3.5%
Kang	257	668	202	644		1.37	[1.09; 1.72]	5.3%	4.6%
Ko	59	162	200	616		1.19	[0.83; 1.71]	2.1%	3.3%
Lee	331	940	262	616		0.73	[0.60; 0.90]	6.4%	4.8%
Liu	135	276	102	274		1.61	[1.15; 2.27]	2.4%	3.5%
Lu	210	500	218	600		1.27	[0.99; 1.62]	4.7%	4.4%
Ohyauchi	132	424	158	692		1.53	[1.16; 2.00]	3.8%	4.1%
Pan	264	616	266	616		0.99	[0.79; 1.24]	5.4%	4.6%
Qadri	92	260	118	400		1.31	[0.94; 1.83]	2.5%	3.5%
Ramis	9	18	34	76		1.24	[0.44; 3.46]	0.3%	0.7%
Savage (a)	85	176	357	858		1.31	[0.95; 1.81]	2.6%	3.6%
Savage (b)	292	574	439	856		0.98	[0.80; 1.22]	6.2%	4.8%
Shirai	118	362	306	936		1.00	[0.77; 1.29]	4.1%	4.3%
Song	112	250	116	280		1.15	[0.81; 1.62]	2.3%	3.4%
Taguchi	279	792	149	504		1.30	[1.02; 1.65]	4.8%	4.5%
Vinagre	106	204	80	206		1.70	[1.15; 2.52]	1.8%	3.0%
Ye	116	306	132	412		1.30	[0.95; 1.77]	2.9%	3.7%
Zeng	228	412	192	392		1.29	[0.98; 1.70]	3.6%	4.1%
Zhang	517	1038	437	1008		1.30	[1.09; 1.54]	9.1%	5.2%
Fixed effect model		10572		16000		1.12	[1.06; 1.18]	100%	--
Random effects model						1.16	[1.05; 1.27]	--	100%

Heterogeneity: I-squared=85.7%, tau-squared=0.0363, p<0.0001

Martorell-Marugan J, Toro-Dominguez D, Alarcon-Riquelme ME, Carmona-Saez P. *MetaGenyo: A web tool for meta-analysis of genetic association studies.* BMC Bioinformatics. 2017

IMAGEO : Integrative Meta-Analysis from GEO Data



D. Toro-Domínguez, J. Martorell-Marugán, R. López-Dominguez, A. García-Moreno, V. González-Rumayor, M. E Alarcón-Riquelme, P. Carmona-Sáez. *ImaGEO: Integrative Gene Expression Meta-Analysis from GEO database*. **Bioinformatics** 2018

IMAGEO: Integrative Meta-Analysis from GEO Data



<http://bioinfo.genyo.es/imageo/>

ImaGEO: Integrative Meta-Analysis of GEO Data Step 1: Data Input Step 2: Assign samples to each group Report Help

Input

Enter GEO IDs (one ID per line, maximum 10)

[Load examples](#)

And/or upload your own data

Only tab-separated files (.txt and .tsv) are admitted

[Download a sample file \(GPL570\)](#)

Browse... No file selected

Analysis parameters

Meta-analysis method Effect size P-value

Select a model for effect size estimation

Fixed effect model Random effect model

Allowed missing values (%)

Adjusted P-value threshold

Group 1 name

Group 2 name

Functional analysis

Email

If you use ImaGEO, please include this reference:

D. Toro-Domínguez, J. Martorell-Marugán, R. López-Domínguez, A. García-Moreno, V. González-Rumayor, M. E Alarcón-Riquelme, P. Carmona-Sáez. *ImaGEO: Integrative Gene Expression Meta-Analysis from GEO database*. **Bioinformatics** 2018

WE NEED LONG TERM INITIATIVES SUSTAINABLE RESEARCH

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis Repository

Quick Search Manage Sets Login Cart 0 GDC Apps

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

Search: e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary

Data Release 13.0 - September 27, 2018

Category	Count
PROJECTS	43
FILES	358,092
PRIMARY SITES	69
GENES	22,147
CASES	33,096
MUTATIONS	3,142,246

Cases by Major Primary Site

Primary Site	Cases
Adrenal Gland	~100
Bile Duct	~100
Bladder	~100
Blood	~100
Bone	~100
Bone Marrow	~100
Brain	~100
Breast	~3,500
Cervix	~100
Colorectal	~2,500
Esophagus	~100
Eye	~100
Head and Neck	~100
Kidney	~100
Liver	~100
Lung	~4,500
Lymph Nodes	~4,000
Nervous System	~100
Ovary	~100
Pancreas	~100
Pleura	~100
Prostate	~100
Skin	~100
Soft Tissue	~100
Stomach	~100
Testis	~100
Thymus	~100
Thyroid	~100
Uterus	~100

GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

- Data Portal
- Website
- Data Transfer Tool
- API
- Data Submission Portal
- Documentation
- Legacy Archive

Bioinformatics Week in Granada

<http://jbi2018.ugr.es/>

XIV SYMPOSIUM ON BIOINFORMATICS

GRANADA 14TH-16TH NOVEMBER, JBI 2018.

<http://oncothon.ptsgranada.com/>



The image shows a website banner for the Oncothon event. At the top, a dark blue navigation bar contains the following menu items: Home, About, Challenges, Program, Inscription, Venue, People, and Partners. Below the navigation bar, the main content area features a world map on the left with various icons representing different regions and topics. The text "BENEFICIARIES OF THE PROJECT" is centered below the map. On the right side, the word "ONCOTHON" is written in large, bold, orange letters. To its right is the logo for "Interreg Sudoe ONCONET SUDOE", which includes a stylized yellow and blue icon. Below the "ONCOTHON" text, the dates "12th – 13th November" and the location "PTS GRANADA" are displayed in orange. At the bottom of the banner, there is a row of logos for various partner organizations, including PTS, Universidad de Granada, Universidad de Sevilla, Biocat, CMO, ICO, and NAVARRABOMED.



Dr. Pedro Carmona

Jordi Martorell

Daniel Toro

Raúl López

Adrian García

Dr. Marta Alarcon. GENyO

Dr. Julio Saez-Rodriguez. U. of Heidelberg

Dr. Michelle Petri. Johns Hopkins Lupus Center

Dr. Joaquín Dopazo. Bioinformatics Area. FPS



Follow us in twitter:

[@pcarmonasaez](#)

[@BioinfoGenyo](#)



The problem: Characterization of Gen2.2

Dendritic Cells are antigen presenting cells with a key role in the immune response

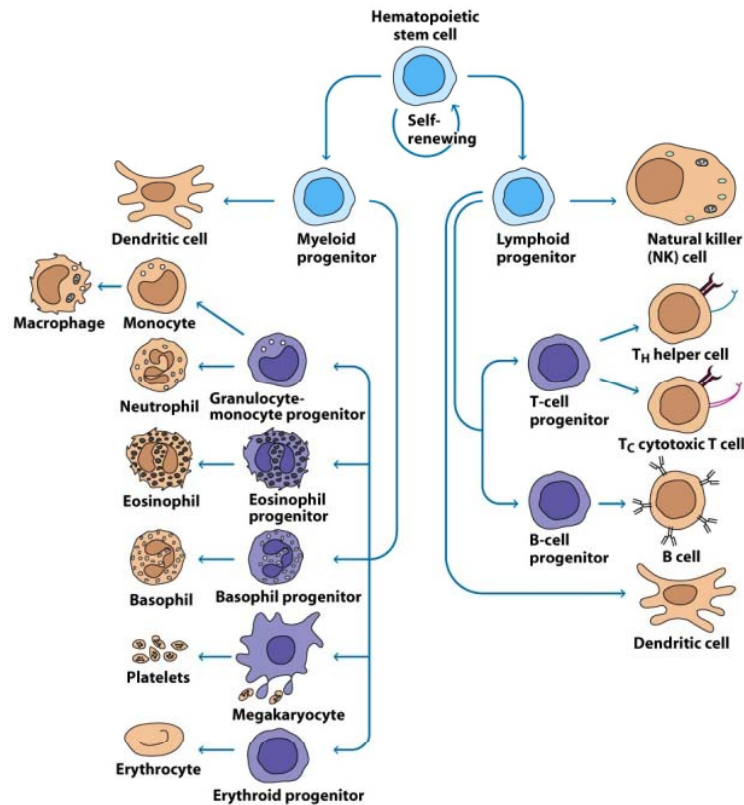


Figure 2-2
Kuby IMMUNOLOGY, Sixth Edition
© 2007 W. H. Freeman and Company

Various DC precursors, such as DC1 (myeloid origin) or plasmacytoid dendritic cells (pDCs).

pDC are difficult to isolate (less than 0.5% of the circulating cells). Therefore, good cells models are required

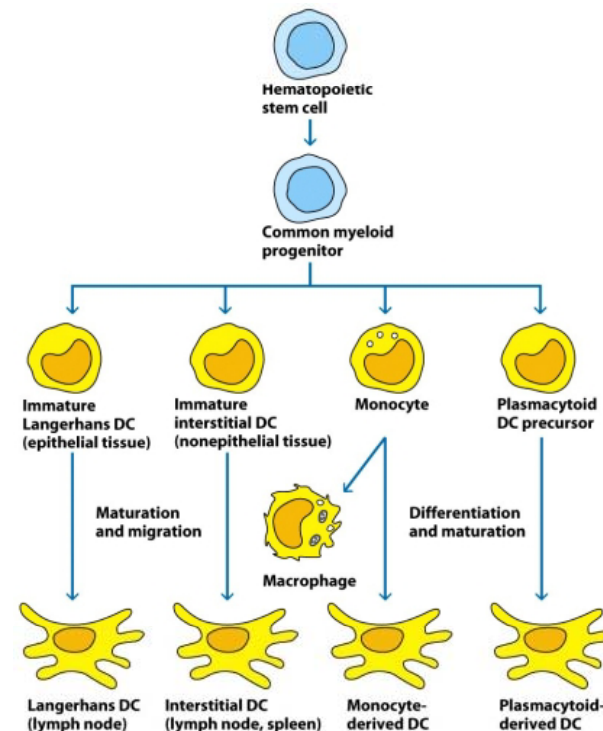


Figure 2-10
Kuby IMMUNOLOGY, Sixth Edition
© 2007 W. H. Freeman and Company

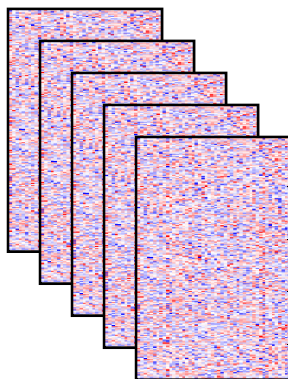
Connecting Phenotypes: Similar profiles

Compare gene expression profiles of Gen2.2 with previously published immune system gene expression datasets to get insights into similarities among cell lines



Gene expression compendium of Immune cell populations

GSE12507
GSE15215
GSE28490
GSE28491
GSE35457



Rank-based Normalization
Join Datasets by Common Genes
Combined Analysis

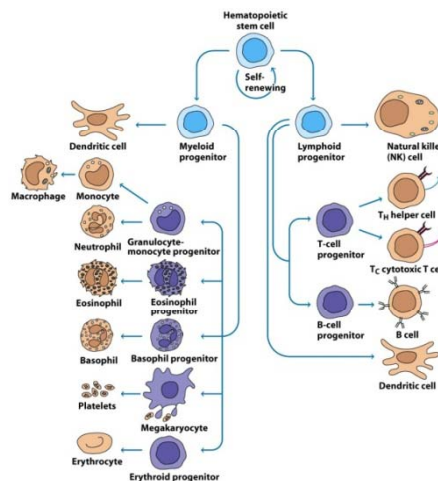
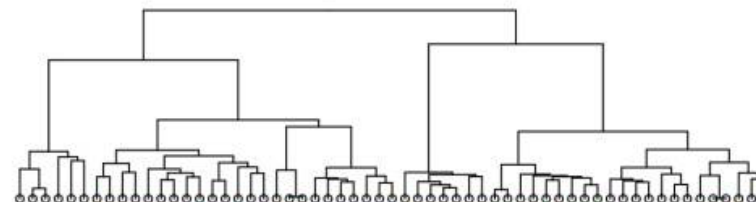
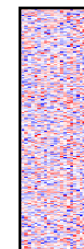
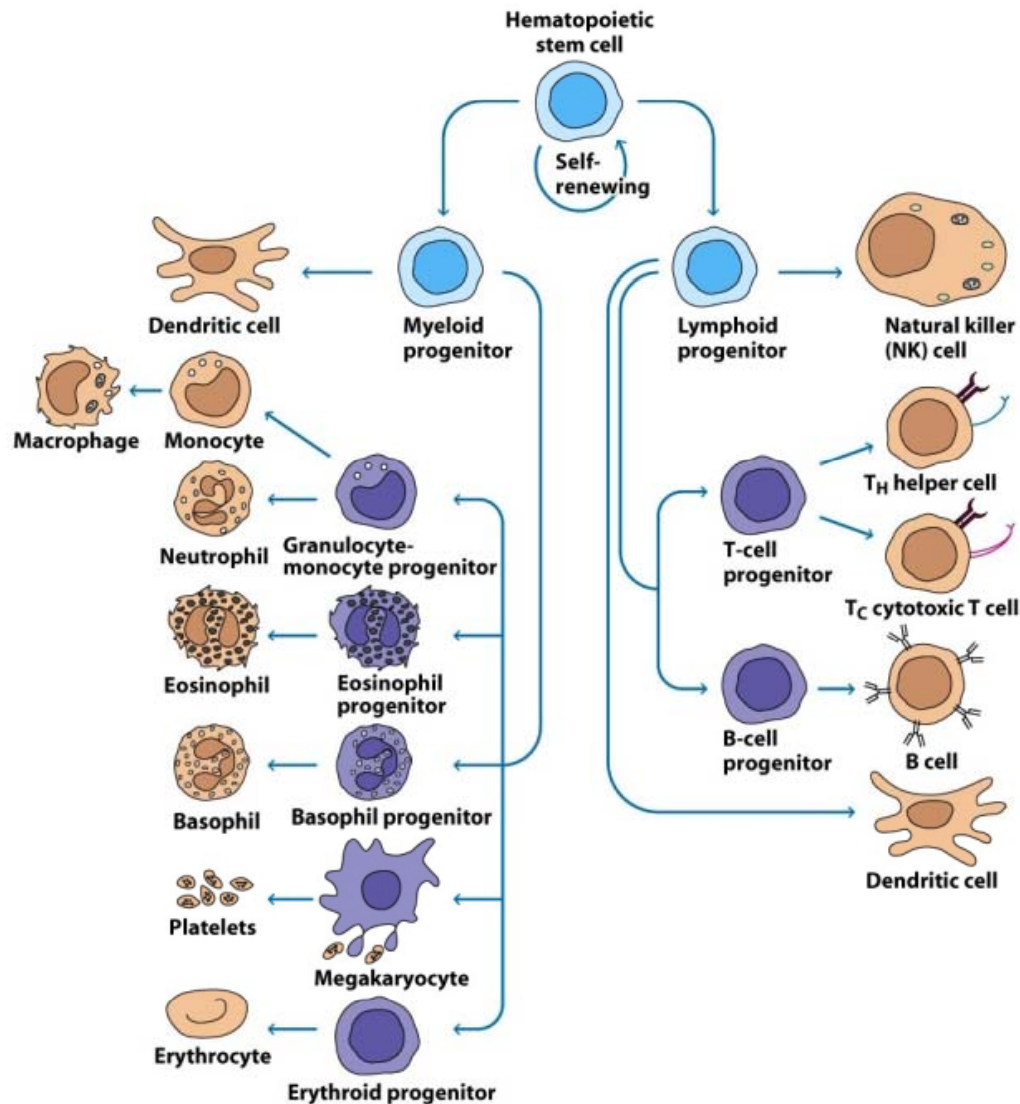


Figure 2-2
Ruby IMMUNOLOGY Sixth Edition
© 2007 W. H. Freeman and Company

Gene 2.2 expression profile
Three replicates Illumina HT-V4

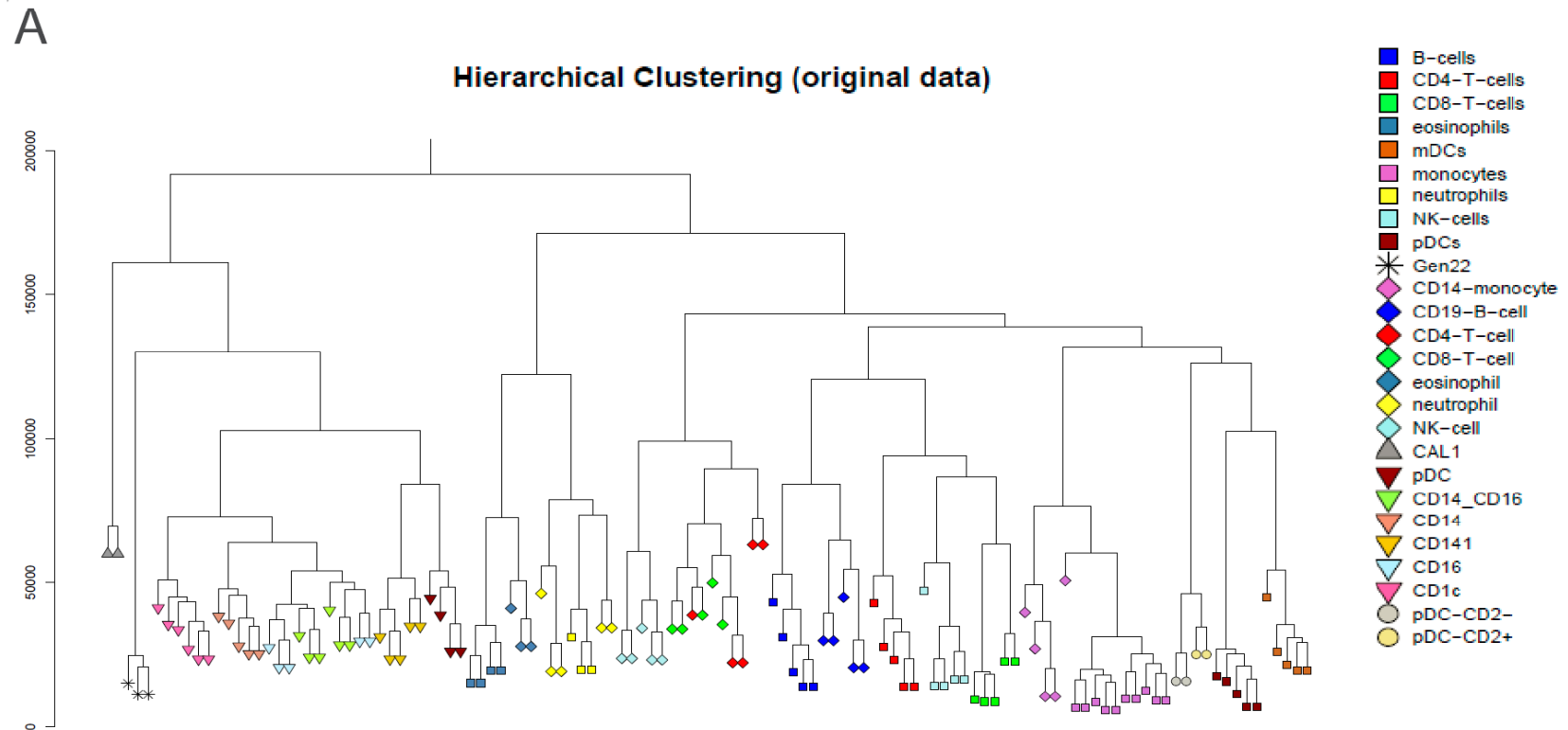


Connecting Phenotypes: Similar profiles



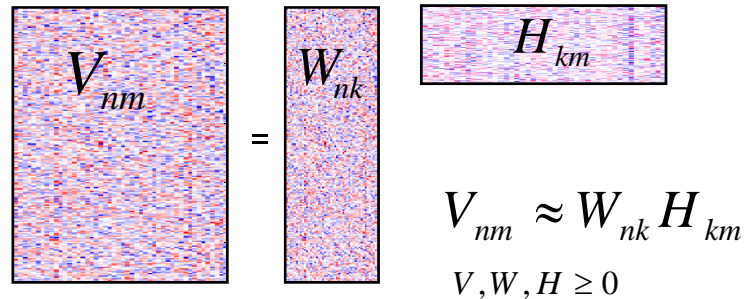
GEO Datasets	Cell Types
GSE28490	Monocytes
GSE28491	B cells
	CD4+ T cells
	NK cells
	CD8+ T cells
	Eosinophils
	mDC
	Neutrophils
	pDC
GSE12507	pDC cell line (CAL1)
	immature T cell line (MOLT4)
GSE15215	pDC CD2-
	pDC CD2+
GSE35457	pdc
	cd14_cd16-mono
	cd14-mono
	cd141-dc
	cd16-mono
	cd1c-dc

Connecting Phenotypes: Similar profiles



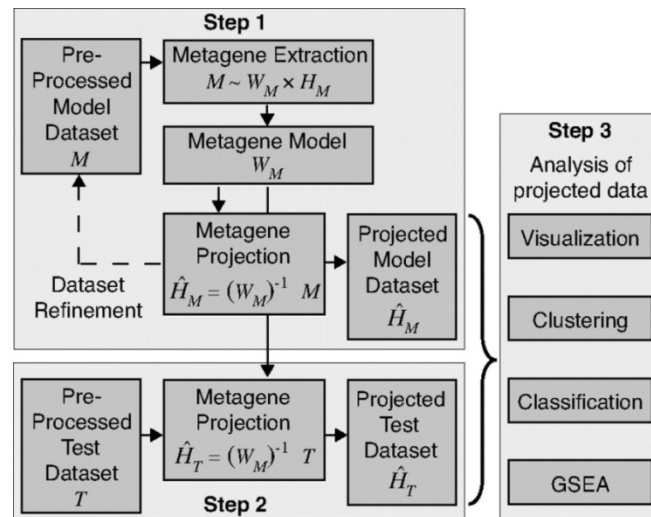
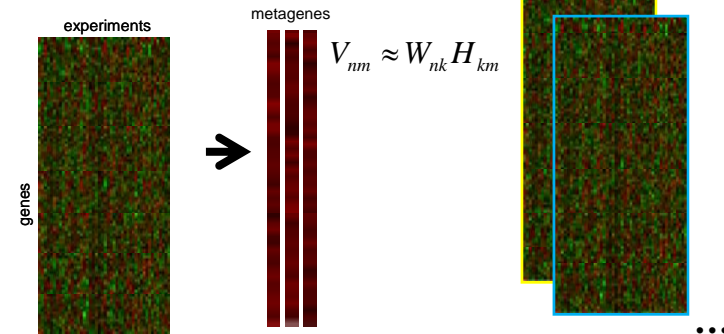
Carmona-Sáez P, et al. Metagene projection characterizes GEN2.2 and CAL-1 as relevant human plasmacytoid dendritic cell models. **Bioinformatics**. 2017

Connecting Phenotypes: Similar profiles



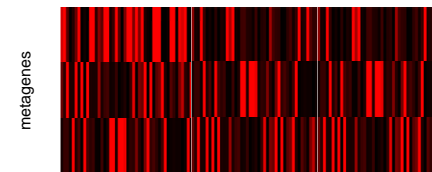
Model Dataset

Test Datasets



$$H_{km} = (W_{nk})^{-1} V_{nm}$$

All Samples



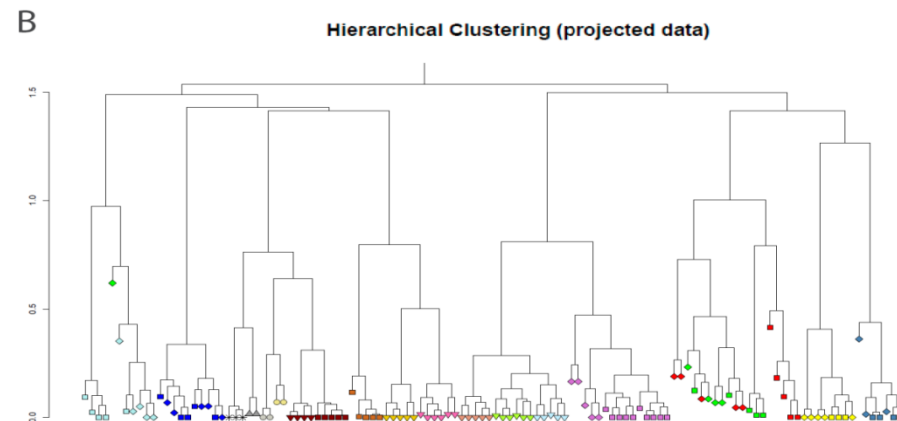
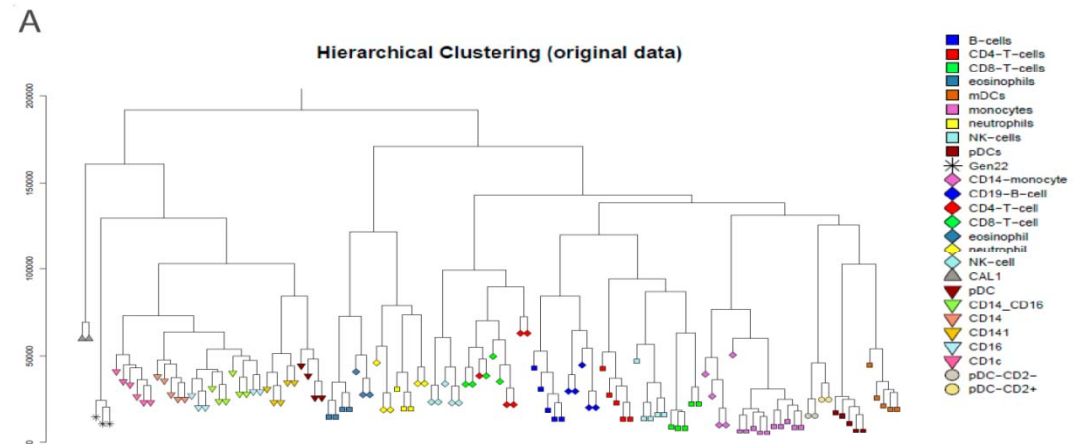
Metagene projection for cross-platform, cross-species characterization of global transcriptional states

Pablo Tamayo*, Daniel Scandfeld*, Benjamin L. Ebert*, Michael A. Gillette*¹, Charles W. M. Roberts², and Jill P. Mesirov*⁵

*Eli and Edythe L. Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02141; ¹Pulmonary and Critical Care Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114; and ²Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115

Communicated by Edward M. Scolnick, The Broad Institute, Cambridge, MA, February 6, 2007 (received for review December 7, 2006)

Connecting Phenotypes: Similar profiles



Carmona-Sáez P, et al. Metagene projection characterizes GEN2.2 and CAL-1 as relevant human plasmacytoid dendritic cell models. **Bioinformatics. 2017**