Functional synergies between profiles of T cell responses to chronic infections and BET bromodomain inhibition across pediatric malignancies

Gabriela Alexe, Ph.D.

Dana-Farber Cancer Institute
Harvard Medical School
Broad Institute

February 13th, 2014

Overview

1

- Metagene Projection Techniques for discovering functional dimensions in large scale and micro scale biological data

2

- JQ1 - a BET bromodomain inhibitor targets MYCN in NeuroBlastoma (Stegmaier Lab, DFCI)

3

- CD8 tetramer specific T Cell transcriptional responses in chronic HIV and HCV (Haining Lab, DFCI)

Knowledge Discovery from Biological Datasets:

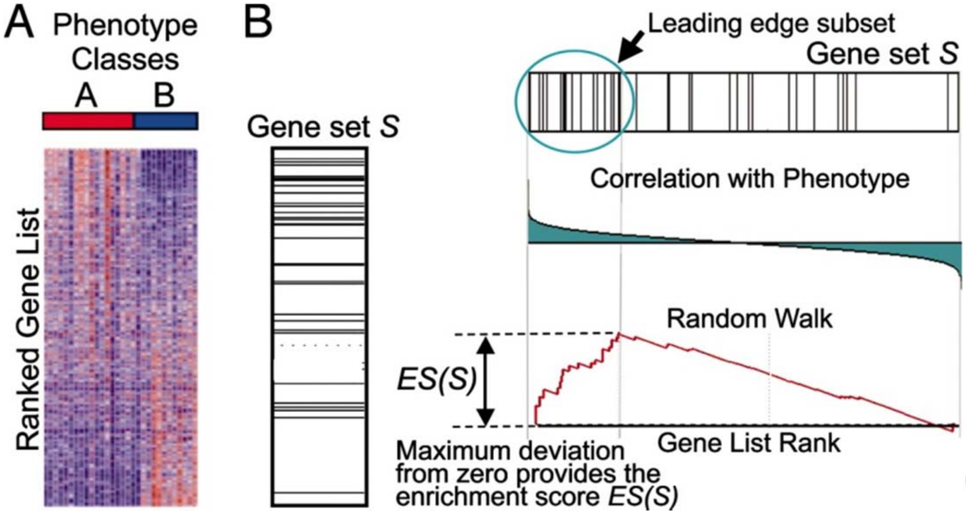
Evaluate data at different levels of **functional** complexity

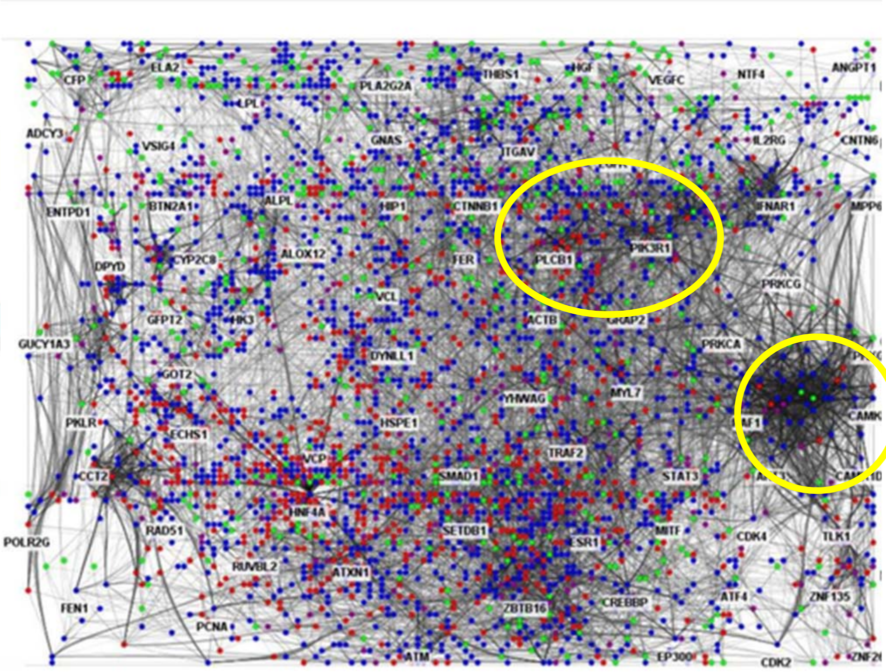
Genes

Pathways

Networks: Gene hubs

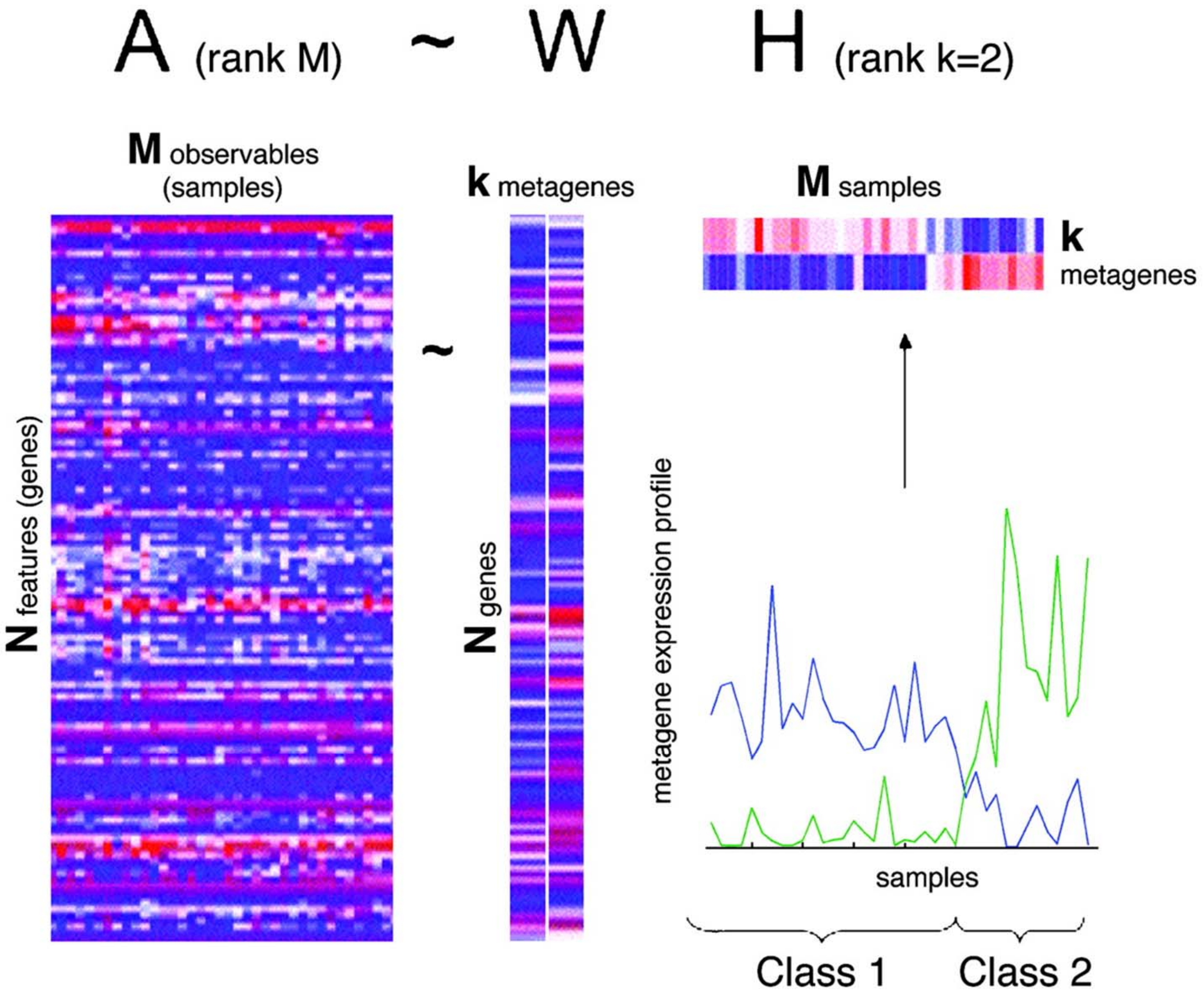
GSEA





Subramanian A et al. PNAS 2005;102:15545-15550

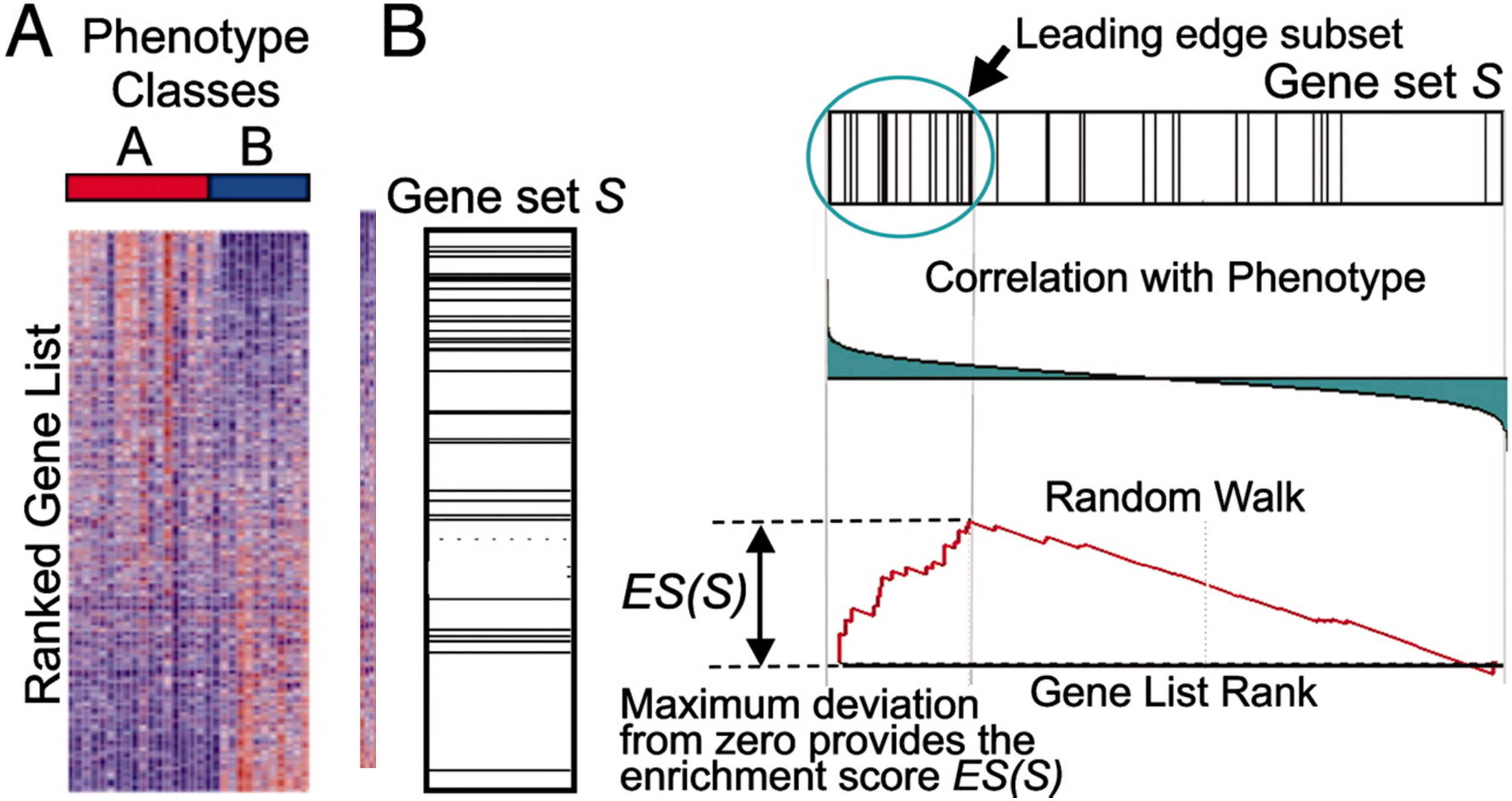
Metagene Projection Strategy:
Gene expression data described in a simple way:



Brunet J et al. PNAS 2004;101:4164-4169

(1) Single Sample Gene Set Enrichment Analysis (ssGSEA*)

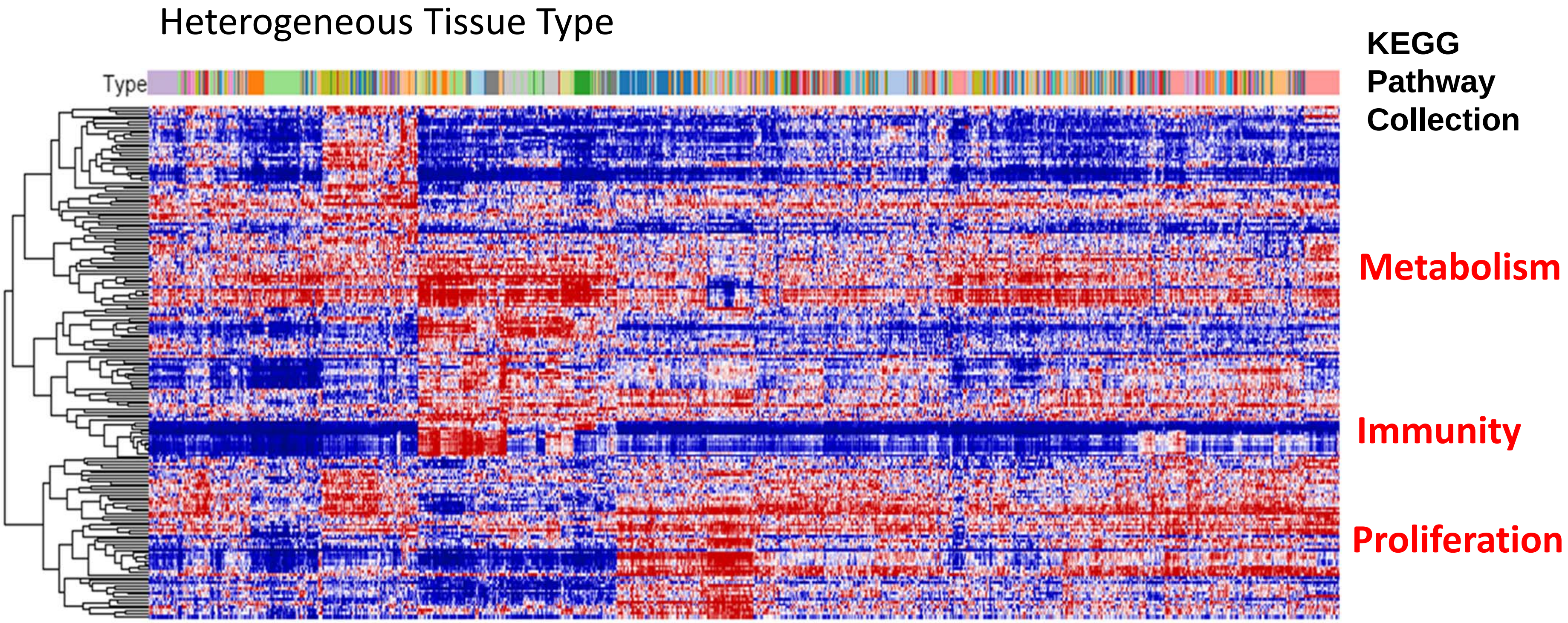
A GSEA overview illustrating the method.



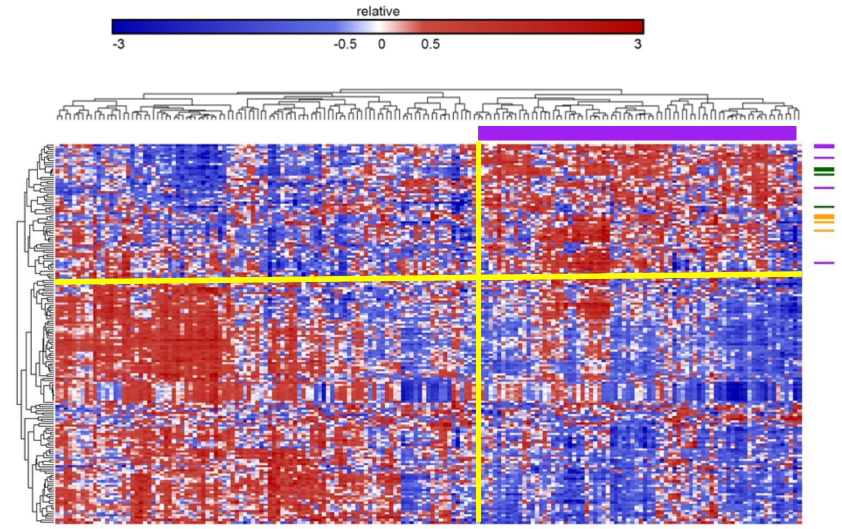
GSEA: Subramanian A et al. PNAS 2005;102:15545-15550

* **ssGSEA: Barbie *et al, Nature* 2009: available in GenePattern / Projections**

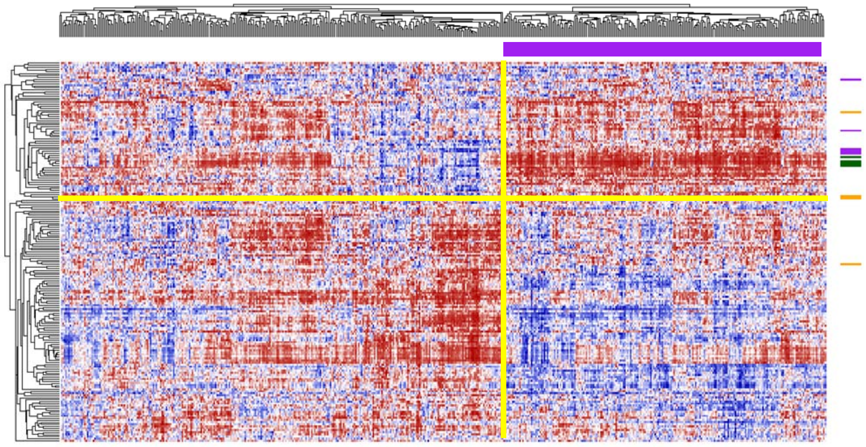
ssGSEA applied to Collection of Cell Line Encyclopaedia (CCLE) data: Identifies similarities of functional enrichment across multi-tissue cell lines



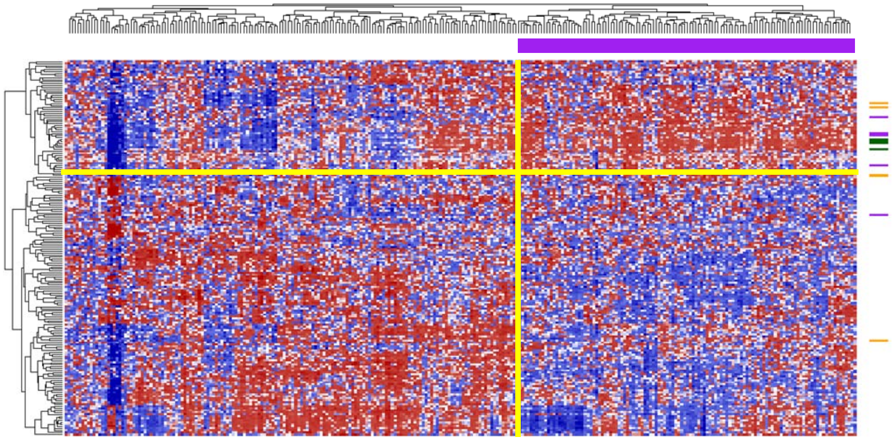
Metagene Projection Analysis: *Single Sample GSEA*
ssGSEA identifies a robust “metabolic” cluster in primary tumor and cell line AML



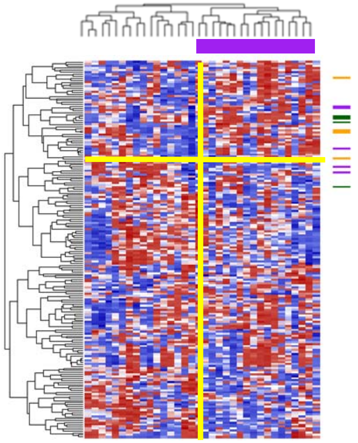
TCGA_LAML: 179 samples RNASeq



Wouters: 526 samples Affy U133A



Tomasson: 279 samples Affy U133A



CCLE_AML

Pathways Up-regulated in the Metabolic Cluster

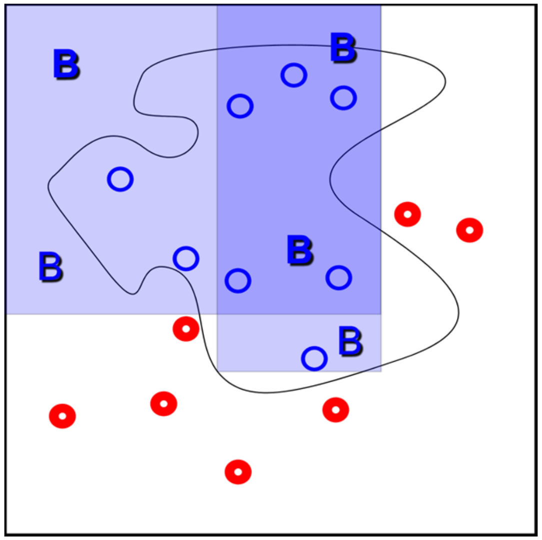
NUCLEOTIDE METABOLISM
OXIDATIVE PHOSPHORYLATION

FAT METABOLISM

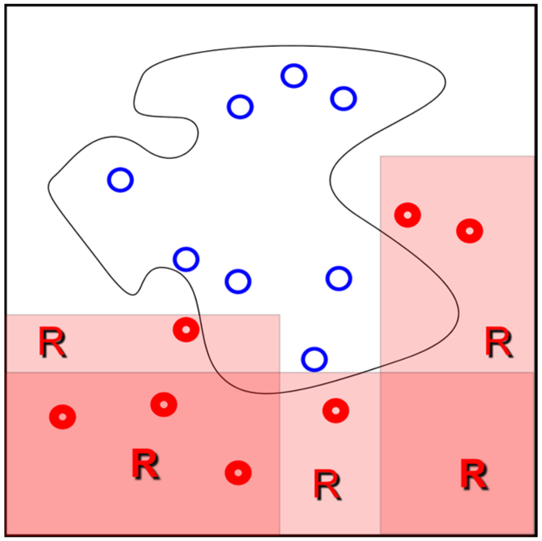
CELL_CYCLE
DNA_REPLICATION
MISMATCH_REPAIR
NUCLEOTIDE_EXCISION_REPAIR

(2) Logical Analysis of Data:
Supervised Classification Models based on Combinatorial Patterns*

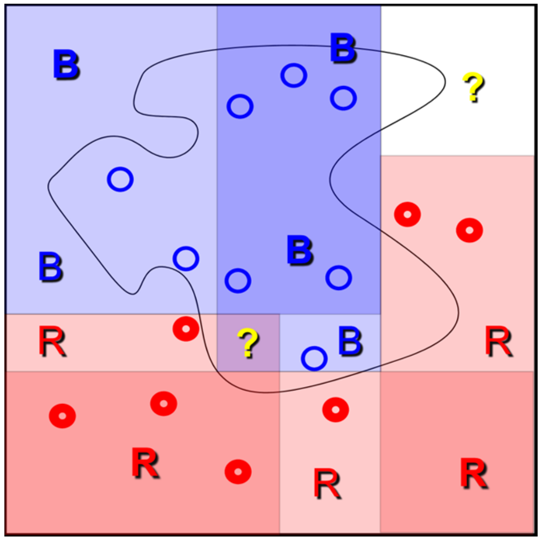
Positive patterns



Negative patterns



Model



Parameters:

Degree (# of defining markers)

Prevalence (# +/- covered cases)

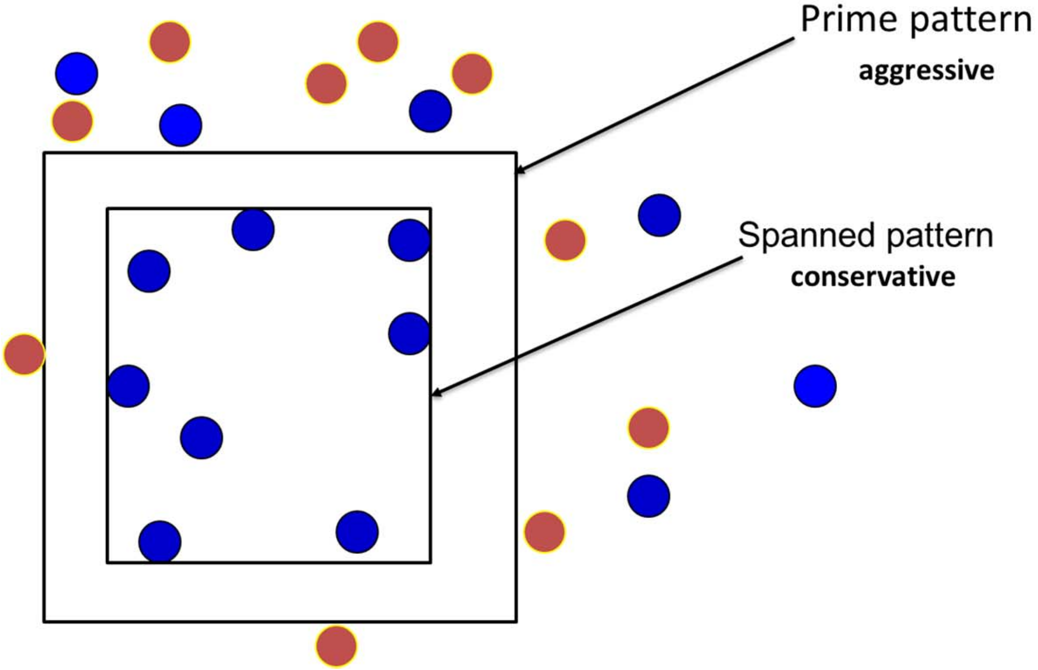
Homogeneity (proportion of +/- cases)

Significance (sample permutation p-value)

Reliability (robustness to validation)

- Alexe G. Combinatorial enhancements of Logical Analysis of Data and Biomedical Applications, 2004.
- Alexe G. et al .Comprehensive vs. comprehensible classifiers in Logical Analysis of Data. Discrete Applied Mathematics 2008,

Prime and spanned patterns



δ-patterns: discover differentially expressing genes between two phenotypes, identification of co-regulated genes on biological pathways

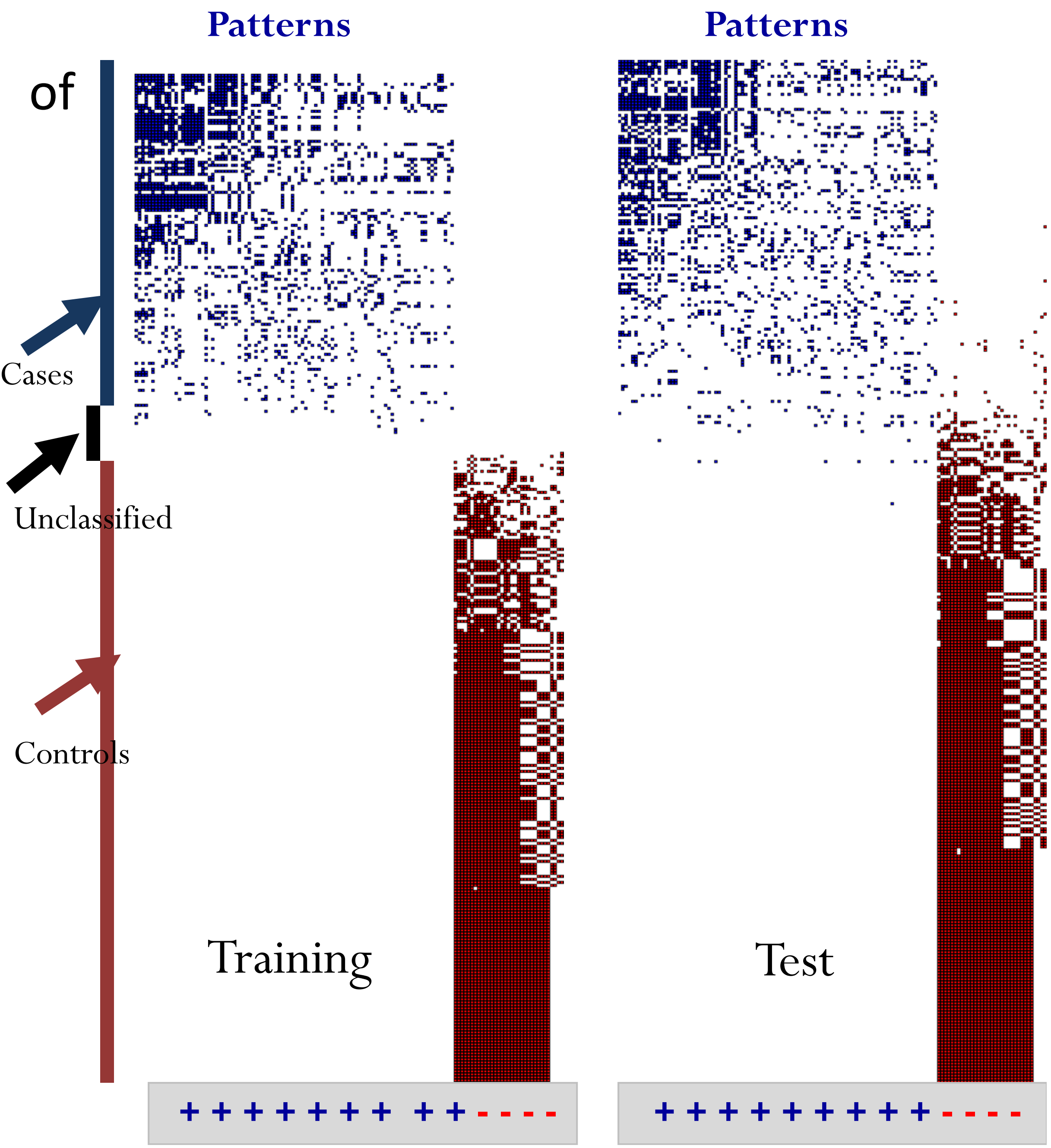
Spanned patterns: conservative, high confidence, diagnosis, gene feature analysis

Prime patterns: aggressive, large sample clusters detection, combinatorial biomarkers

Alexe G, Hammer P: Spanned patterns in Logical Analysis of Data. Discrete Applied Mathematics 2006.
Alexe G, Alexe S, Hammer PL, Vizvari B: Pattern-based feature selection in genomics and proteomics. Annals of Operations Research 2006.
Alexe G, Alexe S, Hammer P: Pattern-based clustering and attribute analysis. Soft Computing 2006.

Modular Representation of Genomic Data: “Pattern” Space

To each genomic data sample *s* and pattern *P:*
assign pattern-sample “activation score” and significance p-value



- Robustness, noise reduction
- Dimensionality reduction
- Libraries of signatures
- Integration of data from multiple sources

*Alexe, Alexe, Hammer, Kogan, *Discr Math* 2007

Logical Analysis of Data: Methodology

- **Discretization/Binarization**: Minimal set of grid cutpoints
- **Support Set Selection:** Collection of reliable attributes
- **Pattern Generation**: Exhaustive enumeration strategy *

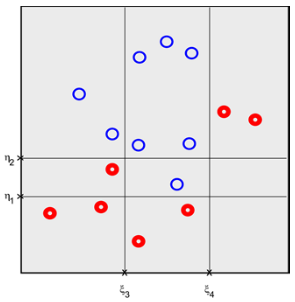
 # operations ≤ 2^n *c*

$$c = \frac{1}{2^n}\prod_{i=1}^{n}(k_i+1)(k_i+2)$$

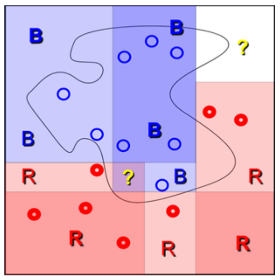
n – number of attributes
k_i – number of cutpoints per attribute i

- **Model Construction:** Optimal **Set Covering of data with collections of class patterns**.
  Classification Models in Pattern Space.
  Validation.

Discretization

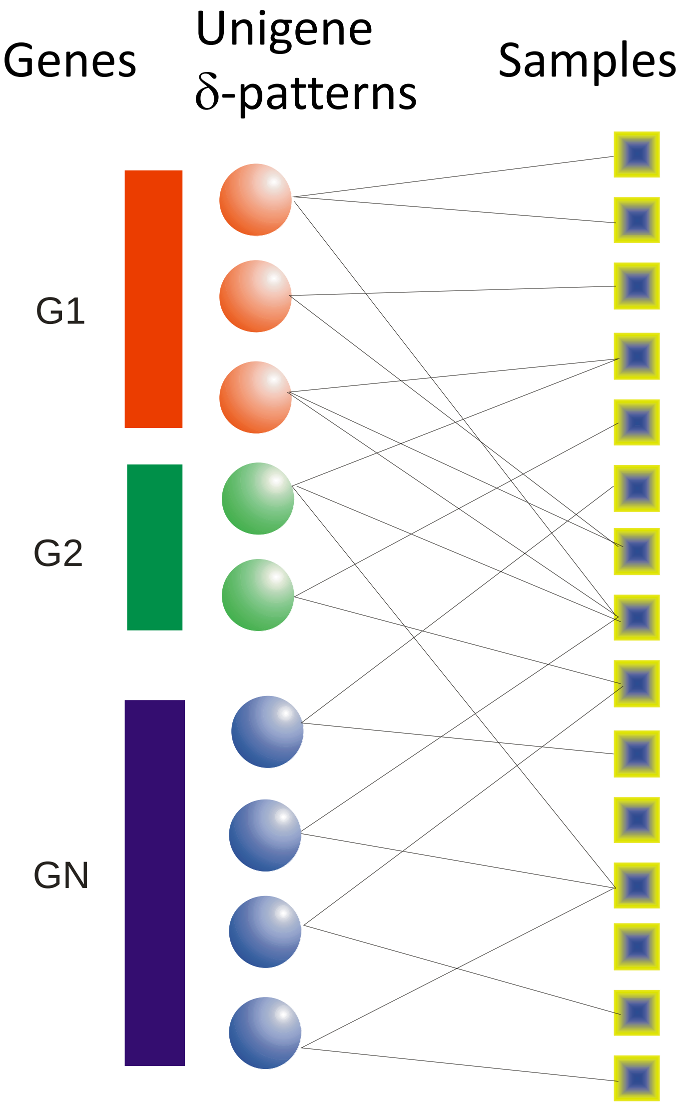


Model



•Alexe & Hammer, *Discr Math* 2006
•Alexe & Hammer, *Discr Appl Math* 2006

Consensus algorithm for δ-pattern extraction



Apply consensus algorithm:

generate all maximal bicliques (*X*,*S*) in the bipartite graph (Unigene δ-pattern, Samples)

(*X* replaced by set of associated genes)

Complexity (serial version):

$$O\big(P\beta(m+m\log\beta)+\beta(m+mP)\big)=$$

$$O\big(Pm\beta\log\beta\big)\leq O(P^2m\beta)$$

(if algorithm is applied from left to right)

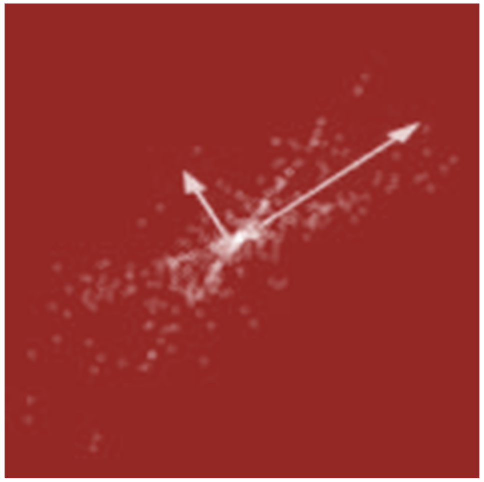
$$O(\min(P,m)Pm\beta)$$

P = # unigene δ-patterns
m = # positive samples
β = # maximal δ-patterns

(3) Unsupervised Metagene Projection Analysis:
Principal Component Analysis vs *Independent* Component Analysis

- Identifies relevant "directions" (meta-genes)
- Reduce the dimensionality to allow identification of hidden relationships in data
- Identify biological mechanisms as "sources" which generate data

Principal Component Analysis (PCA):
Directions of Highest Variation
Linear Associations



Large Scale Biology

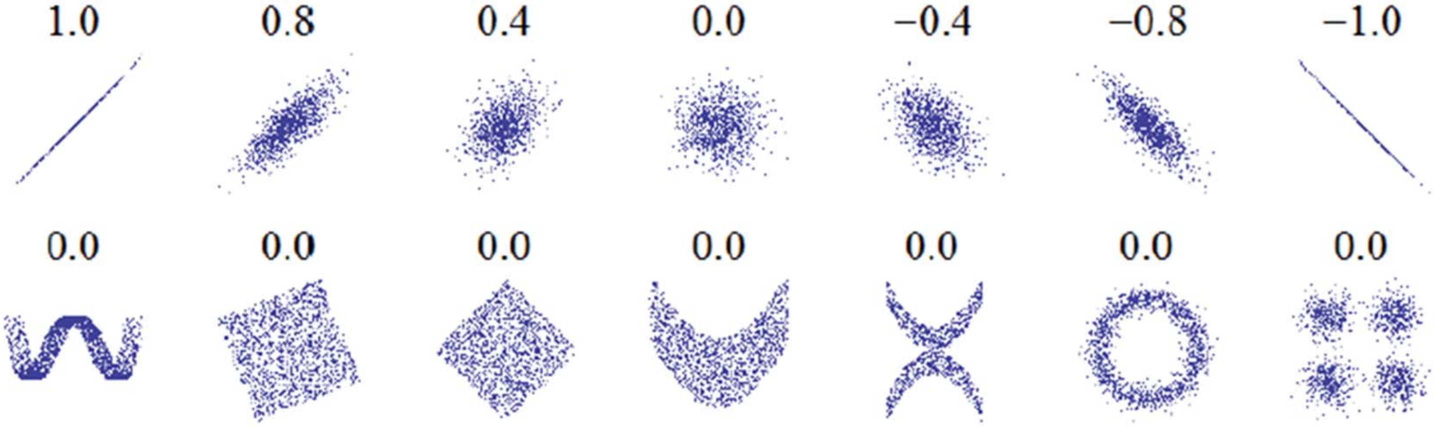
Independent Component Analysis (ICA):
Non-Gaussian Directions
Non-Linear Associations

Micro Scale Biology

Pearson Correlation: does it really capture all patterns of association in transcript data?

Measure of *linear* dependence between r.v.'s X and Y.

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]}{\sigma_x \cdot \sigma_Y}.$$



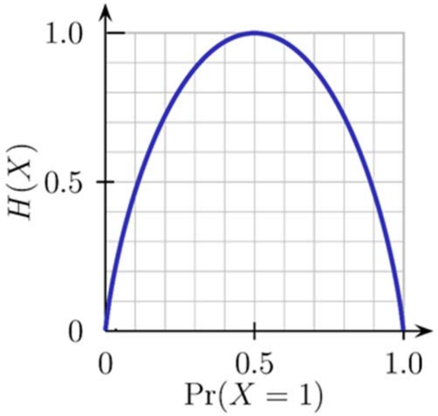
Information Theory Primer

Shannon's Entropy "Quantity of Information to be learned from a distribution X "

Principles:

(1) Monotony: Quantity of information *increases* as the corresponding probability *decreases*.

(2) Additivity: H(*X*,*Y*) = H(*X*) +H(*Y*).



Entropy diagram for a two dimensional distribution

$$H(X) = -\sum_i p_i \log p_i$$

The quantity of information is the *entropy* of the associated probability distribution for the event.

Mutual Information (Shannon) : a general measure of **dependence**

$$I(X;Y) = H(X,Y) - H(X) - H(Y) = H(X) - H(X|Y)$$

Mutual information is the amount of information that knowing either variable provides about the other.

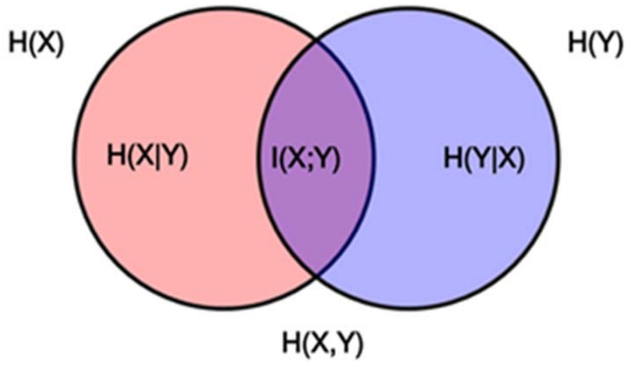
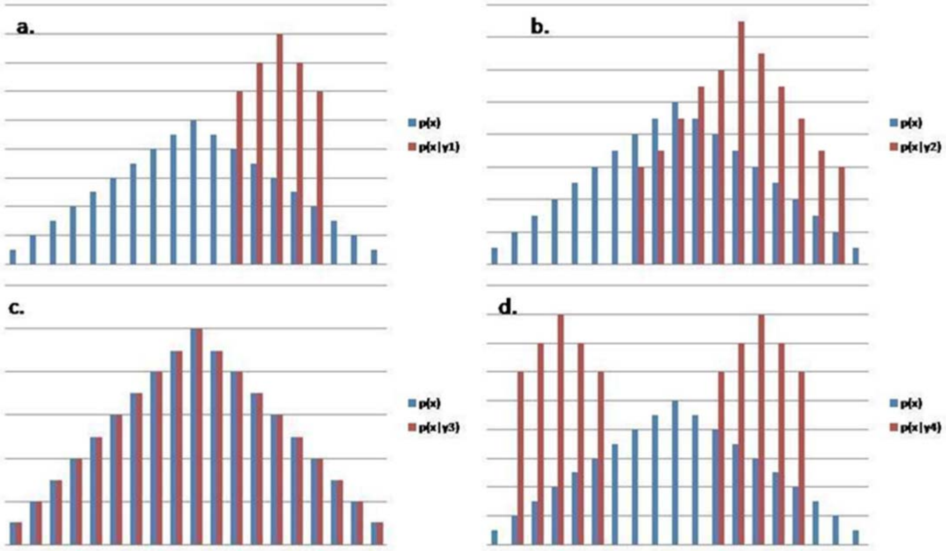


Diagram representation of the mutual information shared by two distributions *X* and *Y*.



This figure demonstrates how *Y* provides information for *X*. The blue is $p(X)$ and the red is $p(X|Y)$ for a specific *Y*.*

---

*Larsen, Richard J. and Marx, Morris L. (2006) An Introduction to Mathematical Statistics and Its Applications.

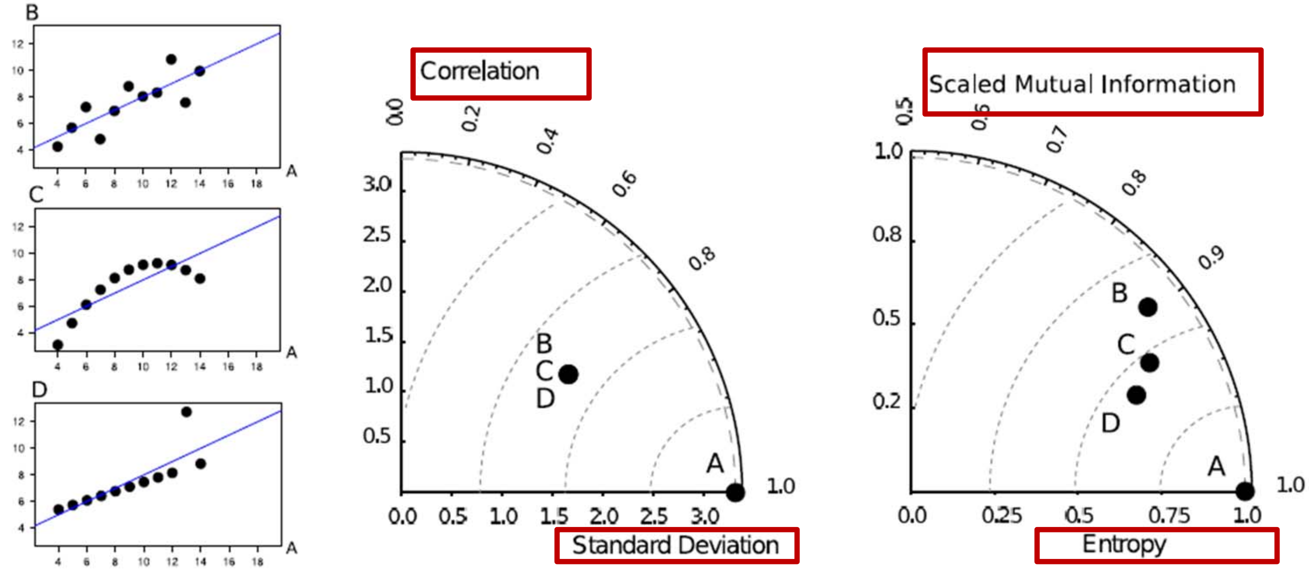
**Kraskov et al. (2003) Hierarchical Clustering Based on Mutual Information.

A classical example: Taylor Diagrams vs Mutual Information Diagram
Pearson correlation vs Mutual Information*

Noise

Non-linear

Outlier



(1) Anscombe's Quartet (2) Taylor Diagram (3) Mutual Information

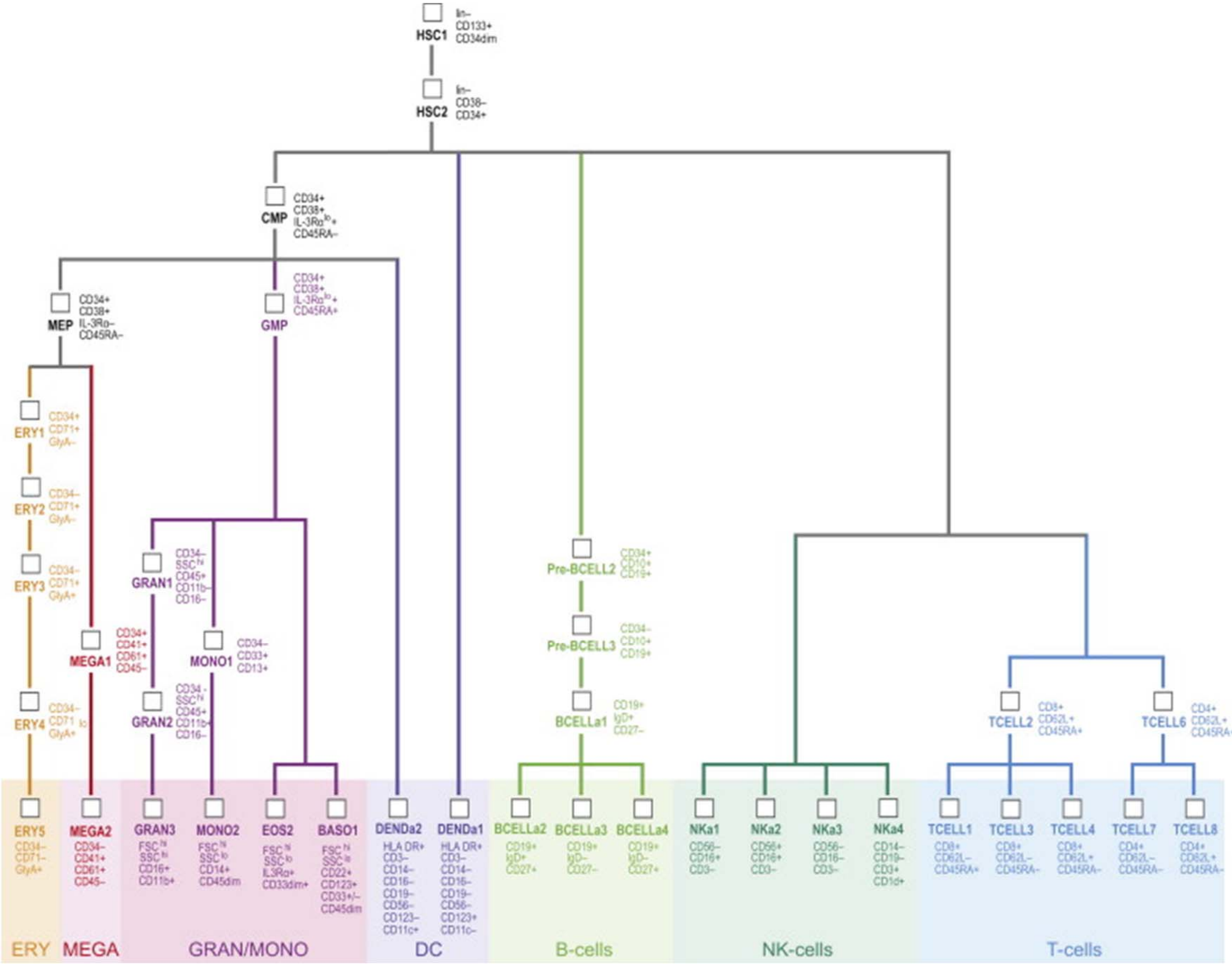
*Correa and Lindstrom 2012

DMAP Portal: A differentiation map of hematopoiesis, from the paper "Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis“ http://www.broadinstitute.org/dmap/home

DMAP:	**Differentiation Map of Hematopoiesis**
Study:	Regulatory circuitry of 38 populations of hematopoietic cells
Organism:	Homo Sapiens
Population:	212 samples (38 populations)
Source:	Human umbilical cord (placenta) and peripheral blood from adult healthy donors for terminal differentiated populations. 4-7 donors per population.
Platform:	HG_U133A 8,887 selected uni-genes (out of 22,944 probesets)
Preprocessing:	Filter QC + RMA + additional ad-hoc normalization per gene (provided as is by DMAP Team)

http://www.broadinstitute.org/dmap/home/789/
Novershtern et al. Cell 2011
Ebert Lab HMS
Regev Lab, Broad, MIT

Hematopoietic Differentiation Map



Novershtern et al. Cell 2011

DMAP ImmunoPhenotypes

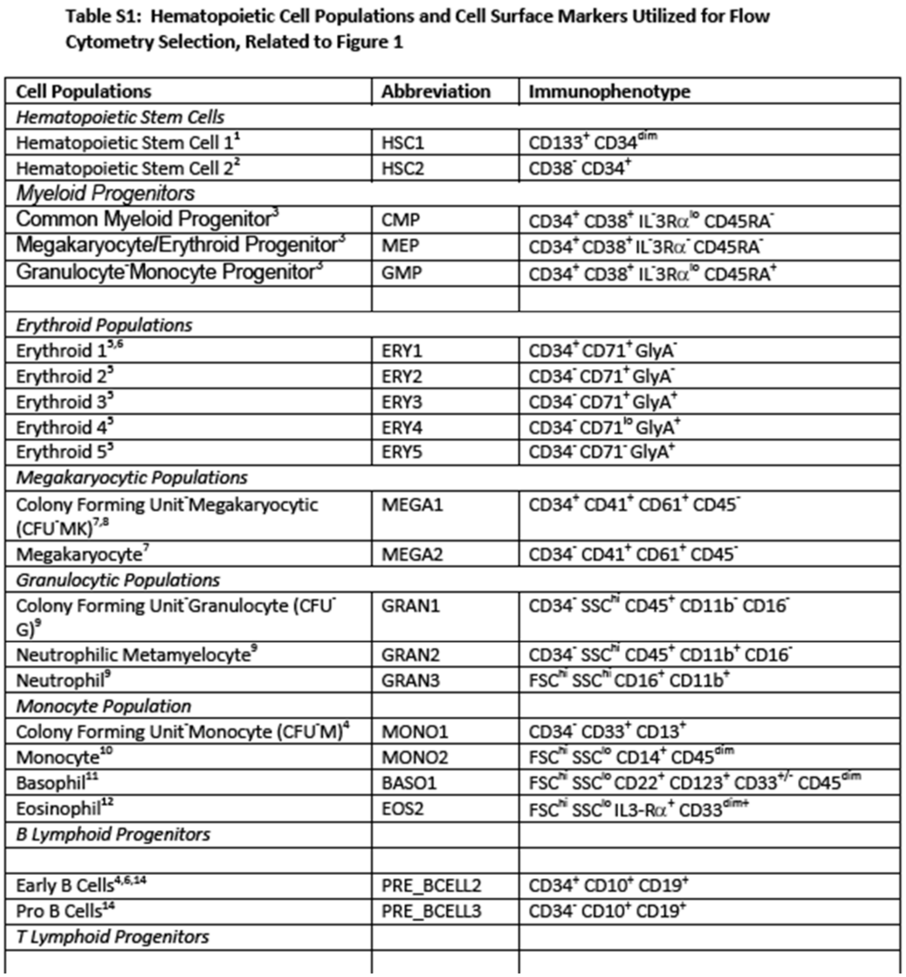
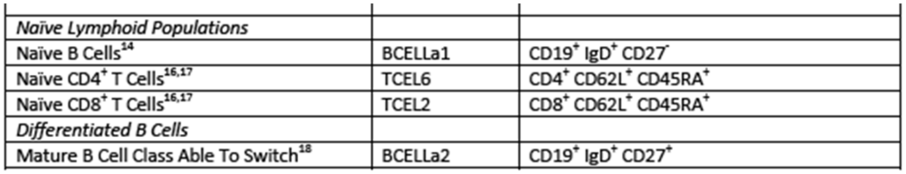
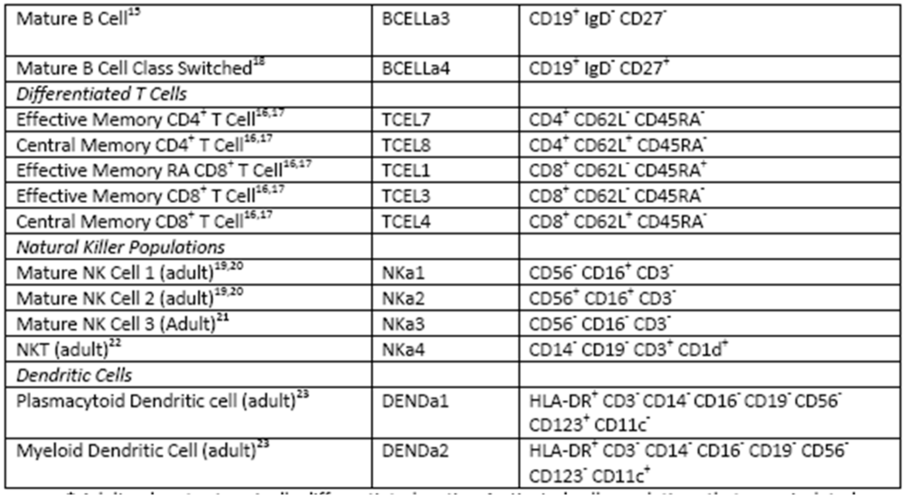
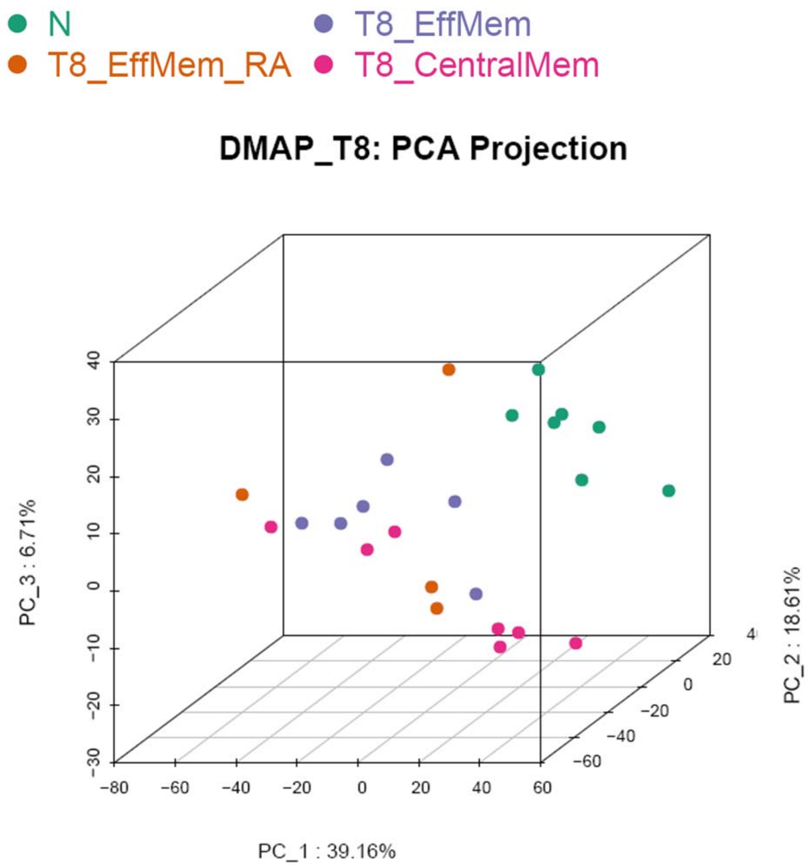




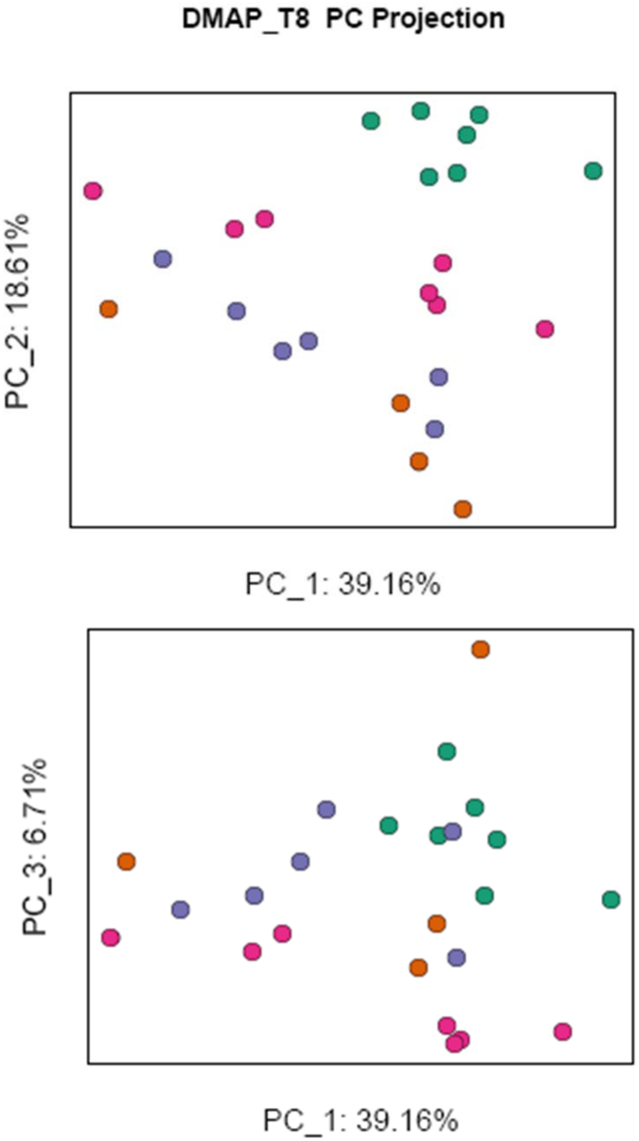
Table S1: Hematopoietic Cell Populations and Cell Surface Markers Utilized for Flow Cytometry Selection, Related to Figure 1

Cell Populations	**Abbreviation**	**Immunophenotype**
Hematopoietic Stem Cells		
Hematopoietic Stem Cell 1[1]	HSC1	CD133^{+} CD34^{dim}
Hematopoietic Stem Cell 2[2]	HSC2	CD38^{-} CD34^{+}
Myeloid Progenitors		
Common Myeloid Progenitor[3]	CMP	CD34^{+} CD38^{+} IL-3Rα^{lo} CD45RA^{-}
Megakaryocyte/Erythroid Progenitor[3]	MEP	CD34^{+} CD38^{+} IL-3Rα^{-} CD45RA^{-}
Granulocyte-Monocyte Progenitor[3]	GMP	CD34^{+} CD38^{+} IL-3Rα^{lo} CD45RA^{+}
Erythroid Populations		
Erythroid 1[5,6]	ERY1	CD34^{+} CD71^{+} GlyA^{-}
Erythroid 2[5]	ERY2	CD34^{-} CD71^{+} GlyA^{-}
Erythroid 3[5]	ERY3	CD34^{-} CD71^{+} GlyA^{+}
Erythroid 4[5]	ERY4	CD34^{-} CD71^{lo} GlyA^{+}
Erythroid 5[5]	ERY5	CD34^{-} CD71^{-} GlyA^{+}
Megakaryocytic Populations		
Colony Forming Unit-Megakaryocytic (CFU-MK)[7,8]	MEGA1	CD34^{+} CD41^{+} CD61^{+} CD45^{-}
Megakaryocyte[7]	MEGA2	CD34^{-} CD41^{+} CD61^{+} CD45^{-}
Granulocytic Populations		
Colony Forming Unit-Granulocyte (CFU-G)[9]	GRAN1	CD34^{-} SSC^{hi} CD45^{+} CD11b^{-} CD16^{-}
Neutrophilic Metamyelocyte[9]	GRAN2	CD34^{-} SSC^{hi} CD45^{+} CD11b^{+} CD16^{-}
Neutrophil[9]	GRAN3	FSC^{hi} SSC^{hi} CD16^{+} CD11b^{+}
Monocyte Population		
Colony Forming Unit-Monocyte (CFU-M)[4]	MONO1	CD34^{-} CD33^{+} CD13^{+}
Monocyte[10]	MONO2	FSC^{hi} SSC^{lo} CD14^{+} CD45^{dim}
Basophil[11]	BASO1	FSC^{hi} SSC^{lo} CD22^{+} CD123^{+} CD33$^{+/-}$ CD45^{dim}
Eosinophil[12]	EOS2	FSC^{hi} SSC^{lo} IL3-Rα^{+} CD33^{dim+}
B Lymphoid Progenitors		
Early B Cells[4,6,14]	PRE_BCELL2	CD34^{+} CD10^{+} CD19^{+}
Pro B Cells[14]	PRE_BCELL3	CD34^{-} CD10^{+} CD19^{+}
T Lymphoid Progenitors		
Naïve Lymphoid Populations		
Naïve B Cells[14]	BCELLa1	CD19^{+} IgD^{+} CD27^{-}
Naïve CD4^{+} T Cells[16,17]	TCEL6	CD4^{+} CD62L^{+} CD45RA^{+}
Naïve CD8^{+} T Cells[16,17]	TCEL2	CD8^{+} CD62L^{+} CD45RA^{+}
Differentiated B Cells		
Mature B Cell Class Able To Switch[18]	BCELLa2	CD19^{+} IgD^{+} CD27^{+}
Mature B Cell[15]	BCELLa3	CD19^{+} IgD^{-} CD27^{-}
Mature B Cell Class Switched[18]	BCELLa4	CD19^{+} IgD^{-} CD27^{+}
Differentiated T Cells		
Effective Memory CD4^{+} T Cell[16,17]	TCEL7	CD4^{+} CD62L^{-} CD45RA^{-}
Central Memory CD4^{+} T Cell[16,17]	TCEL8	CD4^{+} CD62L^{+} CD45RA^{-}
Effective Memory RA CD8^{+} T Cell[16,17]	TCEL1	CD8^{+} CD62L^{-} CD45RA^{+}
Effective Memory CD8^{+} T Cell[16,17]	TCEL3	CD8^{+} CD62L^{-} CD45RA^{-}
Central Memory CD8^{+} T Cell[16,17]	TCEL4	CD8^{+} CD62L^{+} CD45RA^{-}
Natural Killer Populations		
Mature NK Cell 1 (adult)[19,20]	NKa1	CD56^{-} CD16^{+} CD3^{-}
Mature NK Cell 2 (adult)[19,20]	NKa2	CD56^{+} CD16^{+} CD3^{-}
Mature NK Cell 3 (Adult)[21]	NKa3	CD56^{-} CD16^{-} CD3^{-}
NKT (adult)[22]	NKa4	CD14^{-} CD19^{-} CD3^{+} CD1d^{+}
Dendritic Cells		
Plasmacytoid Dendritic cell (adult)[23]	DENDa1	HLA-DR^{+} CD3^{-} CD14^{-} CD16^{-} CD19^{-} CD56^{-} CD123^{+} CD11c^{-}
Myeloid Dendritic Cell (adult)[23]	DENDa2	HLA-DR^{+} CD3^{-} CD14^{-} CD16^{-} CD19^{-} CD56^{-} CD123^{-} CD11c^{+}

DMAP: PCA Projection on top three principal components shows no clear separation of lineages

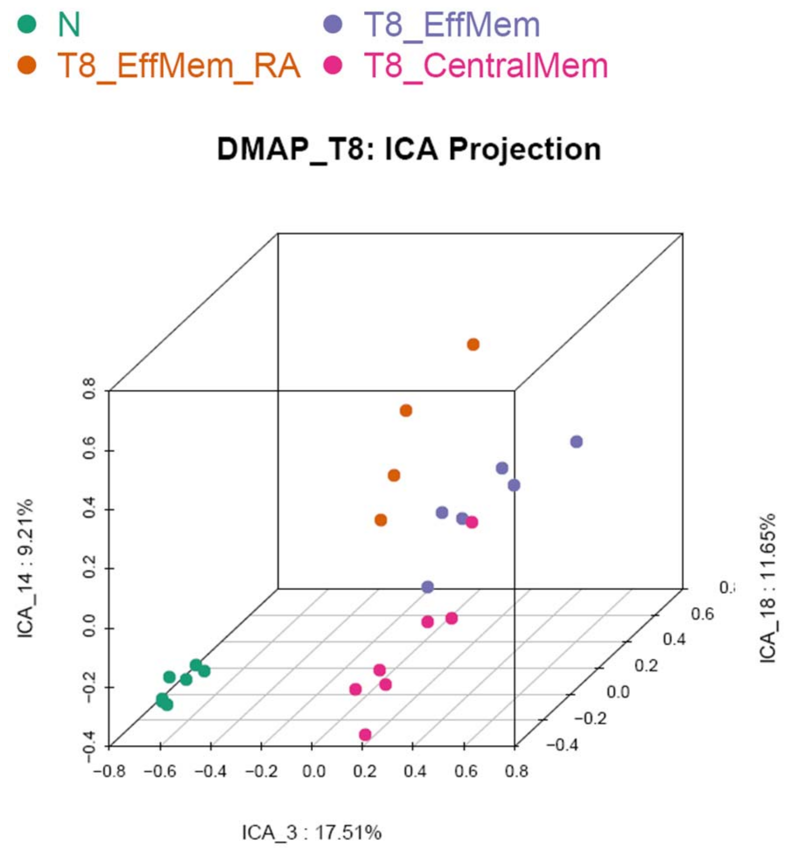


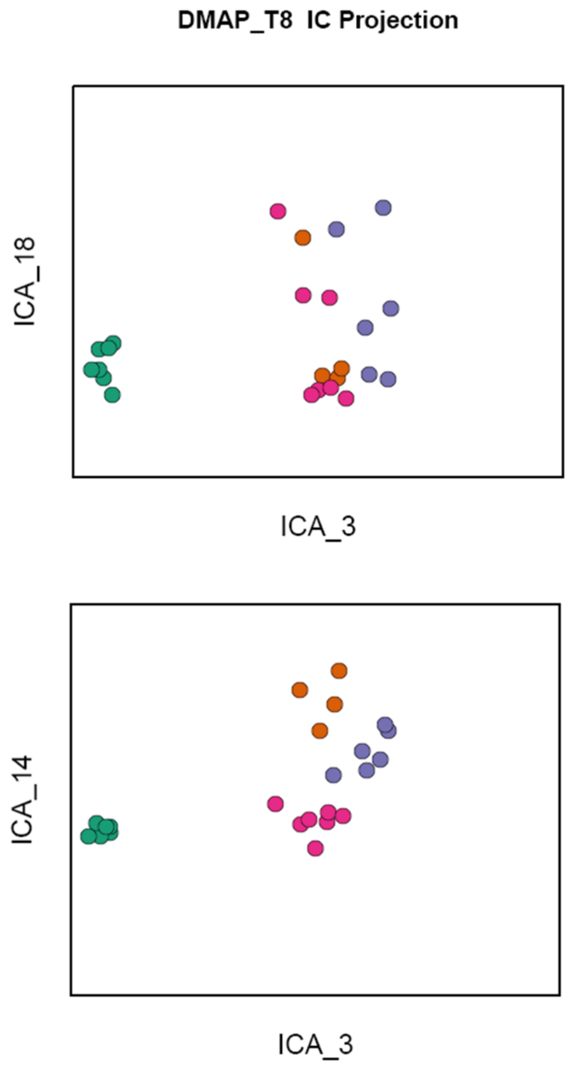
DMAP_T8: PCA Projection



DMAP_T8 PC Projection

DMAP ICA Projection on top variable independent components: ICA 3 and ICA14 as “biological coordinates” are able to distinguish CD8+ T cell phenotypes, while PCA does not

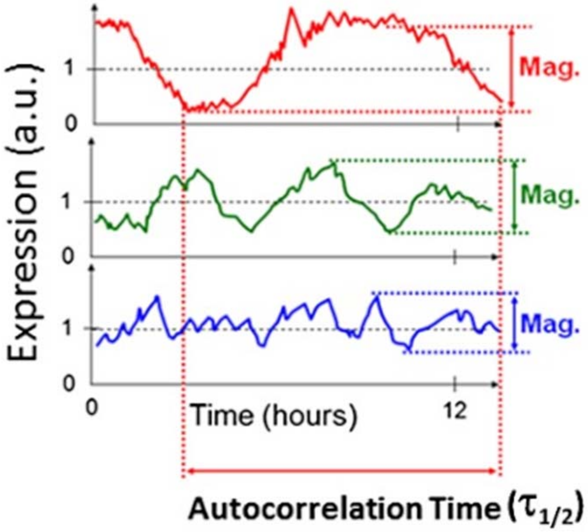




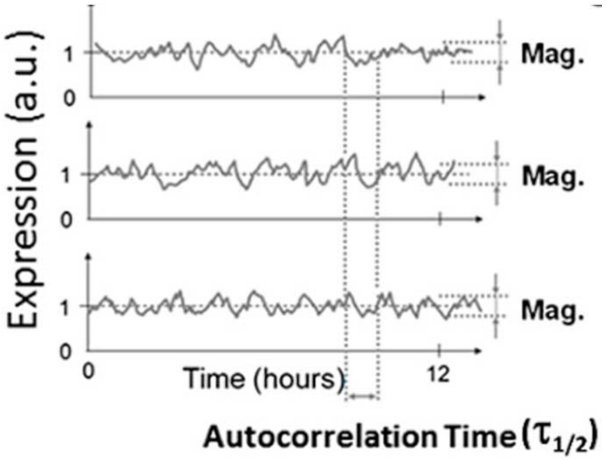
The Fundamental Challenge in Single Cell Transcript-omics

- Fundamental challenge of **pulse formation of eukaryotic transcripts**: Cells from the same biological phenotypic population were found to display high stochastic variation at the transcript level.
- Dar et al, PNAS 109(2012) shown that “ *pulsatile bursting –as opposed to constitutive bursting- is the predominant mode of gene expression and the frequency and the dynamics of the bursts vary equally across the human genome, independent of promoter sequence*.”

Episodic Pulsatile Bursty Gene Expression:
large, slow fluctuations

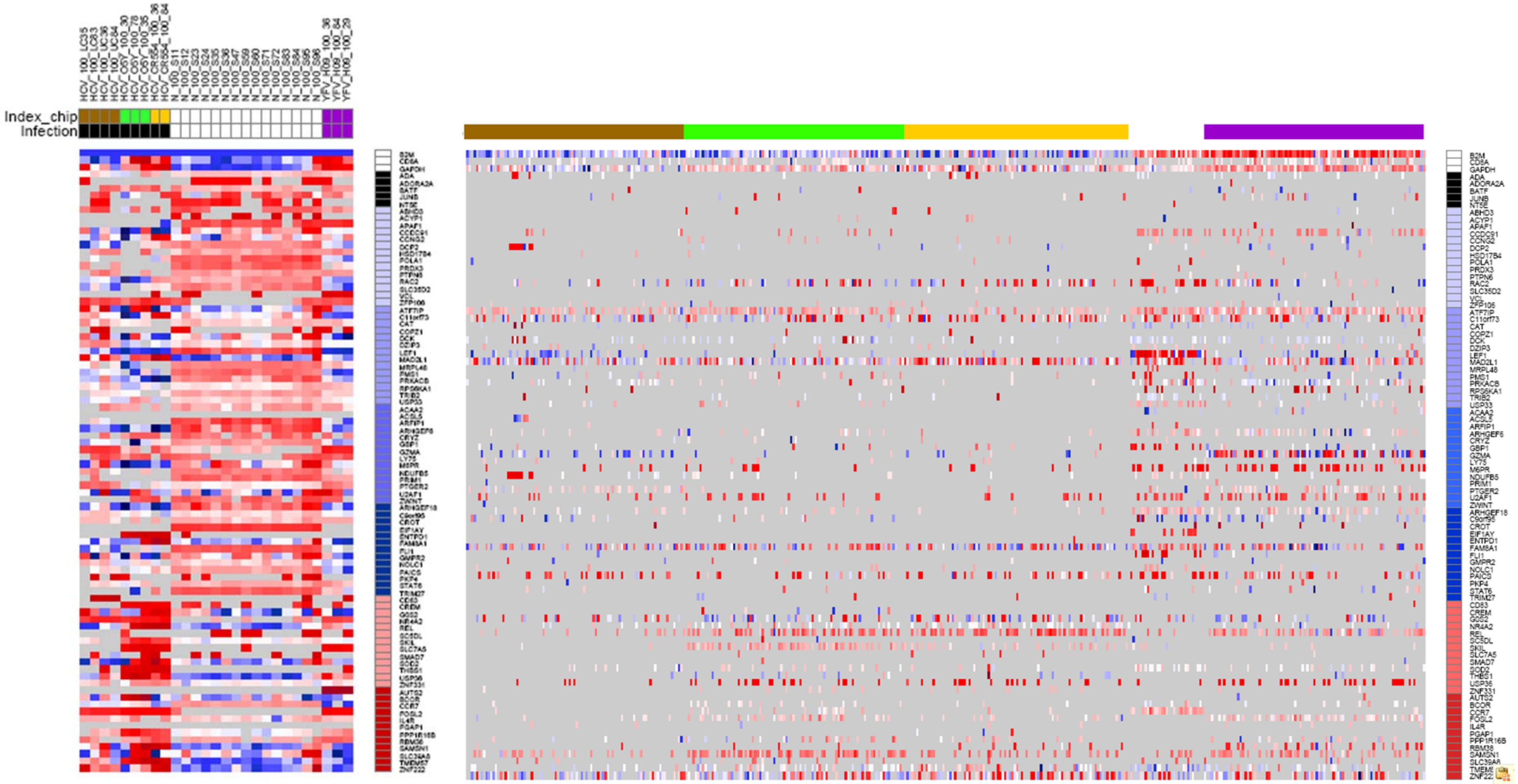


Constitutive Gene Expression Poisson-expression model:
small, fast fluctuations



Dar *et al* PNAS 2012

Bird’s Eye vs High Resolution Layered View of Expression Data - What Information are these data provide?



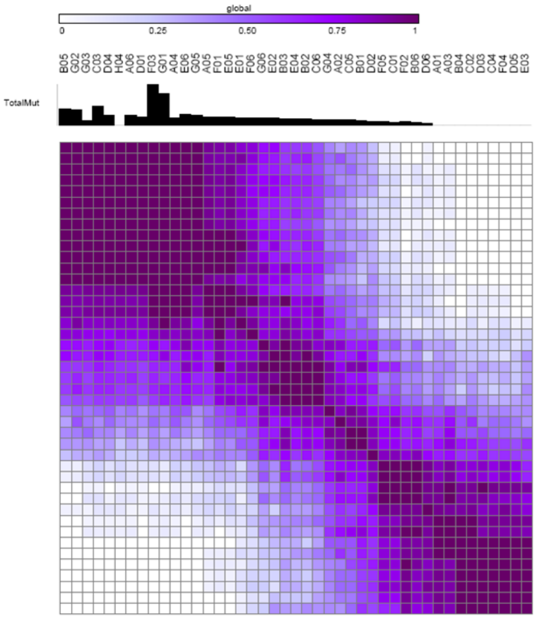
Fluidigm 100 cell expression data

Fluidigm single cell expression data

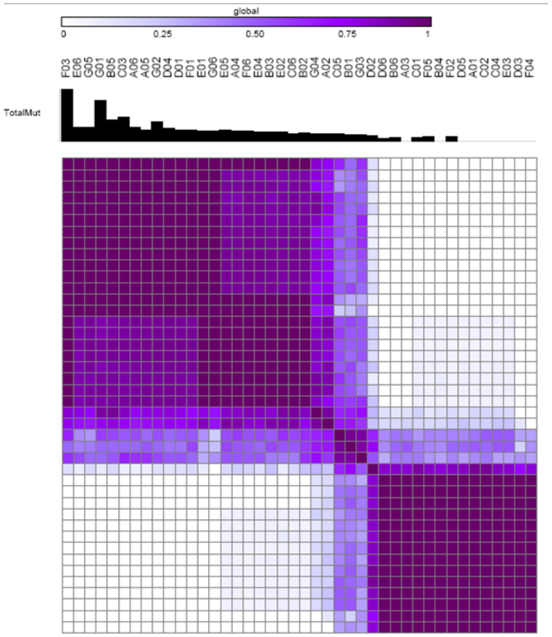
Multi-cell and Single cell Fluidigm data (courtesy of Haining Lab)

Clustering samples:
Hierarchical Clustering based on linear correlation metric is not the appropriate approach
Clusters are evaluated based on biological association with a Mutation Phenotype

Linear Correlation



Mutual Information



Stegmaier Lab
Harvard Medical School

Puissant *et al*
Cancer Discovery 2013

First study shows that
MYCN can be targeted in Neuroblastoma

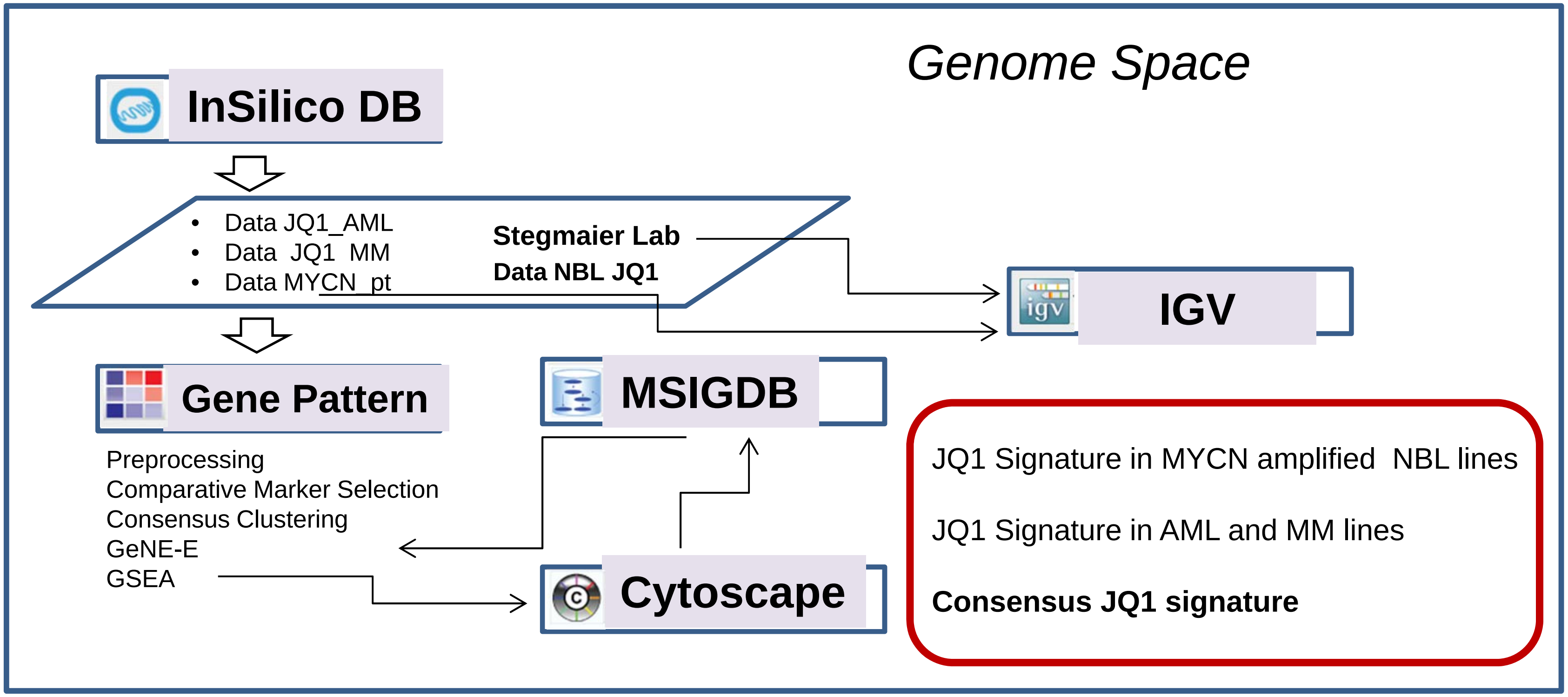
RESEARCH ARTICLE

Targeting MYCN in Neuroblastoma by BET Bromodomain Inhibition

Alexandre Puissant[1,3], Stacey M. Frumm[1,3], Gabriela Alexe[1,3,5,6], Christopher F. Bassil[1,3], Jun Qi[2], Yvan H. Chanthery[8], Erin A. Nekritz[8], Rhamy Zeid[2], William Clay Gustafson[8], Patricia Greninger[7], Matthew J. Garnett[10], Ultan McDermott[10], Cyril H. Benes[7], Andrew L. Kung[1,3], William A. Weiss[8,9], James E. Bradner[2,4], and Kimberly Stegmaier[1,3,6]

Genome Space Pipeline

http://www.genomespace.org/



Genome Space

InSilico DB

- Data JQ1_AML
- Data JQ1 MM
- Data MYCN_pt

Stegmaier Lab

Data NBL JQ1

IGV

Gene Pattern

Preprocessing
Comparative Marker Selection
Consensus Clustering
GeNE-E
GSEA

MSIGDB

Cytoscape

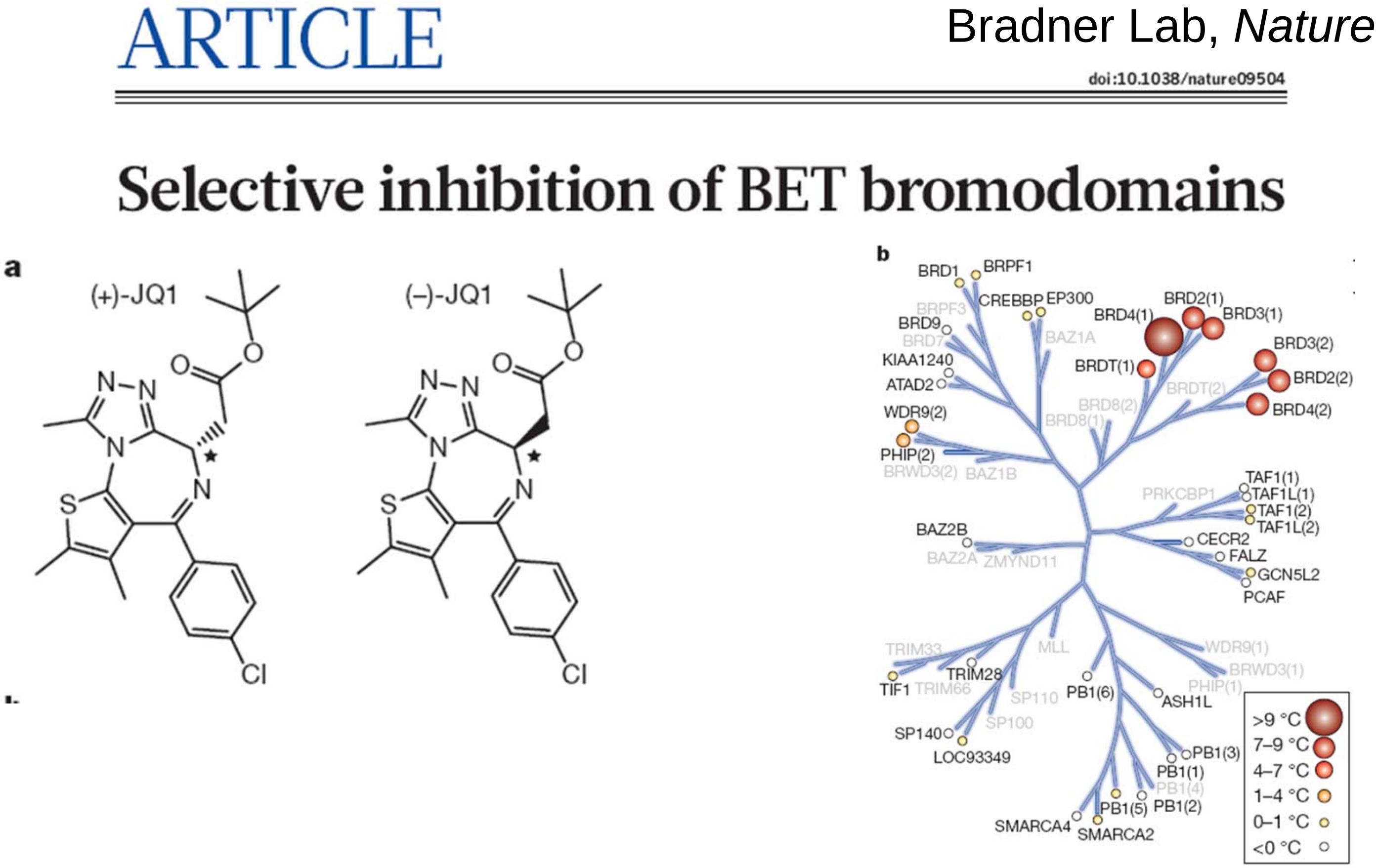
JQ1 Signature in MYCN amplified NBL lines

JQ1 Signature in AML and MM lines

Consensus JQ1 signature

Puissant et al., Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Discov.* 2013 Mar;3(3):308-23.

Bradner Lab, *Nature* 2010



ARTICLE

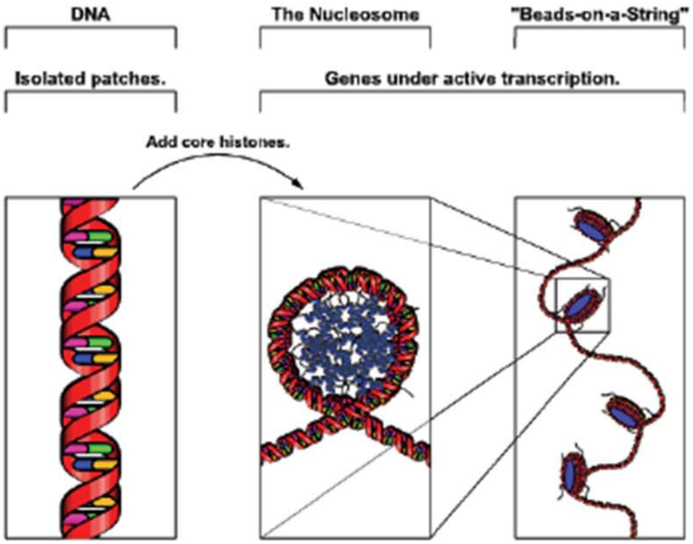
doi:10.1038/nature09504

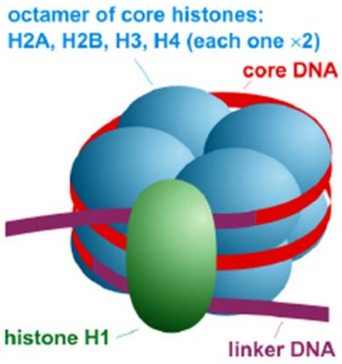
Selective inhibition of BET bromodomains

- JQ1 is a chemical compound under development as a **non-hormonal male contraceptive drug.**

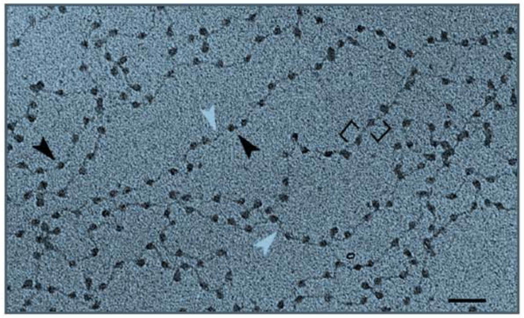
- JQ1 binds competitively to acetyl-lysine recognition motifs or bromodomains and **inhibits BET family genes**. Highest selectivity for BRD4, BRD3, BRD2 and BRD4T (testis).

DNA packaged around core histones (H2A, H2B, H3 and H4) to form nucleosomes, beads on a string.





Nucleosomes plus linker DNA: beads on a string model



Histone tails are prone to modifications through acetylation or methylation

⇩

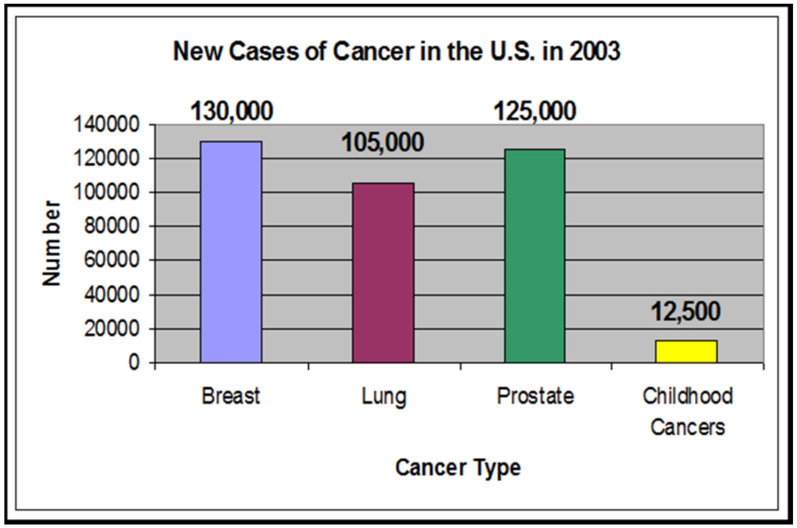
Chromatin alterations around histone spools

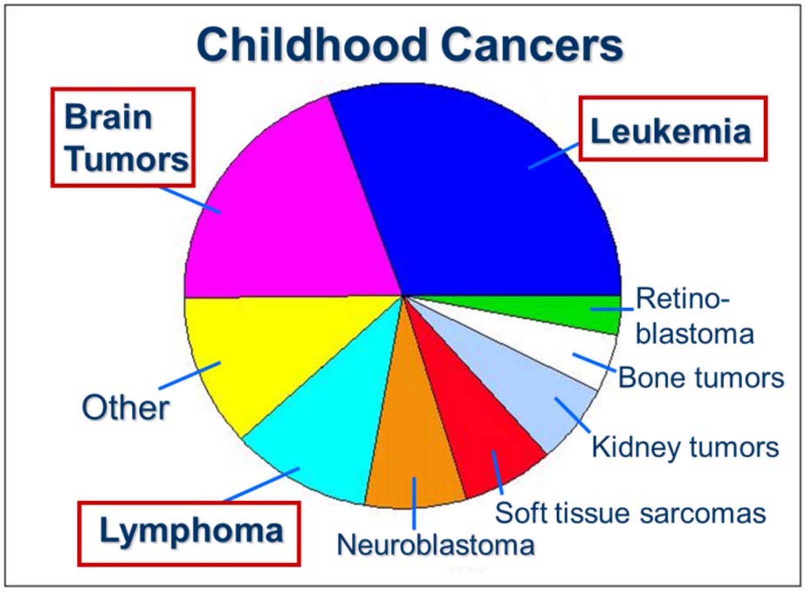
⇩

In cancer: Uncontrolled activation of oncogenes, repression of tumor suppression genes, repression of immune response

B Chang, L Stork: Childhood cancer

Childhood Cancer is a rare disease…

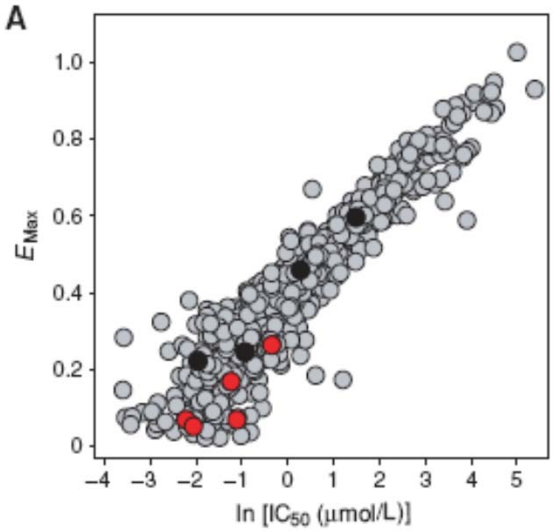




US

- **1 in 330** young adults develops pediatric cancer before age 20
- **1 in 750** survives
- **650 – 700 NBL** cases per year

MYCN amplified cell line tumors are sensitive to JQ1



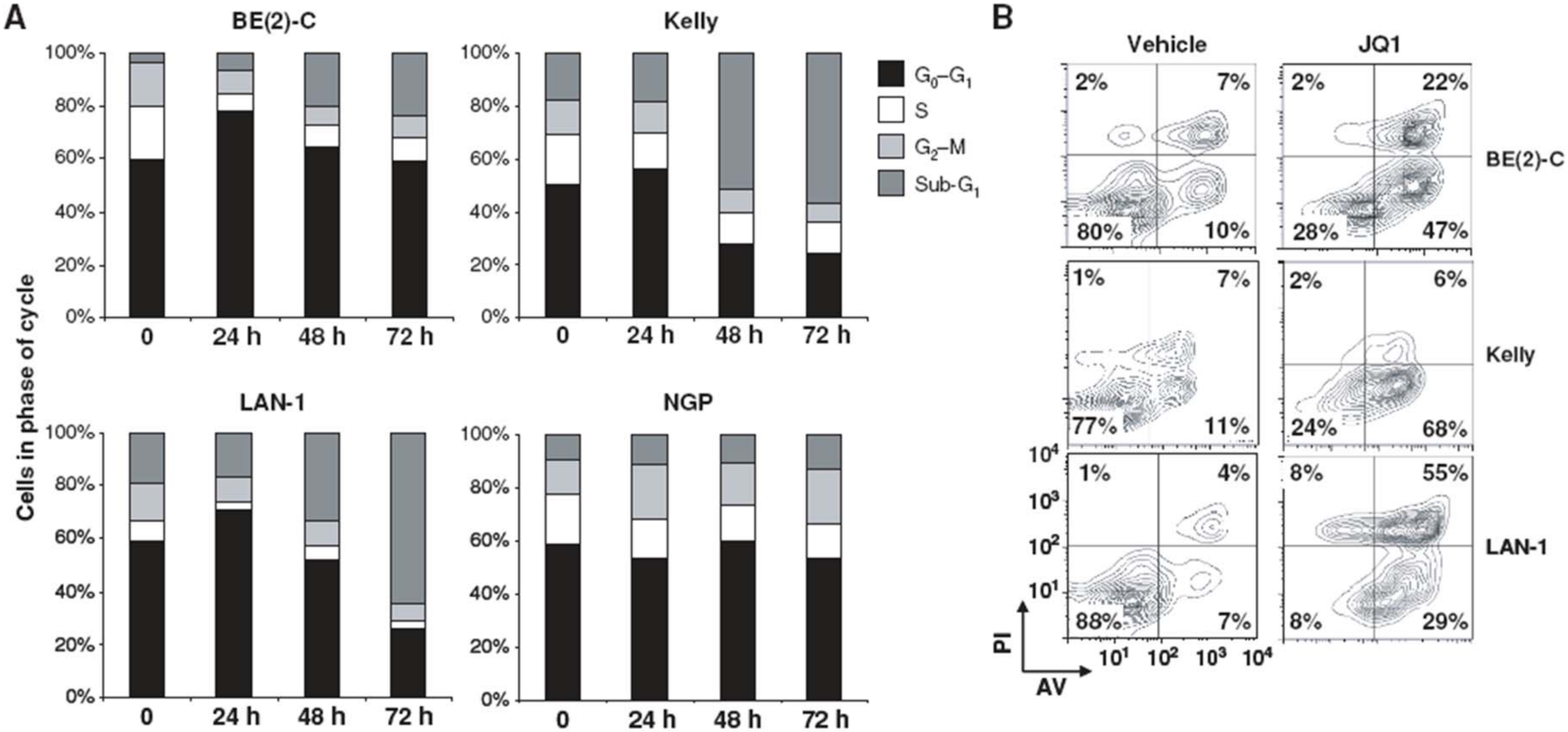
- Antiproliferative activity of JQ1 was profiled in more than 650 cancer cell lines.

Red dots: NBL with MYCN amplification
Black dots: NBL with WT MYCN

- Drug response is presented as ln(IC 50), plotted against the maximum effect corresponding to the minimum measured viability (E max).

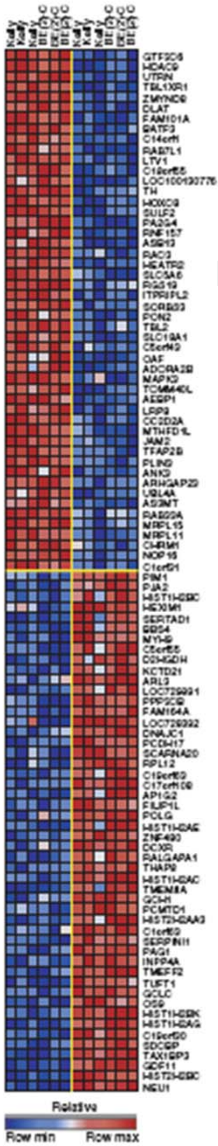
JQ1 treatement induces cell cycle arrest in NBL cell lines with MYCN amplification: BE(2)-C and Kelly

Puissant *et al* (2013)



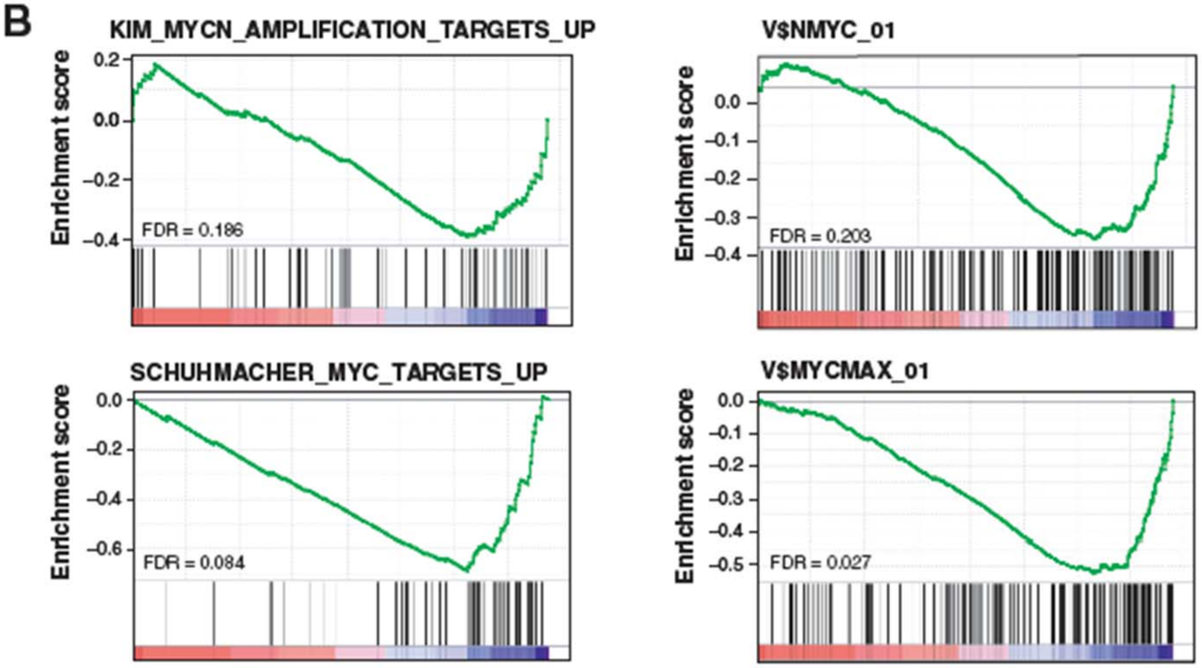
Inhibition of MYCN- and c-MYC–dependent transcription by JQ1 treatment of neuroblastoma cells

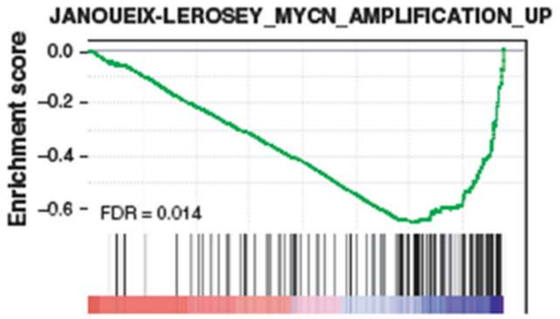
Control **JQ1 (24h)**

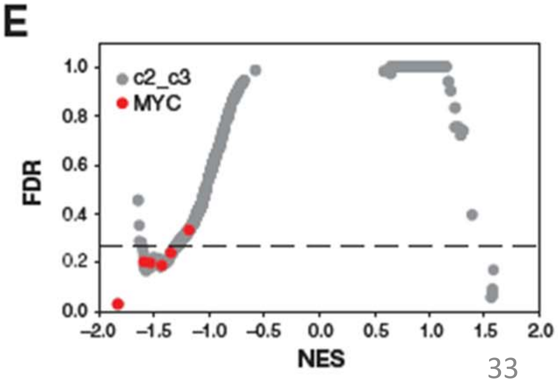


MTHFD1L

MYCN

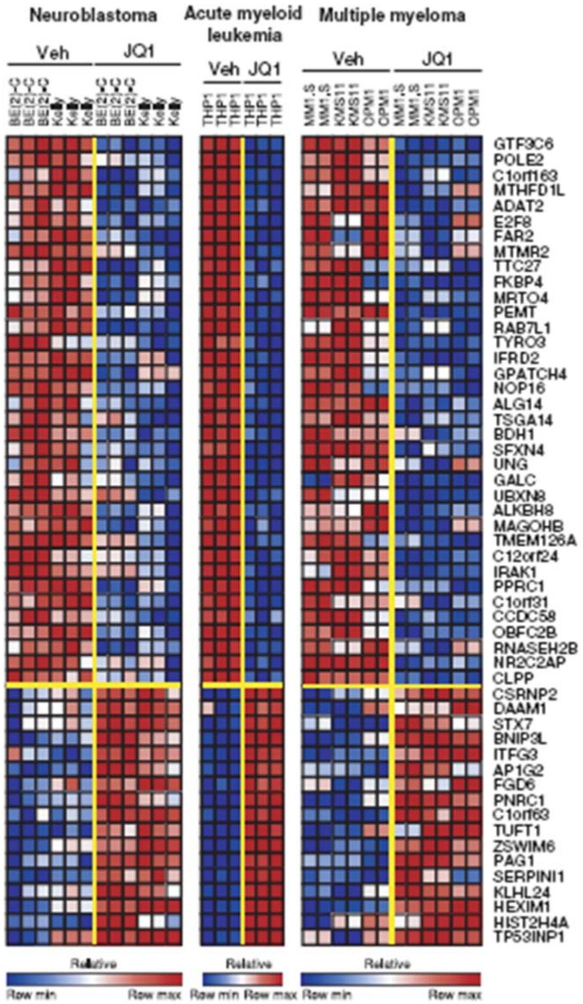




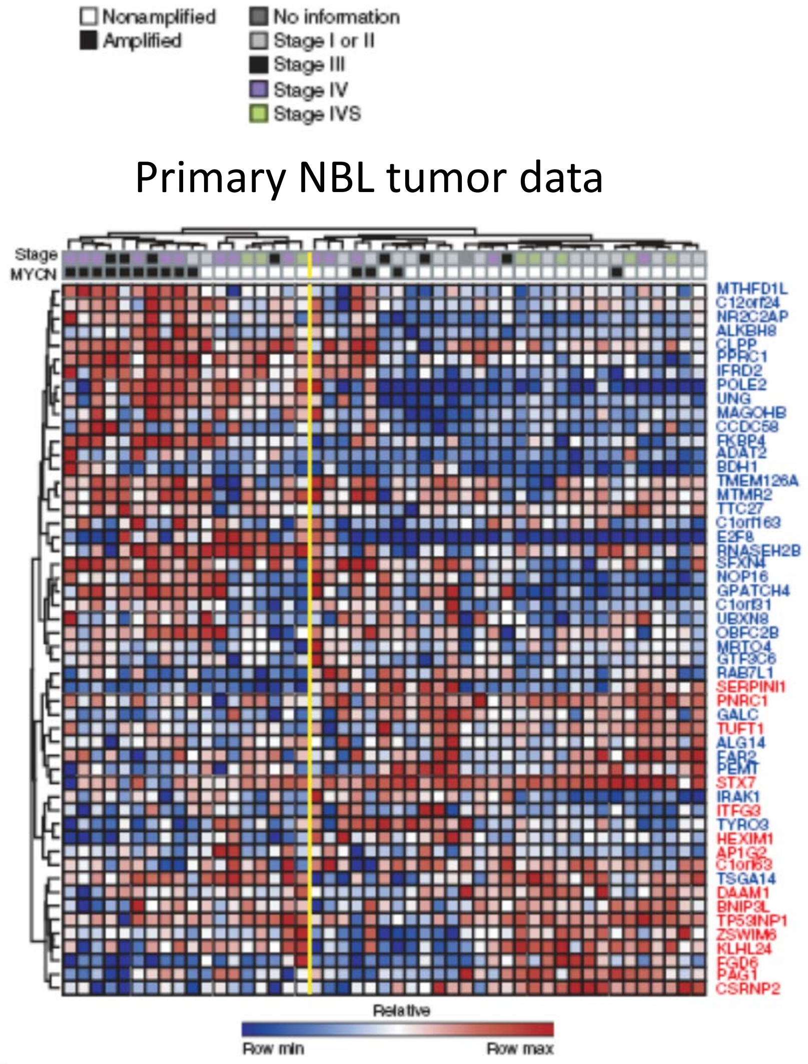


Robust transcriptional changes associated with JQ1 treatment across Neuroblastoma, AML and MM cell lines and primary Neuroblastoma tumors

Cell line data

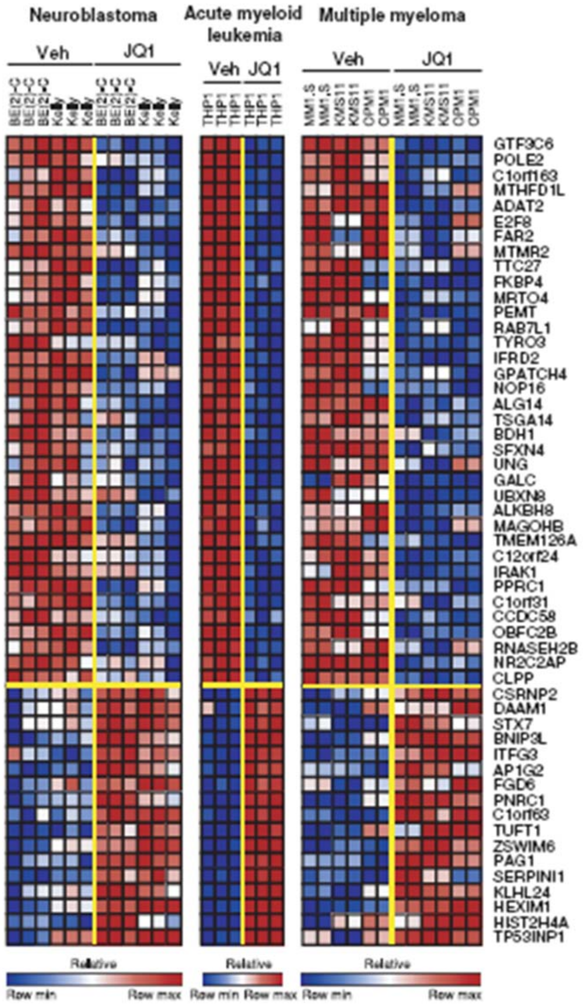


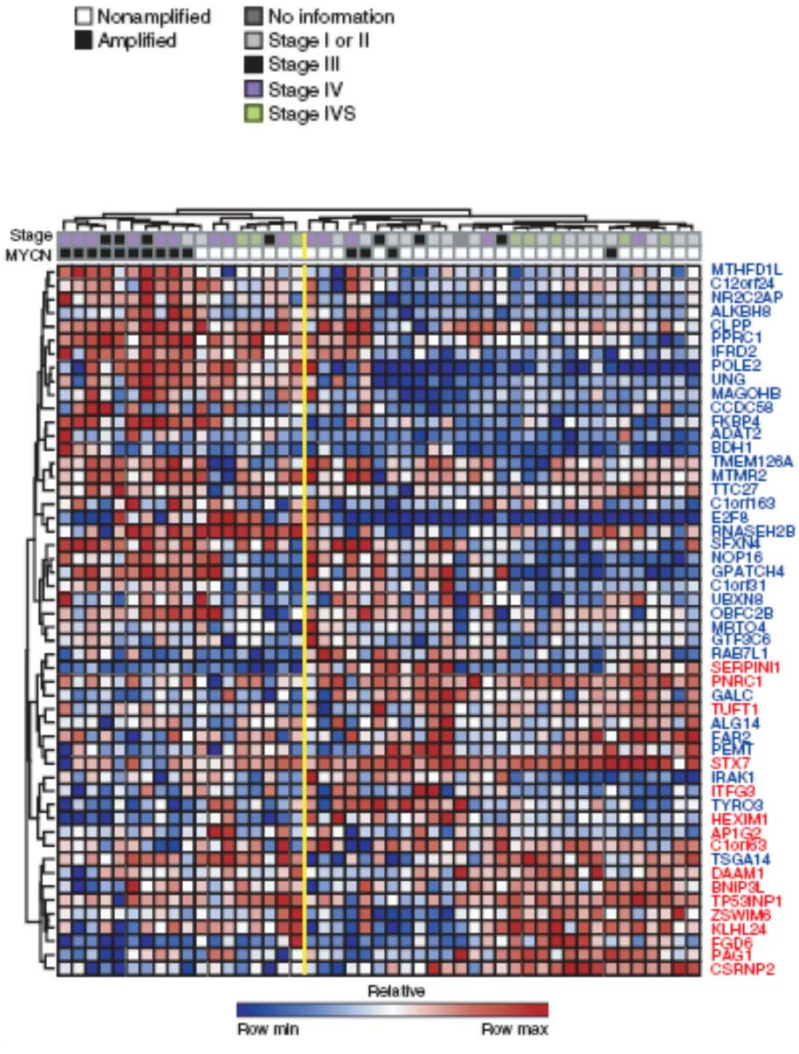
Primary NBL tumor data



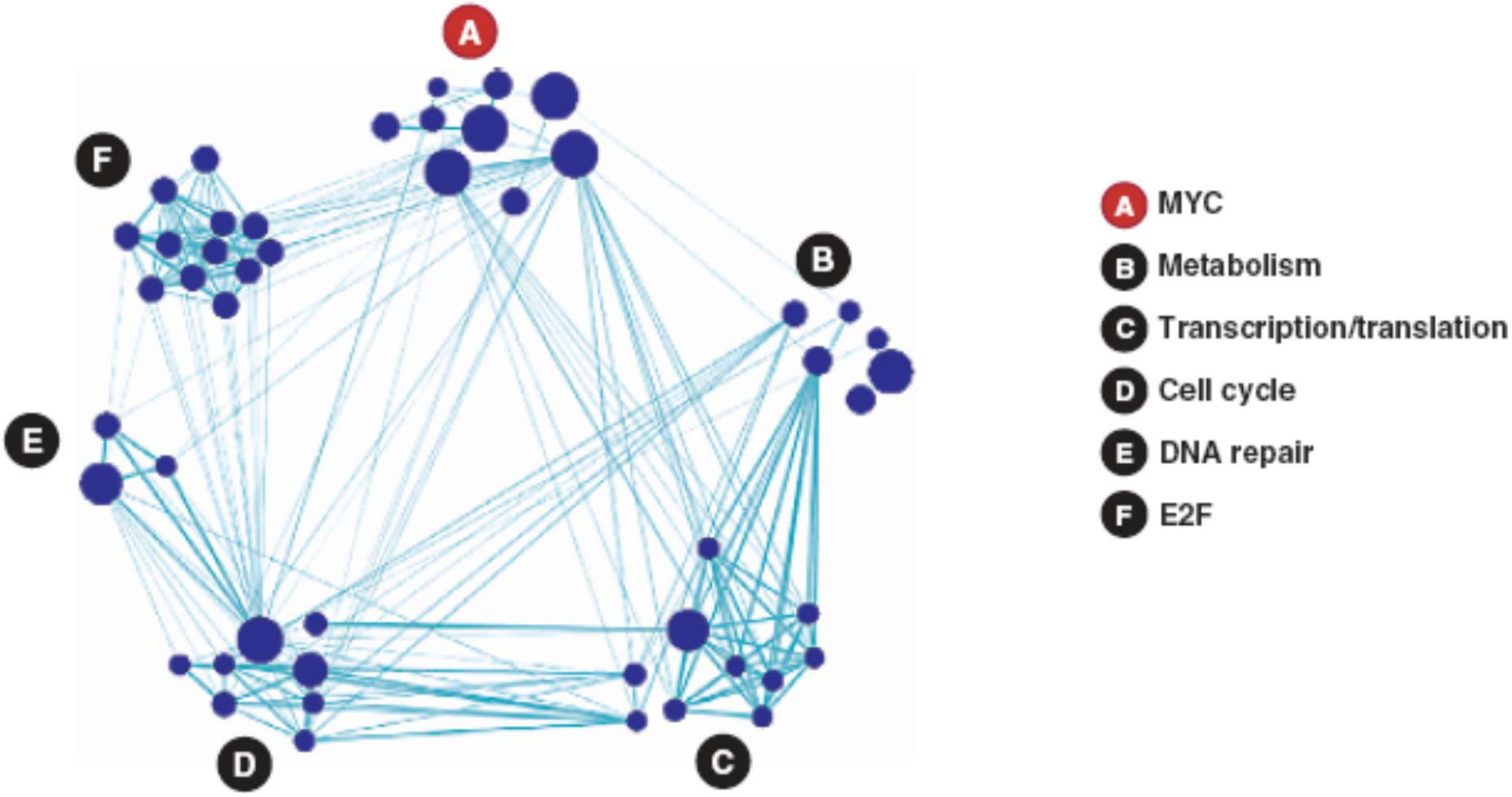
Consensus JQ1 Signature

Transcriptional changes associated with BET bromodomain inhibition by JQ1

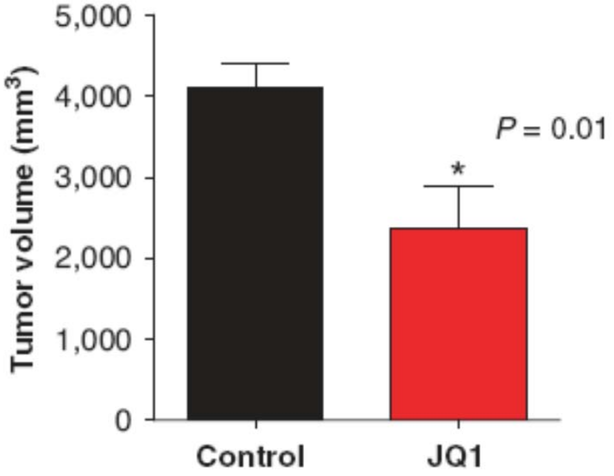


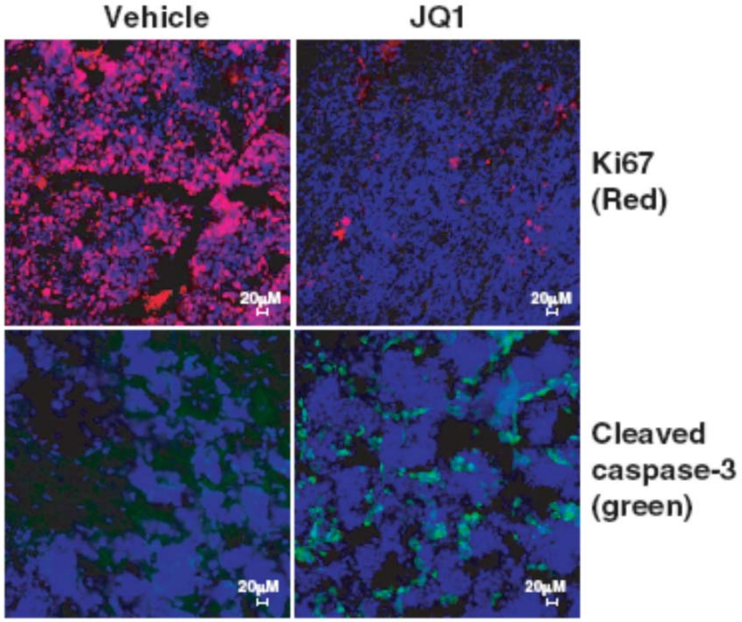


Enrichment Map for JQ1 DN Consensus Signature across Neuroblastoma, AML and MM lines

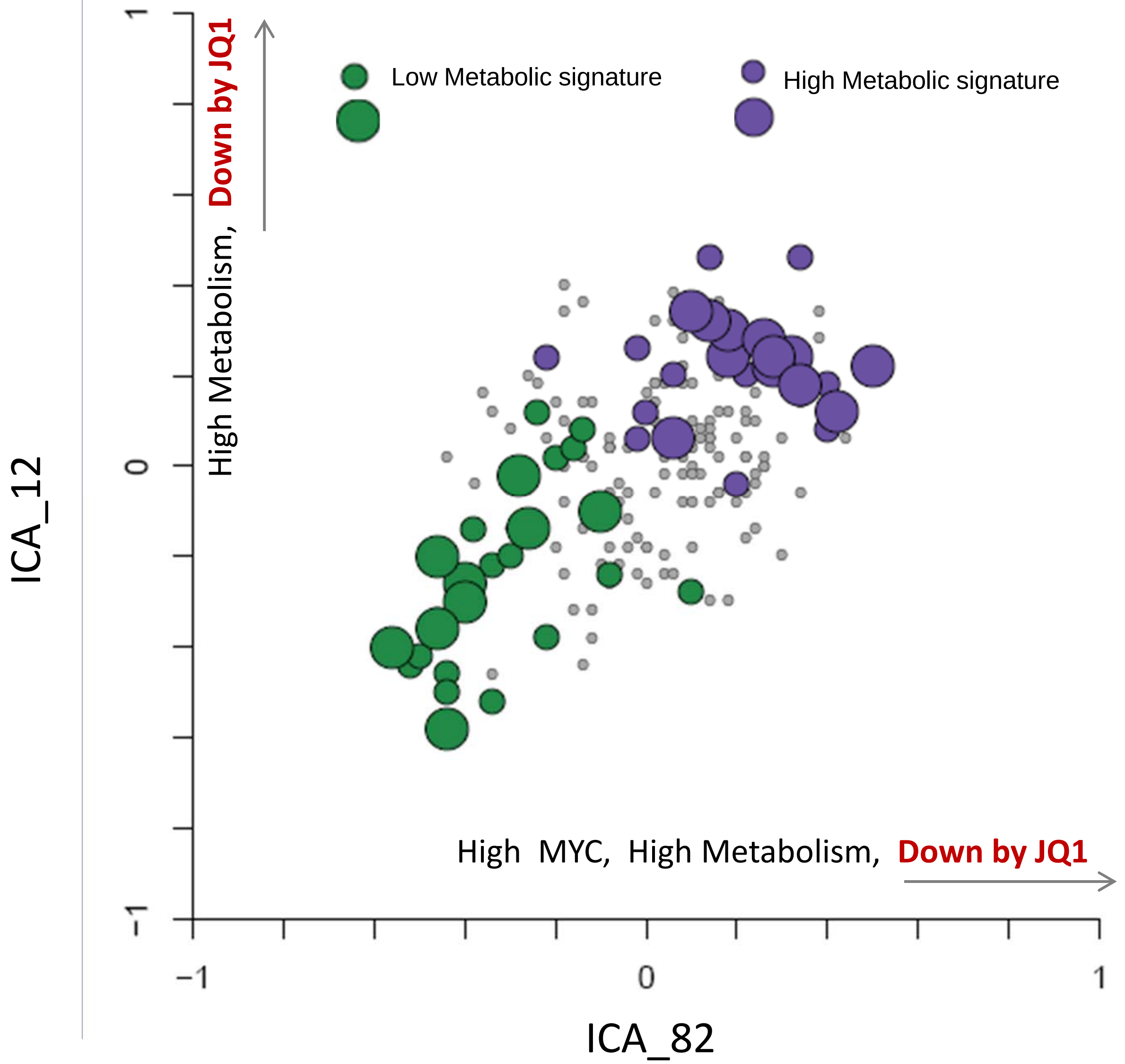


in vivo validation of JQ1 effect in in mice



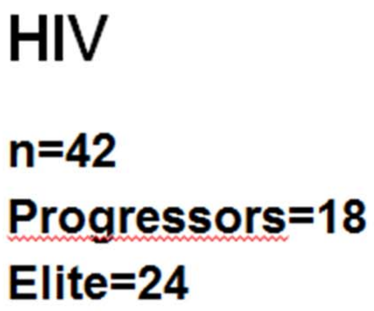


TCGA AML: ICA projection supporting the association JQ1 response and metabolic signature in AML

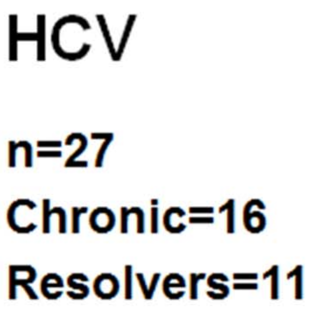


Haining Lab (DFCI):

First tetramer sorted CD8+ T cell data in response to **HIV** and **HCV**


HIV

n=42
Progressors=18
Elite=24

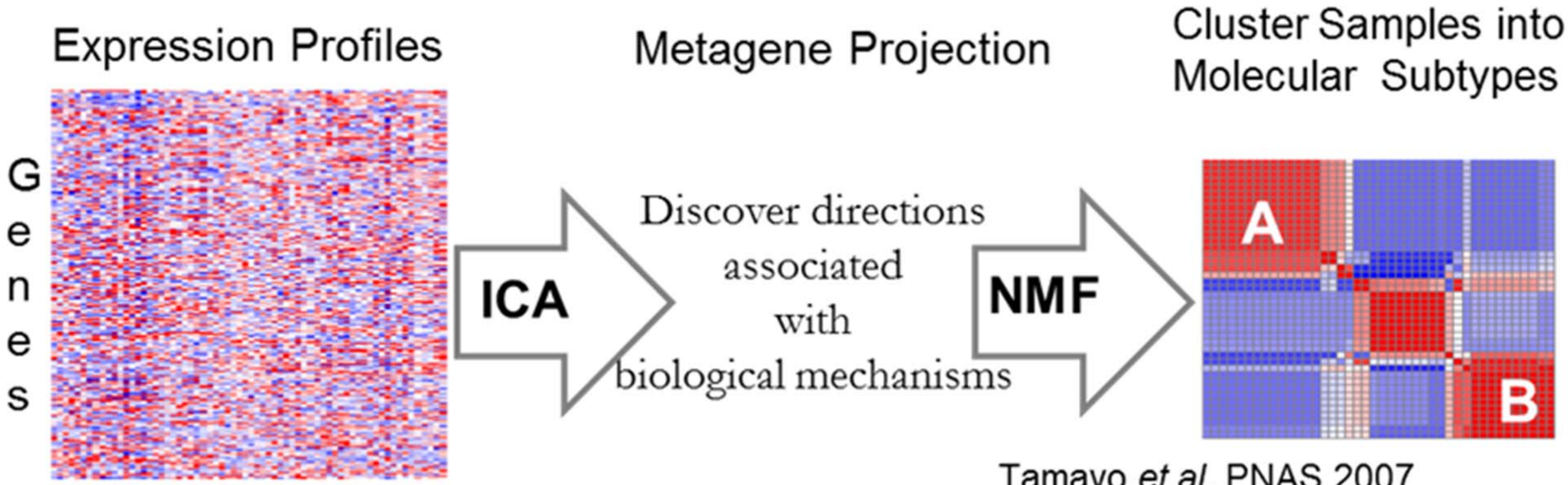

HCV

n=27
Chronic=16
Resolvers=11

Haining Lab, DFCI
Nature Med 2010

Lauer Lab, Mass Gen Hospital

Chart Flow of Metagene Projection Strategy



Expression Profiles

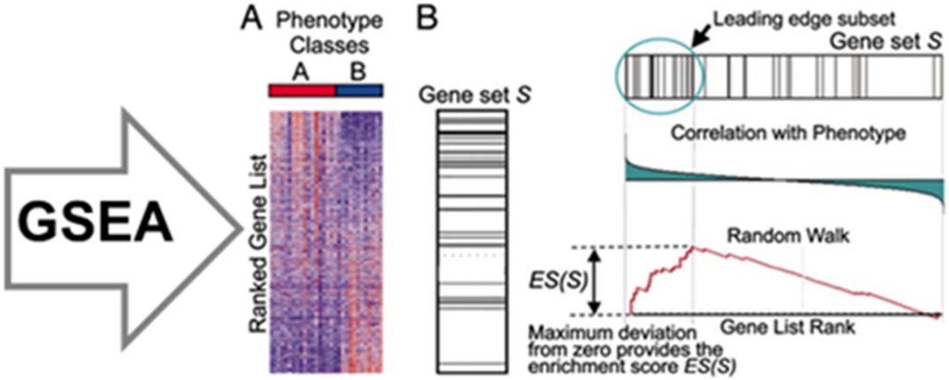
Metagene Projection

Cluster Samples into Molecular Subtypes

Discover directions associated with biological mechanisms

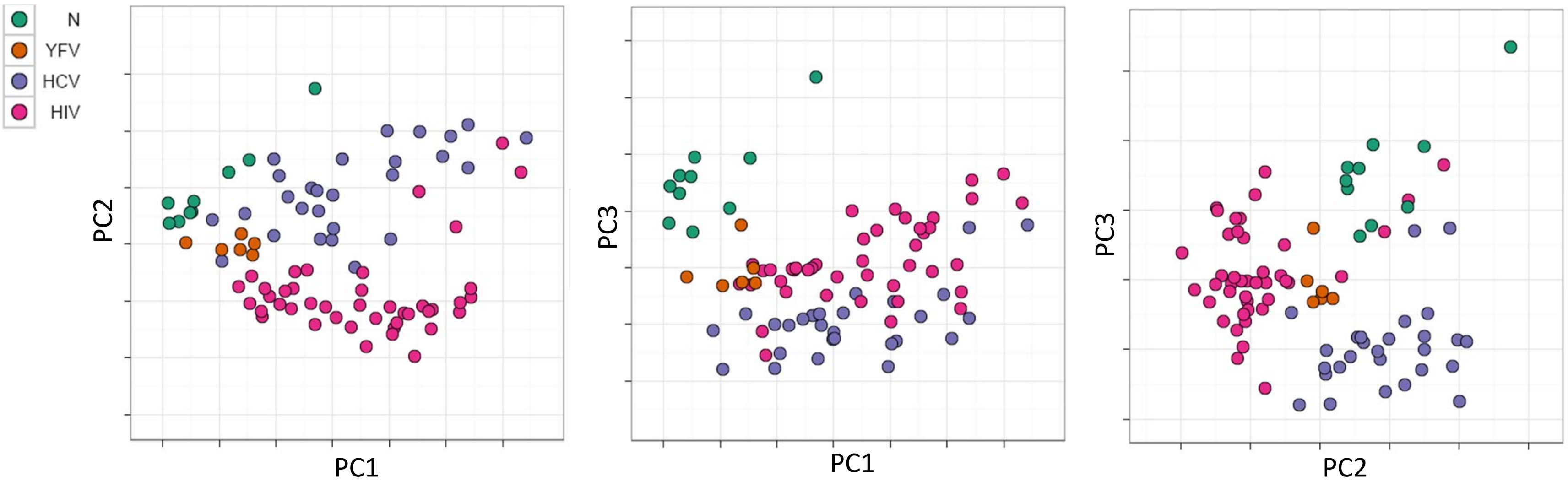
Tamayo *et al*, PNAS 2007

Discover Pathways associated with the Molecular Subtypes

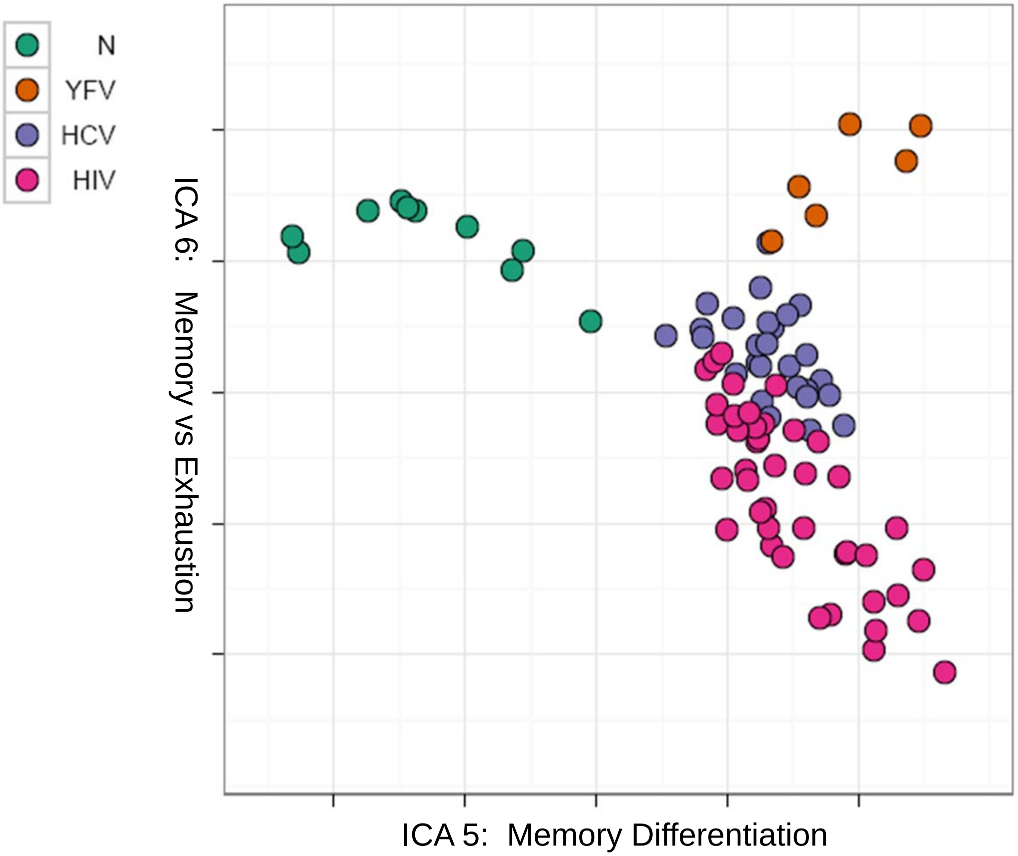


Subramanian *et al*, PNAS 2005

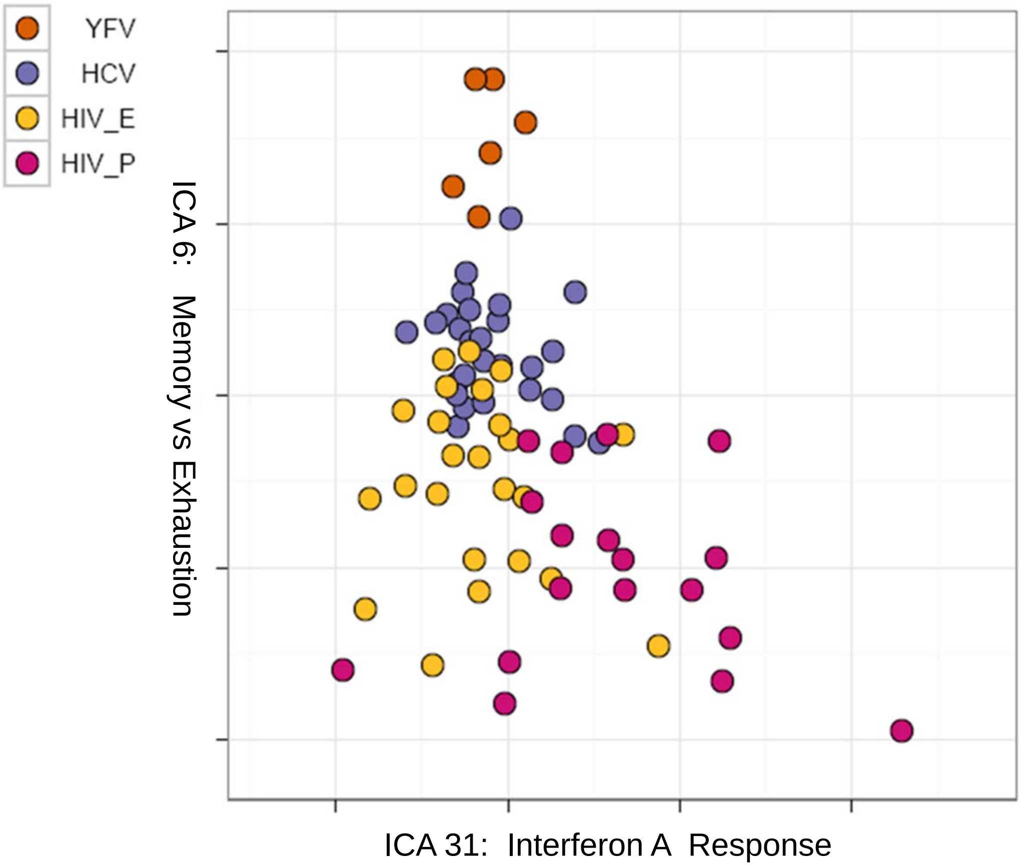
Principal component analysis (PCA) is unable to distinguish between antigen-specific CD8 T responses to acute and chronic viral pathogens. (Supplementary Material)



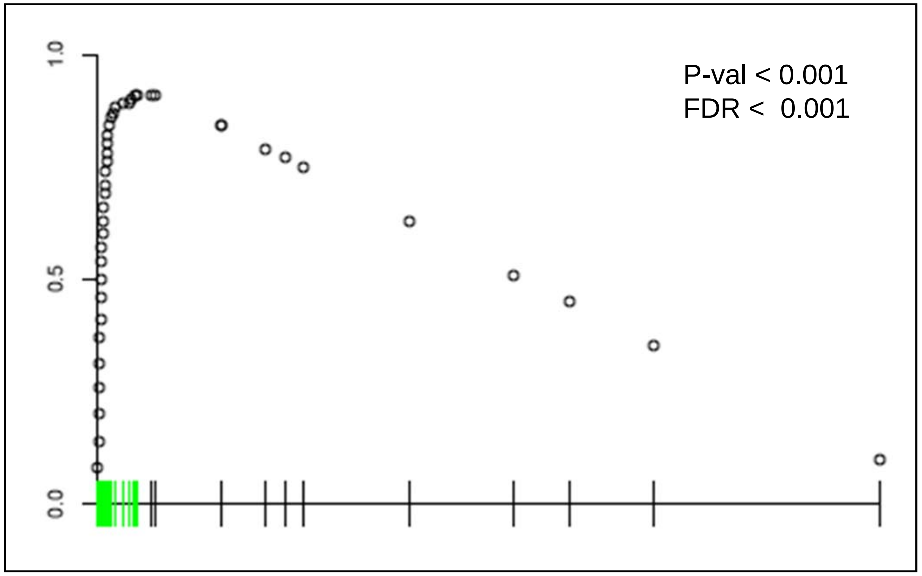
Metagene projection based on independent component analysis (ICA) distinguishes between antigen-specific CD8 T cells responding to different viral pathogens.



Metagene projection based on ICA can also distinguish antigen-specific CD8 T cells from HIV elite vs HIV progressors. The metagene which differentiate elite vs progressor T cell responses is enriched in Interferon signatures.



The component that distinguished between naive and antigen-experienced cells is enriched for known memory signatures.*

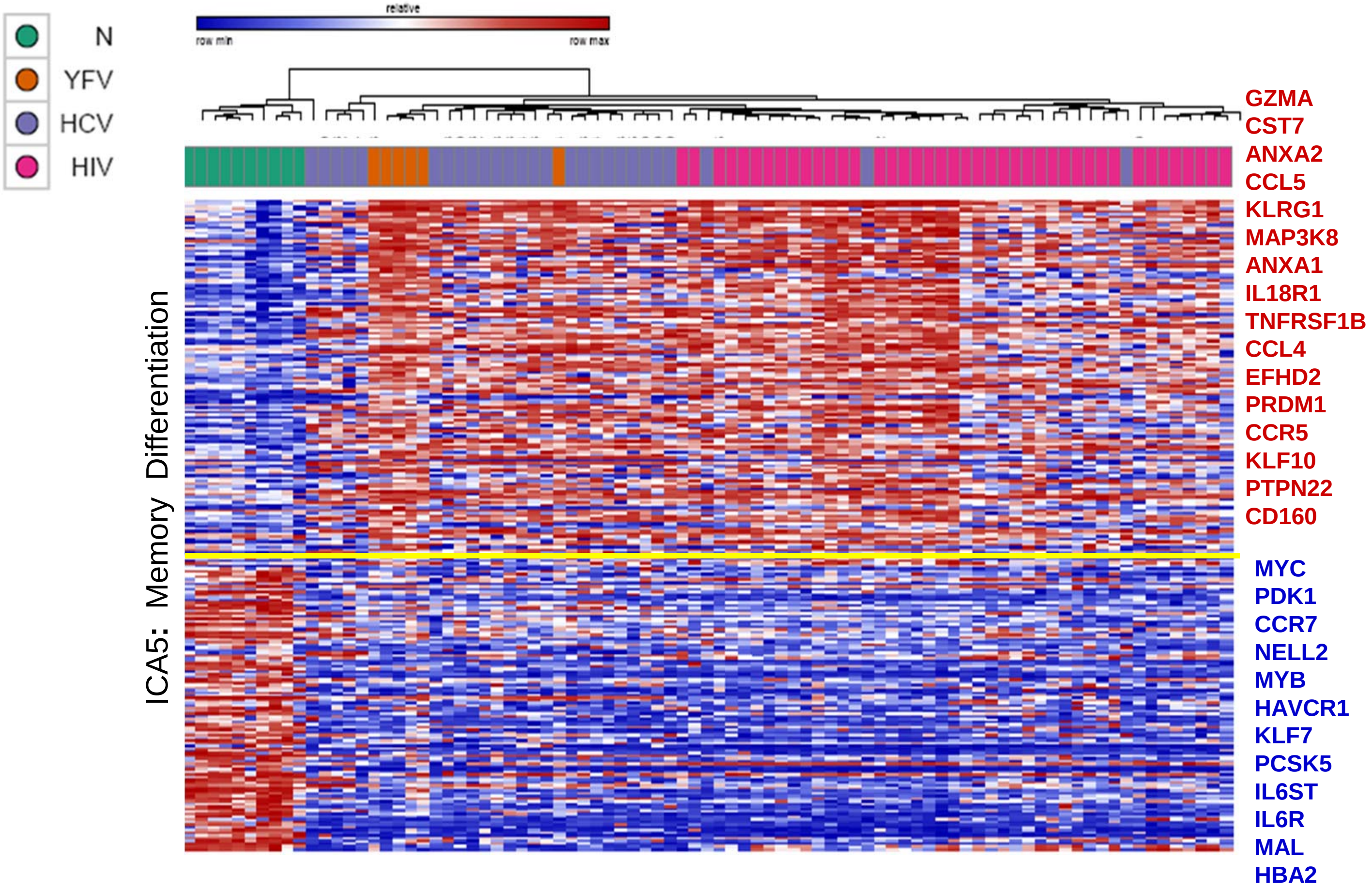


Haining Conserved Differentiation Signature

*Haining et al., *J Immunol* 2008, 181:1859-1868

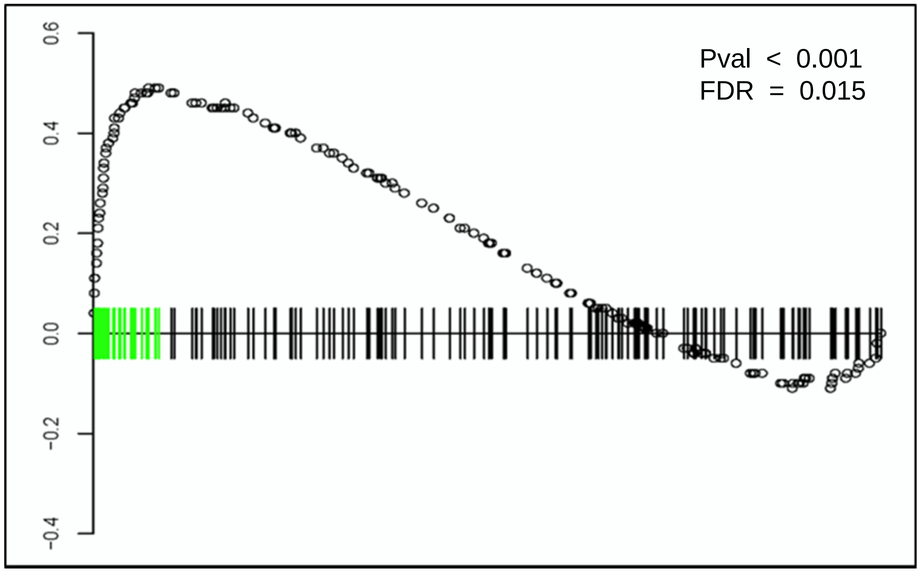
Leading Edge genes are highlighted green

Heatmap showing relevant end point genes in the component that distinguishes between naive and antigen experienced cells is enriched for known memory signatures.



*Haining et al., *J Immunol* 2008, 181:1859-1868

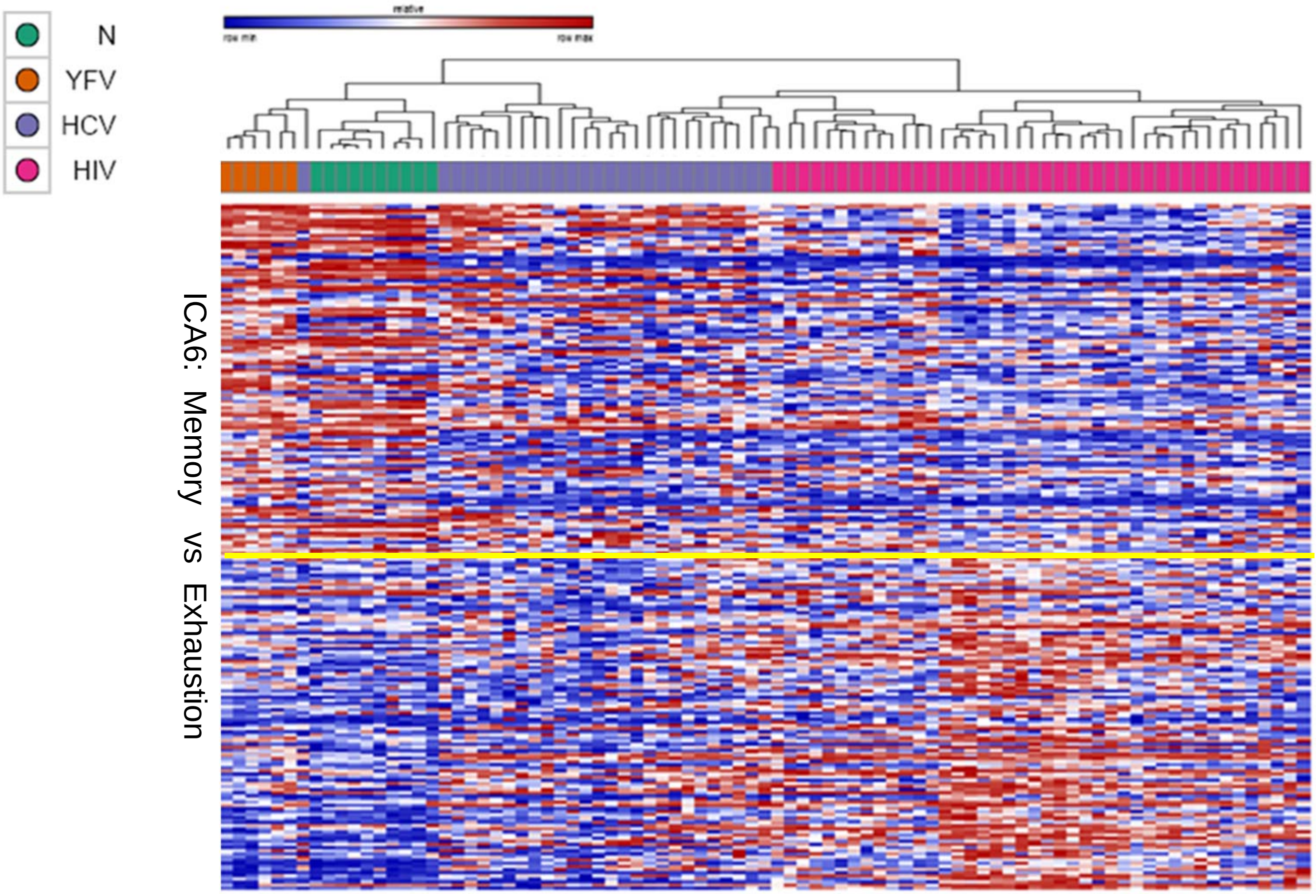
The component that distinguishes between acute and chronic antigen-specific T cells enriches for memory vs. exhaustion.



Wherry CD8 D30 Chronic Up Acute Down

* Wherry et al., *Immunity* 2012

The component that distinguishes between acute and chronic antigen-specific T cells enriches for memory vs. exhaustion signatures.

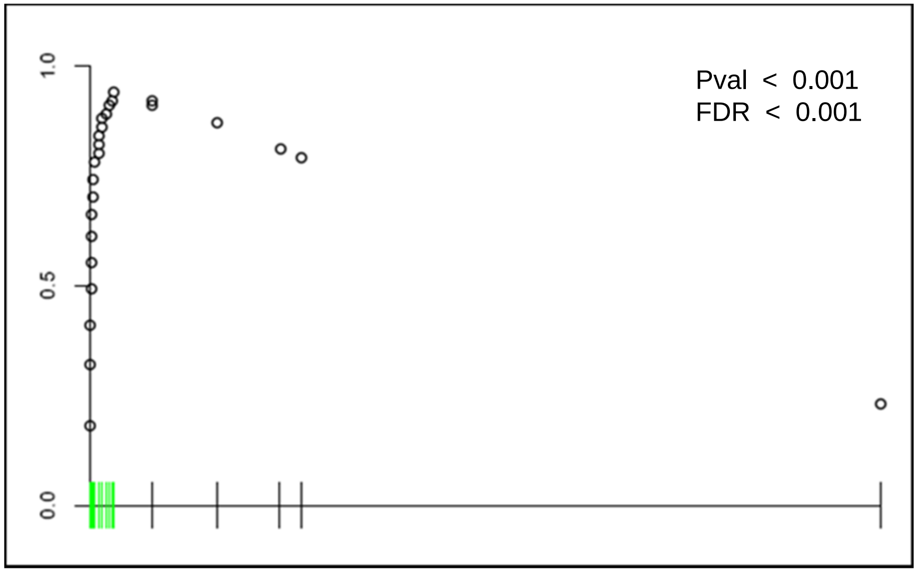


GNLY
ITGAM
LEF1
TNFRSF25
KLRB1
CD28
DPP4
SATB1
IL18RAP
SELL
IL7R
MYC
GADD45A
CCR7
IL18R1

RSAD2
RNASEH2B
IFIH1
TANK
NUSAP1
MX2
CASP1
SH2D1A
EZH2
PTTG1
NAB1
TOX
AIM2
CD244

* Wherry et al., *Immunity* 2012

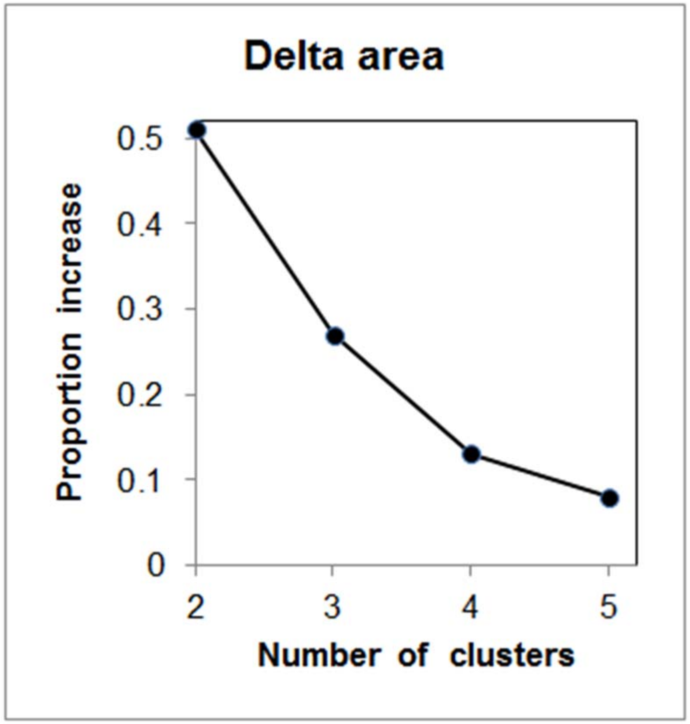
The component that distinguishes between HIV elite and HIV progressor antigen-specific T cells enriches for Interferon alpha signatures.



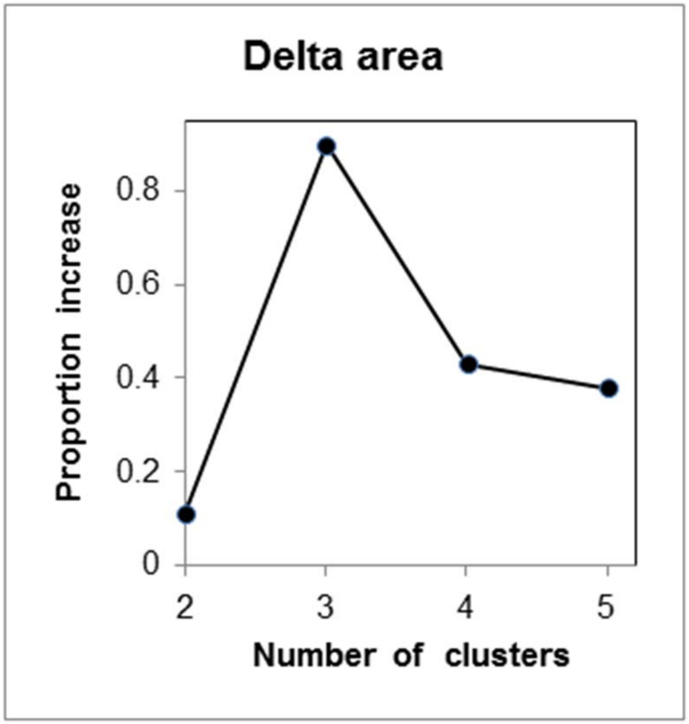
Moserle Interferon A Response Signature

Statistics based on Delta Area* estimate 2 populations of T cell responses to HCV and 3 populations of T cell responses to HIV.

Proportion increase in Delta Area* estimates two clusters in HCV data

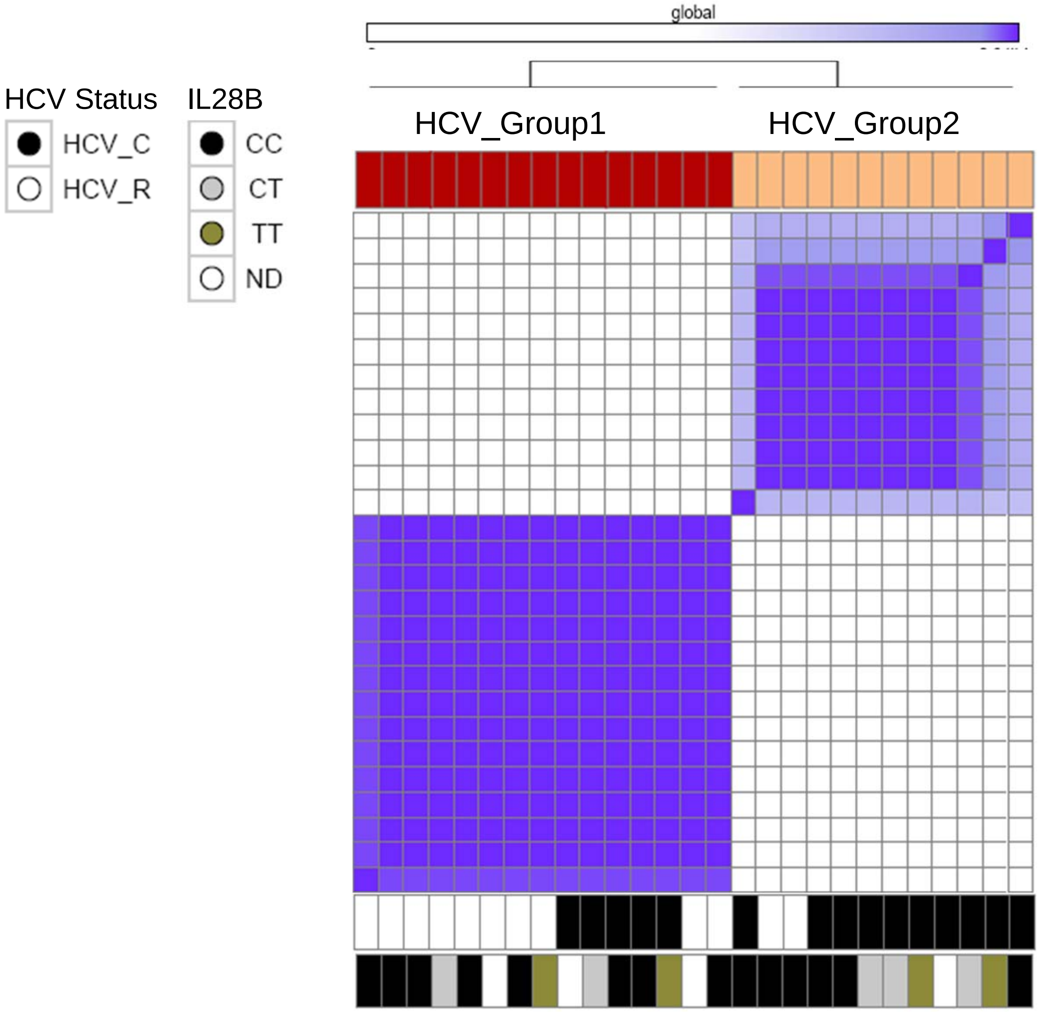


Proportion increase in Delta Area* estimates three clusters in HIV data



*Monti et al, Consensus Clustering, 2005,
GenePattern: Consensus Clustering

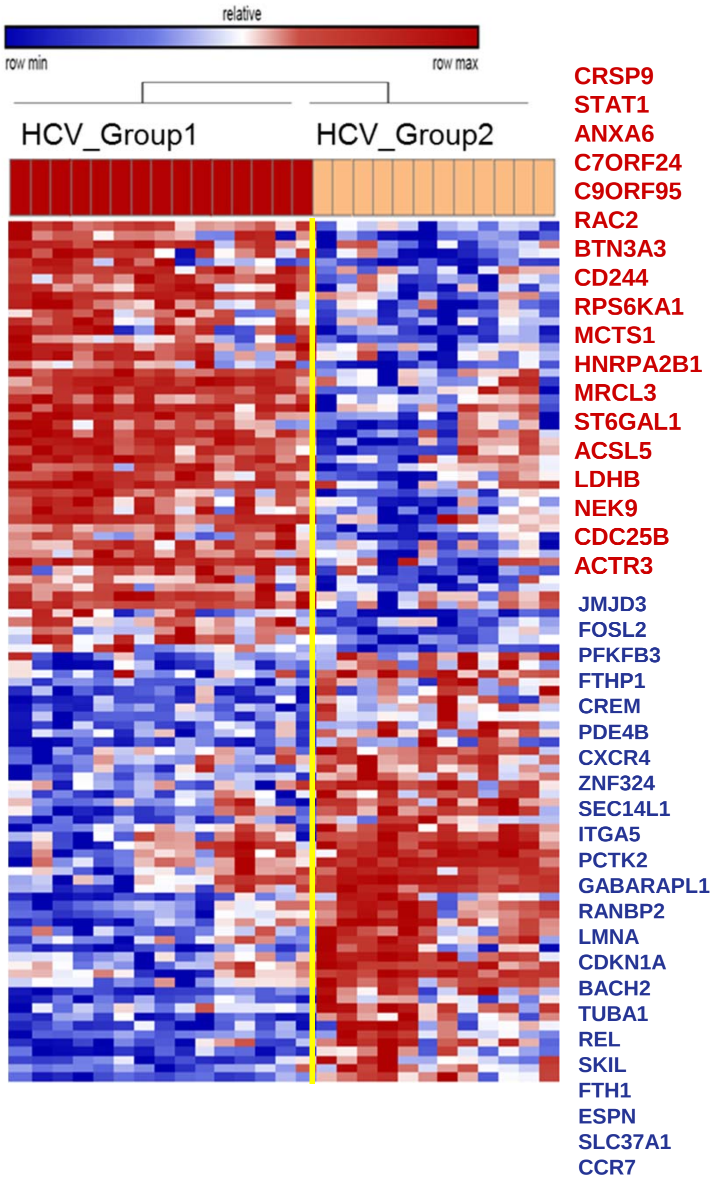
There is heterogeneity in the profiles of T cells responding to chronic infection: Unsupervised analysis shows 2 populations in the profiles of HCV-specific T cell responses.



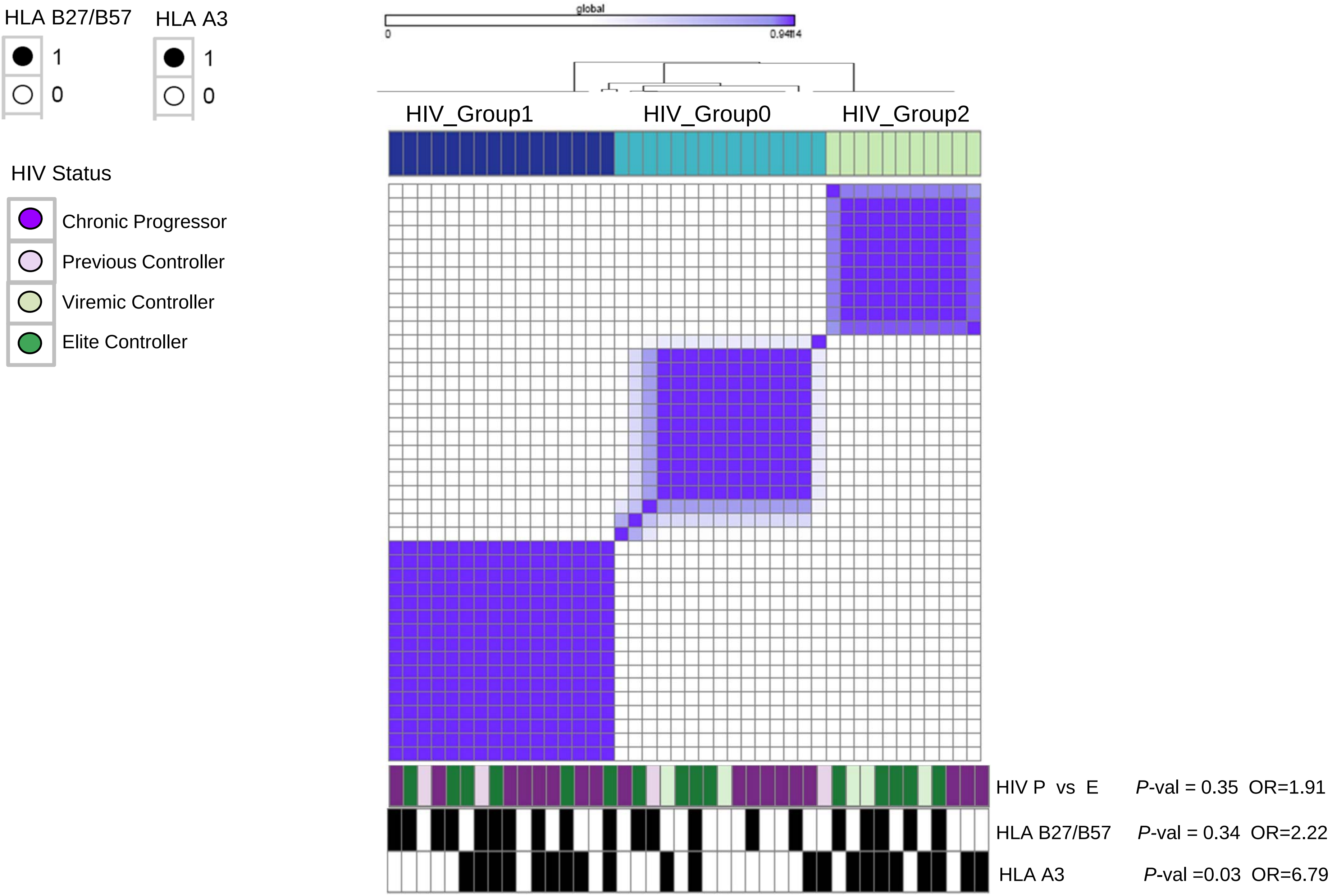
HCV Status: C vs R assoc *P*-val =0.02, OR=9.05

IL28B Genotypes assoc *P*-val > 0.8

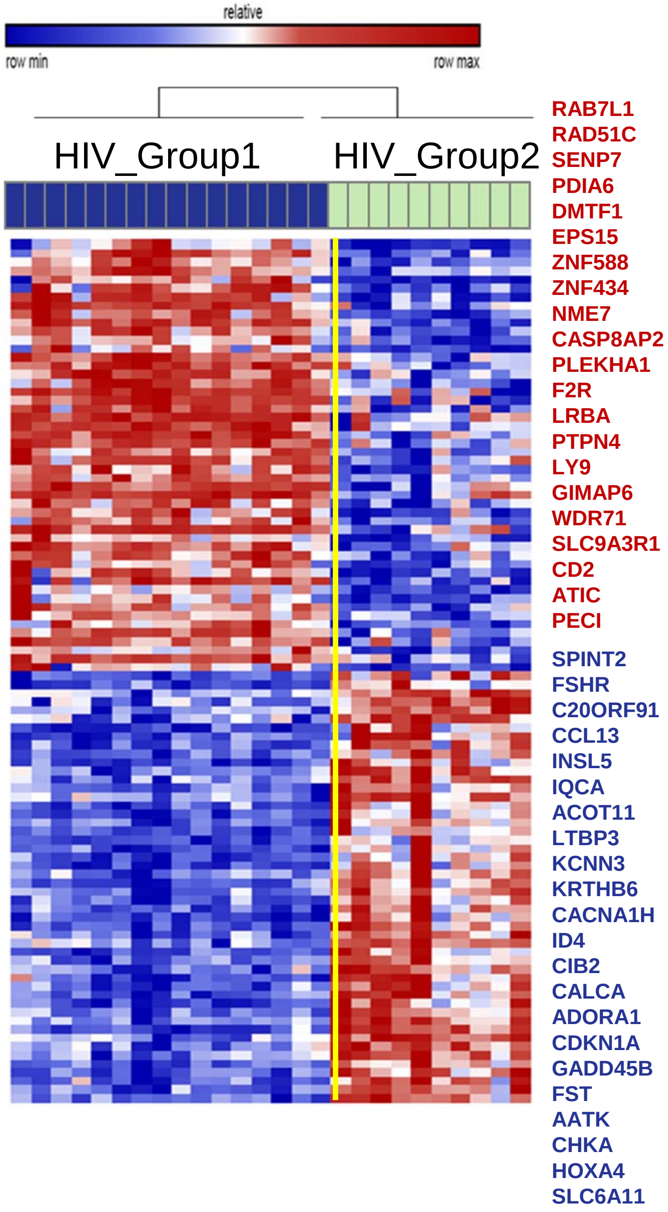
Unsupervised analysis shows 2 populations in profiles of HCV-specific T cell responses. Heatmap of top differentially expressed genes in the HCV Groups. (Supplementary Material)



There is heterogeneity in the profiles of T cells responding to chronic infection.
Unsupervised analysis shows 3 populations of profiles in HIV-specific T cell responses.

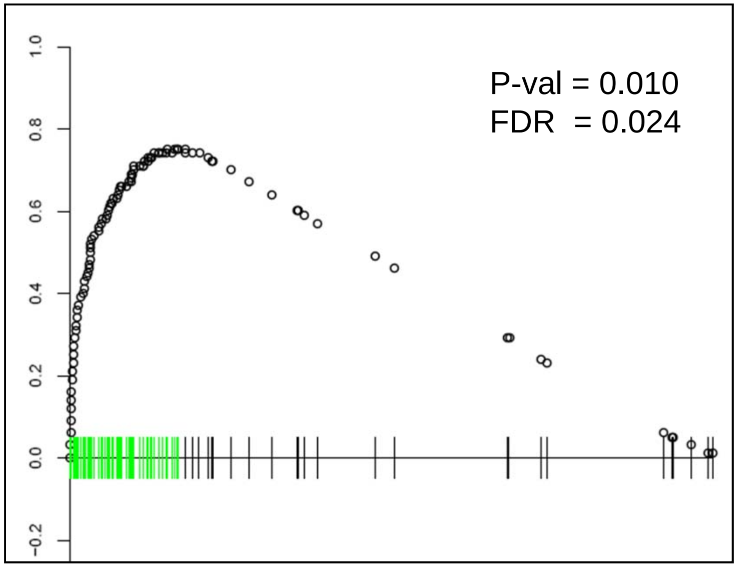


There is heterogeneity in the profiles of T cells responding to chronic infection.
Heatmap of top differentially expressed genes in the HCV groups.
(Supplementary Material)



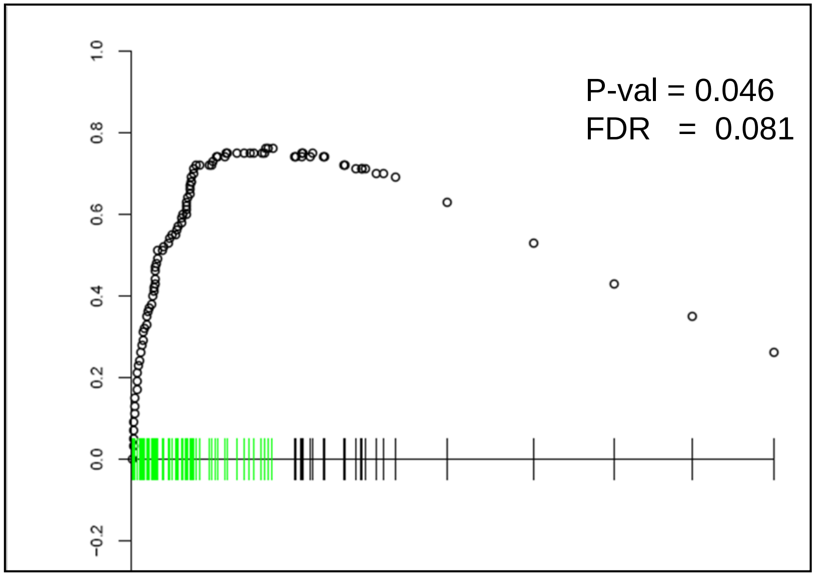
HCV_Group1 and HIV_Group1 expression profiles are significantly similar.

HCV_Group1 vs HCV_Group2



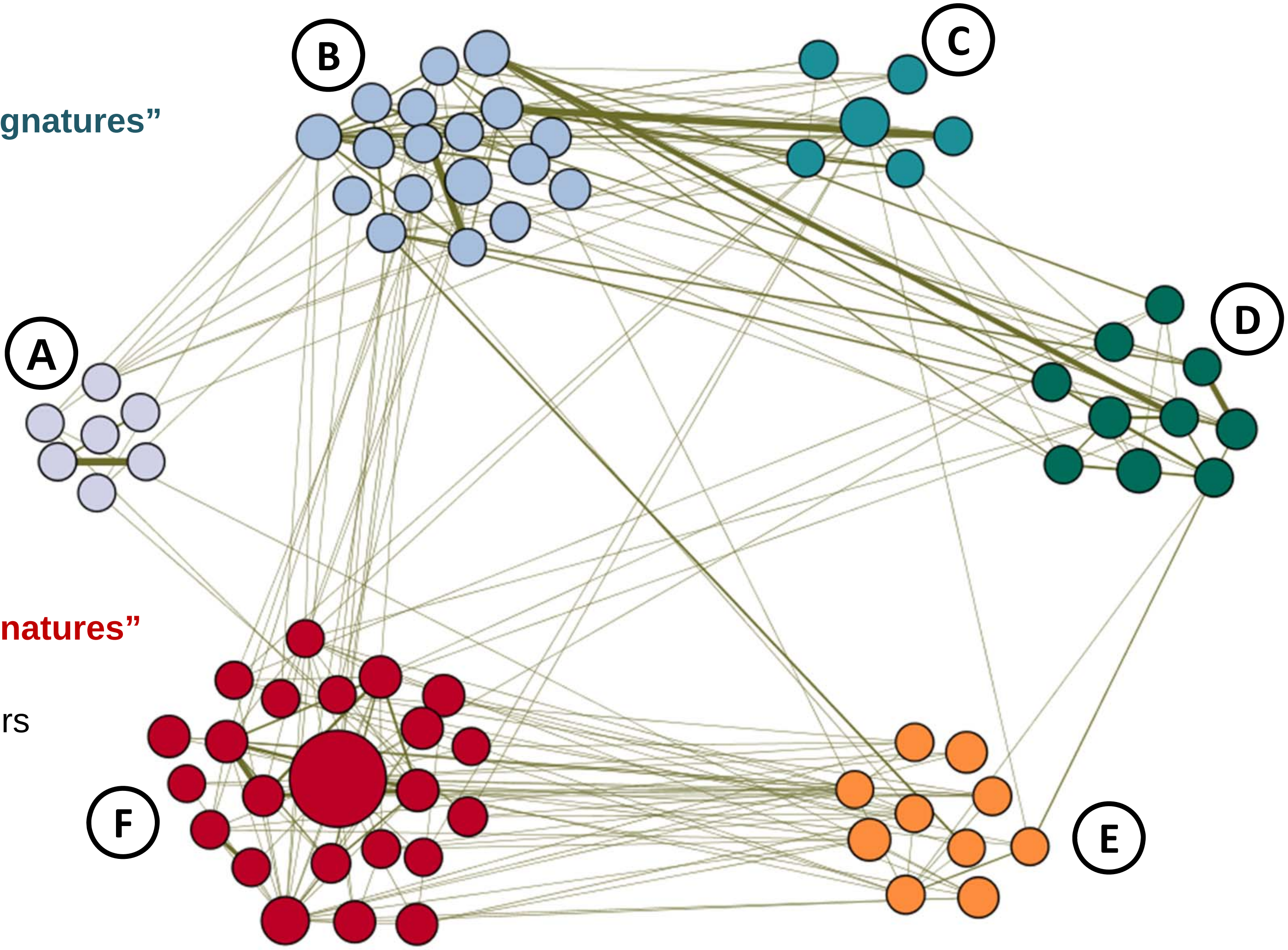
Top 100 genes high in HIV_Group1 vs HIV_Group2

HIV_Group1 vs HIV_Group2



Top 100 genes high in HCV_Group1 vs HCV_Group2

Enrichment map* for the gene set signatures in HCV populations of T cell responses shows four functional clusters of “Good Signatures” and two functional clusters of “Poor Signatures”.



HCV Group 1 : “ Good Signatures”

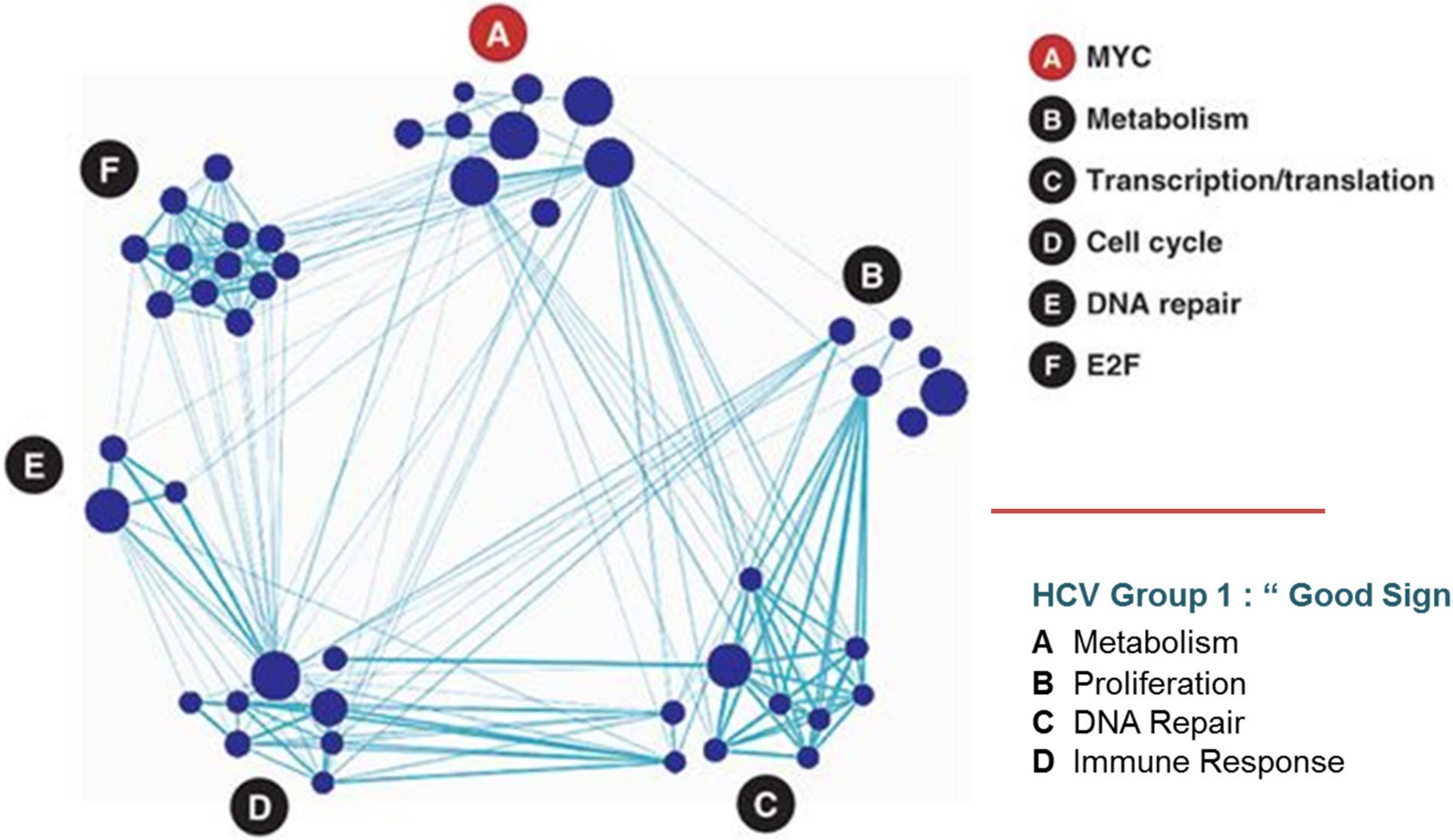
A Metabolism
B Proliferation
C DNA Repair
D Immune Response

HCV Group 2 : “Poor Signatures”

E Inflammation
F AP1 Transcription Factors

*Bader Lab, *PLoS One* 2010

Enrichment Map for JQ1 DN Consensus Signature across Neuroblastoma, AML and Multiple Myeloma cell lines



HCV Group 1 : “ Good Signatures”

A Metabolism
B Proliferation
C DNA Repair
D Immune Response

Conclusions

- Disease states can be described by alterations at various levels of hallmark functionality
- Changes involved in “fundamental” biological hallmarks are synergistic across diseases and tissues at large scale and micro scale levels
- Metagene projection strategies are able to provide sensitive solutions to represent biological data in a “biological”-interpretable way
- Information theoretic strategies provide a “bridge” towards understanding biology at a micro and quantum level.

Collaborators

Harvard Medical School
Stegmaier Lab

- **Alex Puissant**
- **Stacey Frumm**
- **Yana Pikman**

Haining Lab

- **Prakash Gupta**
- Kathleen Yates
- Catia Fonseca
- Yan Tan
- Robert Barnitz
- Jernej Godec

Cancer Institute of NJ
Bhanot Lab

Ganesan Lab

Broad Institute
Mesirov Lab

- **Pablo Tamayo**

Train to discover hidden biological dimensions!

XColony System: Courtesy to Sorin Alexe