Cell-Based Microarrays: Current Status, Future Prospects

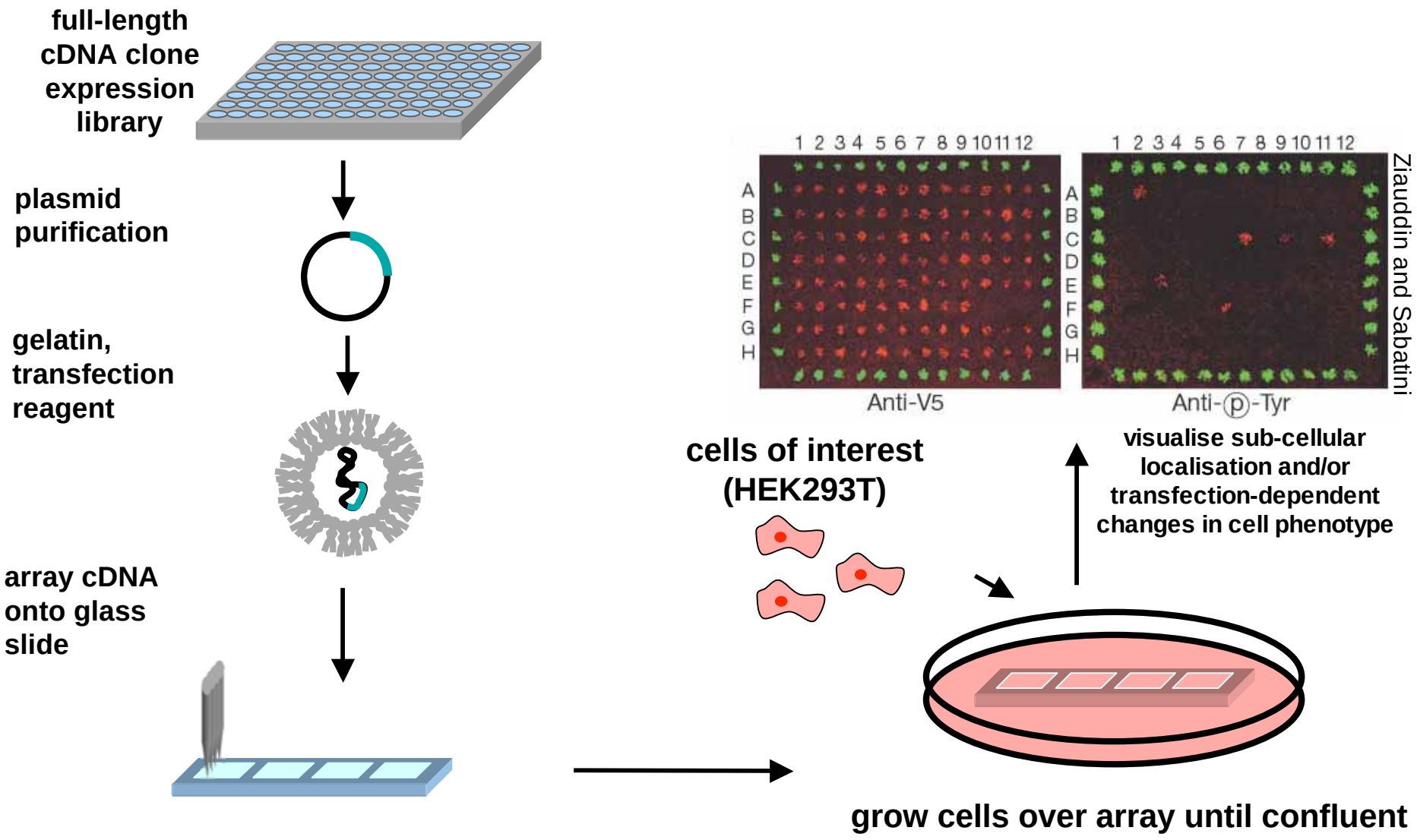
Tom Freeman

Scottish Centre of Genome Technology and Informatics (GTI),

University of Edinburgh Medical School

Reverse Transfection Methodology

Ziauddin and Sabatini Nature **411:**107-110 (2001)



Applications of Reverse Transfection Gain of Function Assays

- high throughput screen for protein sub-cellular localisation

- protein interaction studies

- *in vitro* screen for genes that induce desired effect on cell phenotype

Our interest:

- potential for use in identifying candidate genes for use in gene therapy - search for pro-apoptotic genes for treatment of disease - collaboration with Prof. Andrew Miller (Genetic Therapies Centre, Imperial College, London)

MRC Rosalind Franklin Centre for Genomics Research

Technology Set Up and Investigation into Gene Tagging Strategies

- cDNA clones were selected from the full-length MGC clone collection (IRAT) by searching for genes with interesting functional activity e.g. factors, receptors, kinases
- search refined using databases (EMBL, Locuslink) to pull out sequence and description of function
- 20 genes chosen due to differences in protein size and known function
- open reading frame (ORF) transferred into Gateway (Invitrogen) cloning system
- ORF tagged at both N- and C-terminal with GFP

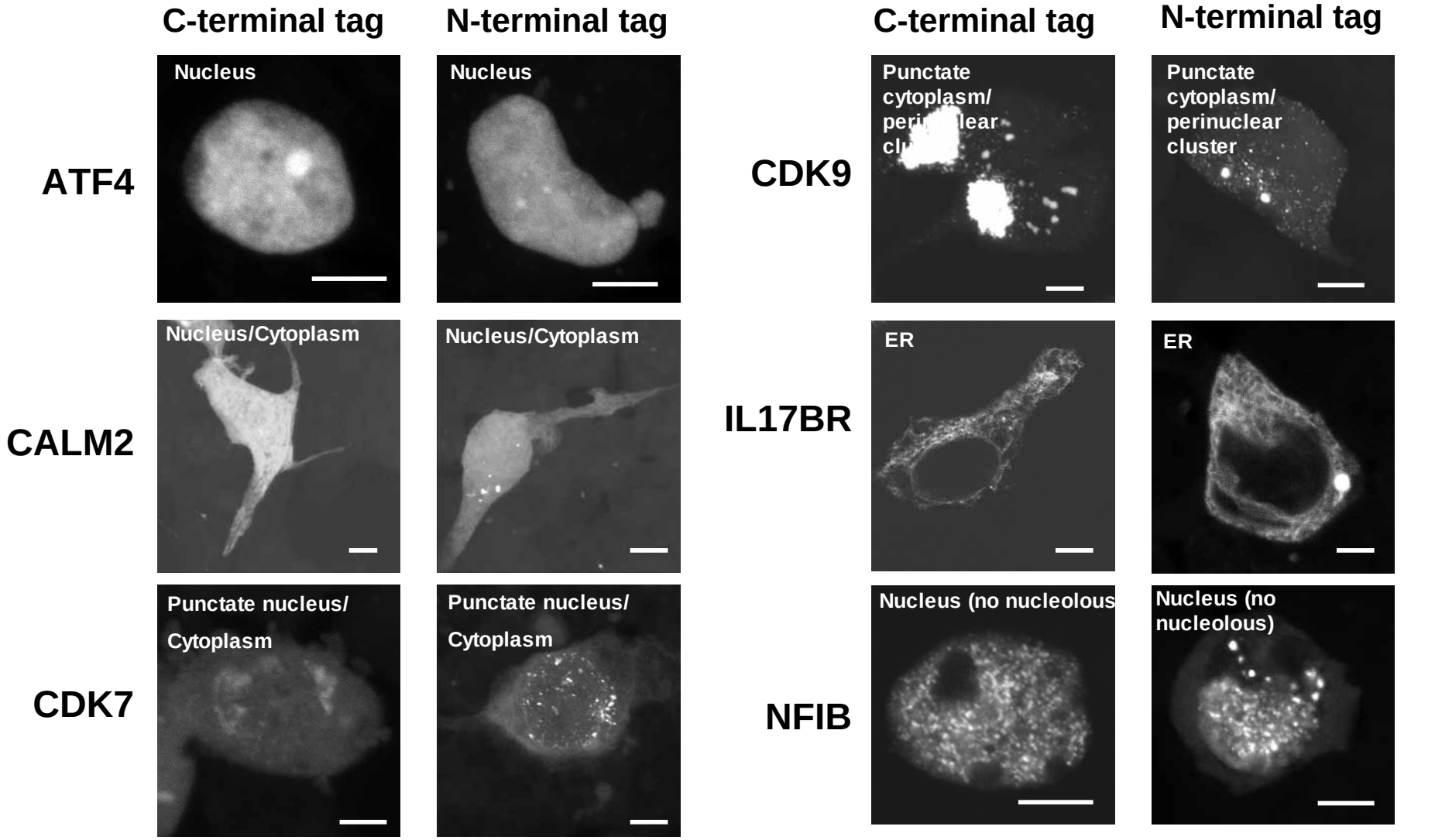
Results

- 16 of the 20 genes were successfully transferred into Gateway vectors

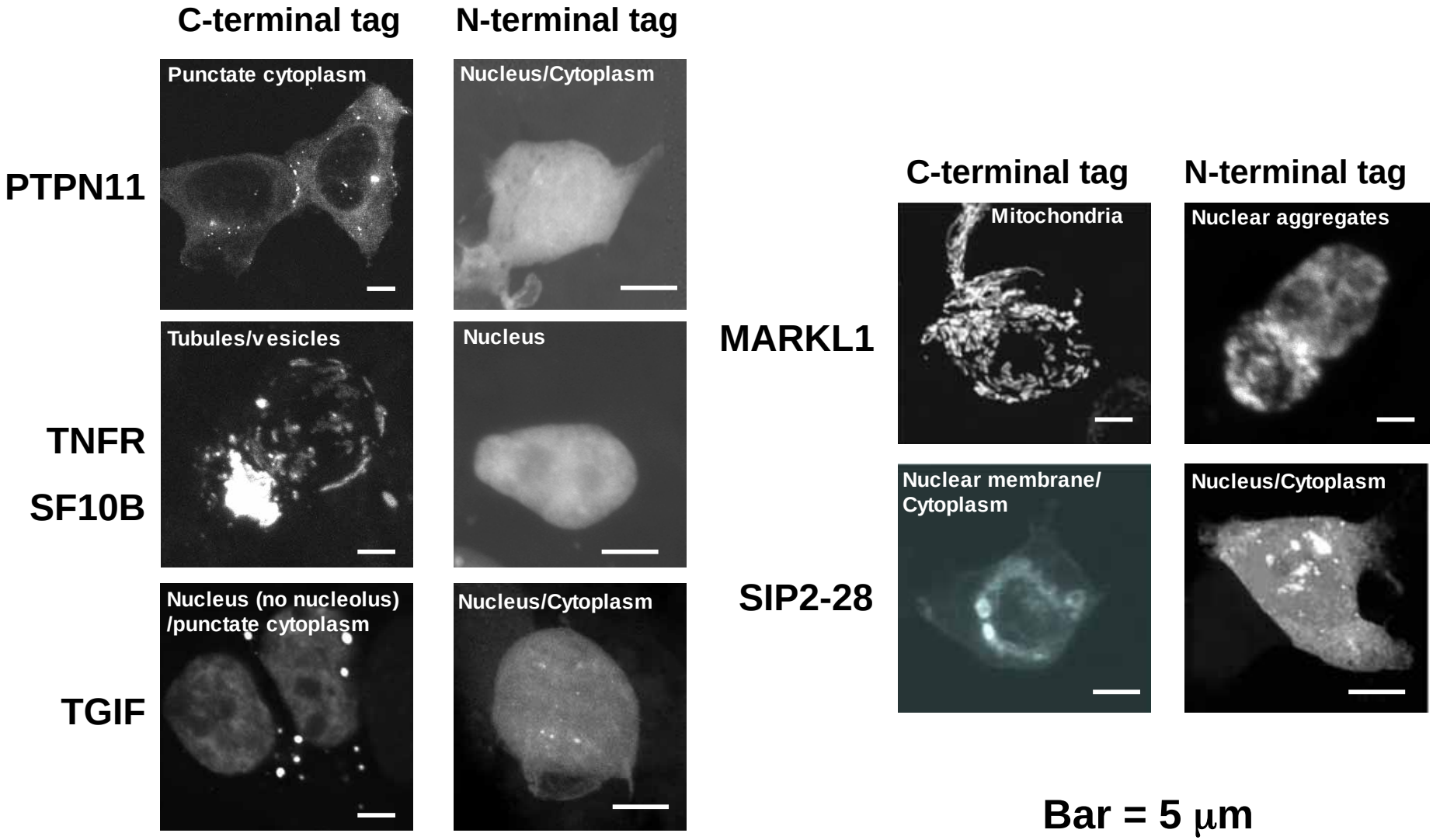
The genes could be divided into 3 classes:

1. C- and N-terminal sub-cellular localisations agreed (6/16)
2. C- and N-terminal sub-cellular localisations did not agree (5/16)
3. no fusion protein observed with N-terminal construct (5/16)

GROUP 1: N and C-terminal same localisation

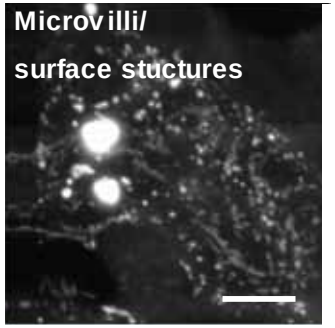


GROUP 2: N and C-terminal different localisations

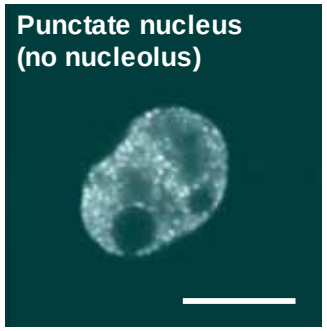


GROUP 3: C-terminal localisations only

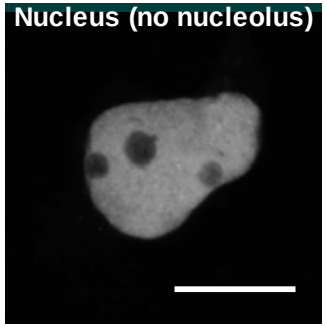
CXADR



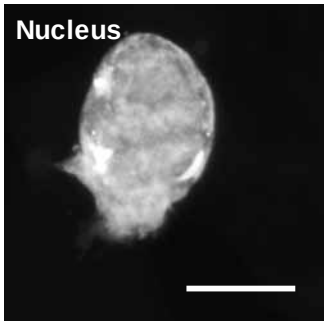
NFIL3



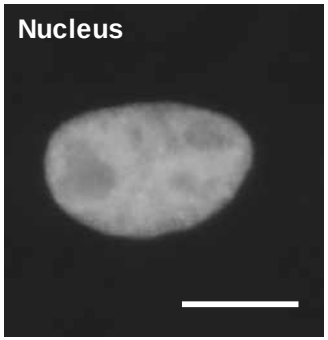
PPARG



STK15



CDKN1B



Bar = 5 μm

Conclusions

- technique works!
- in all cases sub-cellular localisation of C-terminal GFP construct correct as compared to that reported/predicted
- localisations same as those observed with well-based assays
- GFP provides good positive control for transfection and can be visualised in living cells
- other tags e.g. His-tag require post-processing of the arrays to allow visualisation of transfected protein which can damage the cell monolayer

Palmer E. and Freeman T. Investigation into the use of C- and N-terminal GFP fusion proteins for sub-cellular localisation studies using reverse transfection microarrays. *Comp. Funct. Genomics* **5: 342-353 (2004).**

Construction of High Density Reverse Transfection Array

Why?

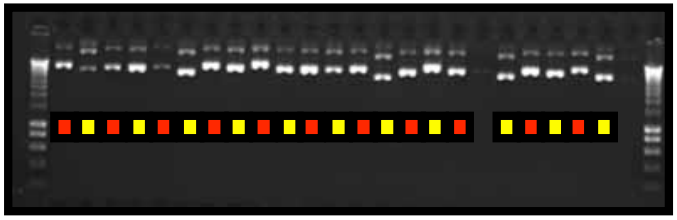
- not much point otherwise
- limited reagents

Issues

- transfer of ORF's from MGC clones into Gateway system time consuming and expensive for large numbers of genes
- transfer may introduce errors in ORF
- gene-tags can effect localisation (therefore function)

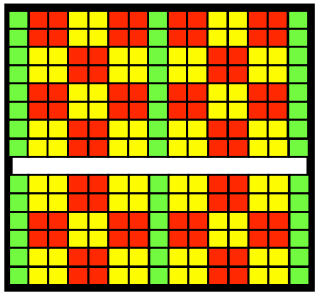
Strategy

- use MGC clones (IRAT) direct for construction of high density array

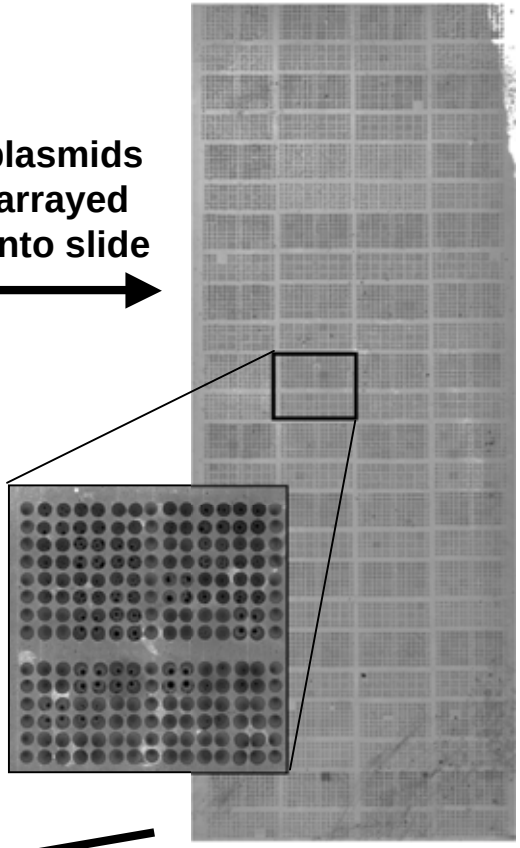


2,796 MGC IRAT clones used for plasmid prep.

arrayed in quadruplet on each slide, each sub-grid being surrounded by a GFP-only expression vector



plasmids arrayed onto slide



(post-array scan) array comprised of 10,080 features, 1,959 clones

Reverse transfection

(fluorescence scan) grids outlined by GFP

High Density Gain of Function Screen

Functional assays

phospho-tyrosine (Ab)

apoptosis inducers (TUNEL)

verification and functional characterisation

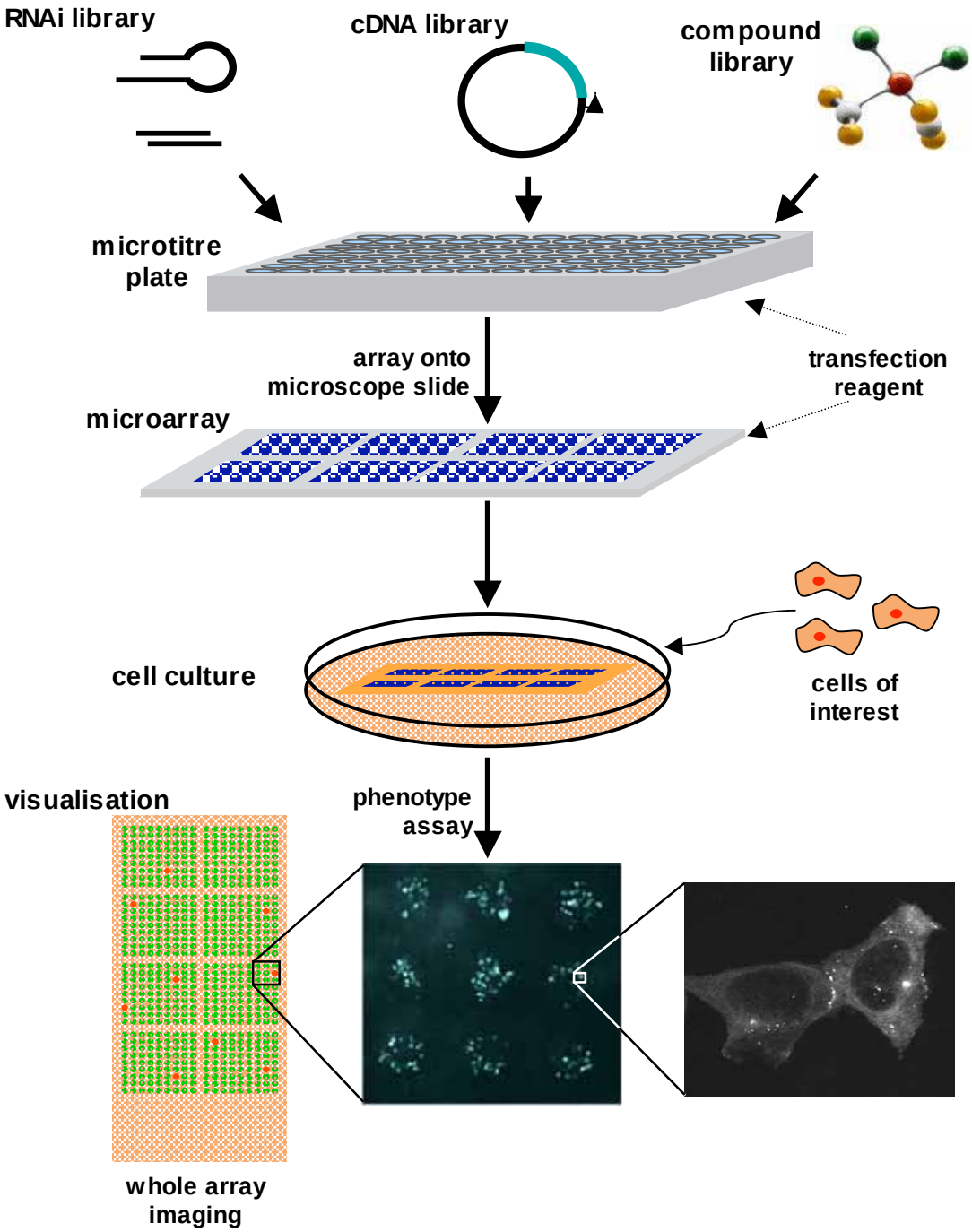
Palmer E.L., Miller A.D. and Freeman T.C. BMC Genomics **7**:145 (2006)

GENE	Percentage of cells undergoing apoptosis									
	12 hour		24 hour		36 hour		48 hour		60 hour	
	TUNEL	CASP3	TUNEL	CASP3	TUNEL	CASP3	TUNEL	CASP3	TUNEL	CASP3
CSTB	5	0	5	30	10	10	10	10	10	0
MGC5439	0	20	10	20	10	20	10	40	60	60
C22orf23	30	10	30	10	60	10	40	20	40	40
AF1Q	10	10	10	10	10	10	10	10	60	40
CCBP2	N/A	10	1	20	1	20	1	30	10	50
LOC134285	N/A	20	5	20	5	20	5	25	30	65
EXOC7	N/A	20	1	30	1	40	60	50	60	60
STK3	0	20	0	20	10	20	10	50	40	60
ACO1	40	10	40	10	40	20	40	30	70	50
XBP1	1	2	1	2	10	5	60	10	60	70
STS	N/A	20	N/A	20	N/A	30	N/A	50	N/A	60
Mock transfection	2	2	2	2	2	2	3	2	4	3

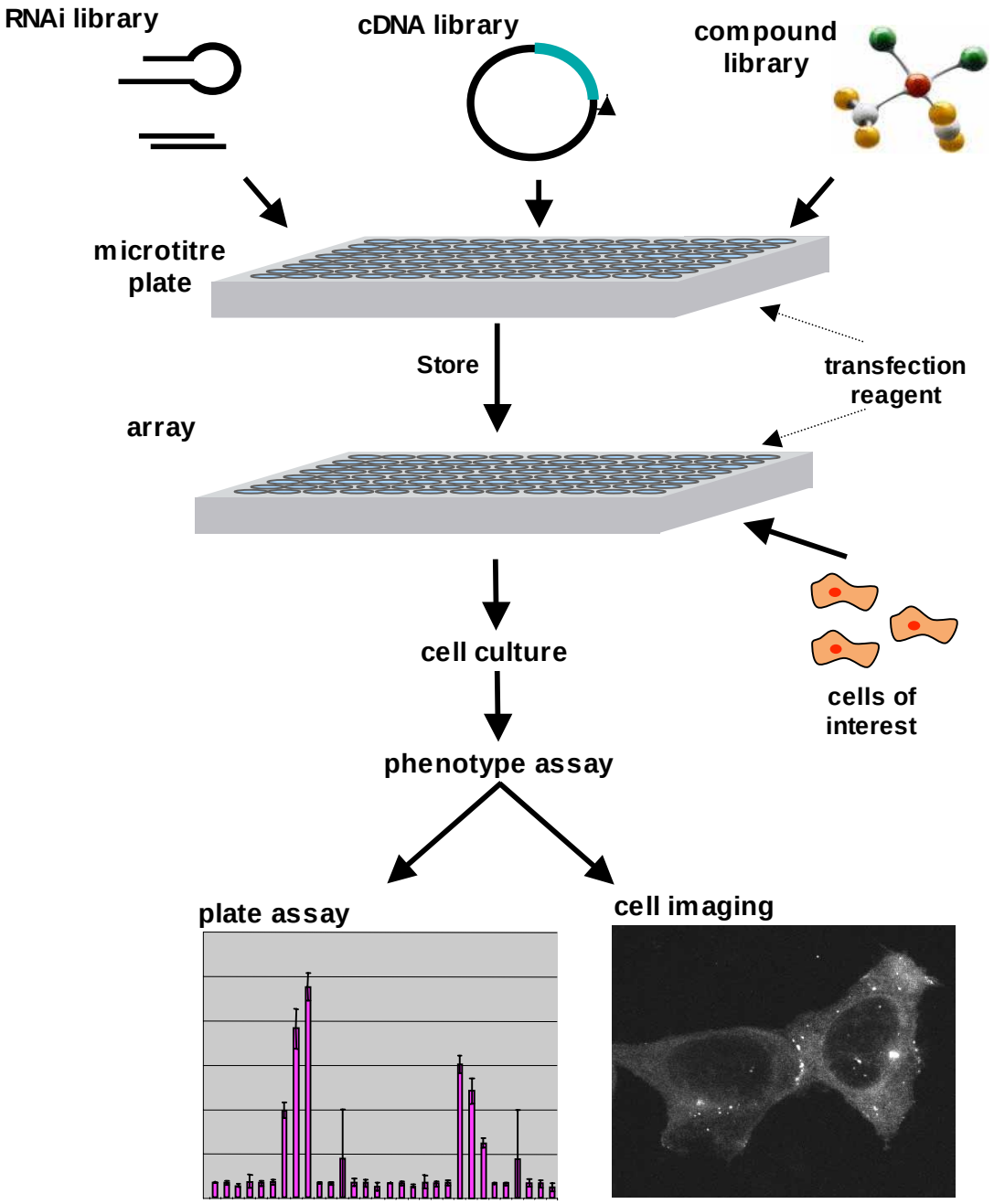
10/79 gene ‘hits’ verified as inducing apoptosis – 7 classified (annotated gene name), 3 unclassified

- **1 gene known to be associated with apoptotic pathway**
- **1 over-expression gives rise to apoptosis through indirect route**
- **2 considered interesting by our collaborators on apoptosis**
- **6 no previous association to apoptosis**

Cell-based Microarrays



Cell-based ‘arrays’



Final Comments on Cell-Based Arrays

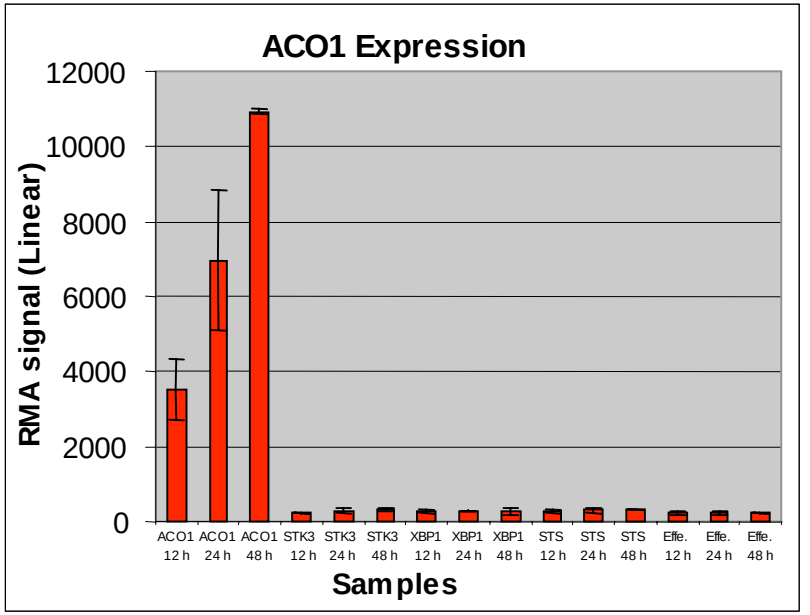
- reverse transfection microarray technology is a challenge to set up and in our hands proved to be a temperamental technique – many variables
- requires significant resources to in terms of access to high quality reagent resources, printing and scanning technology
- many aspects simplified by use of plate-based assays
- need robust phenotype assays and good verification/functional characterisation pipeline
- Having said this, the technology can provide high throughput functional screen properly resourced

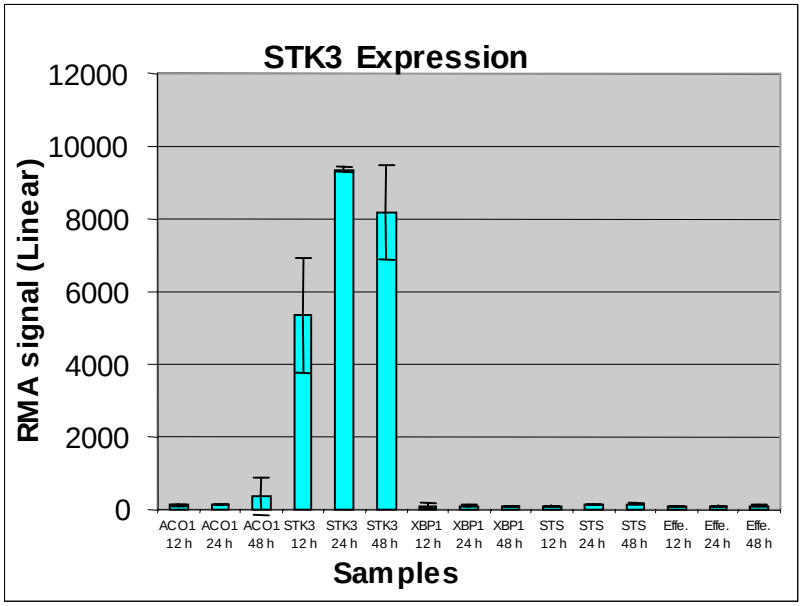
Found Genes but by What Mechanism Do They Induce Apoptosis?

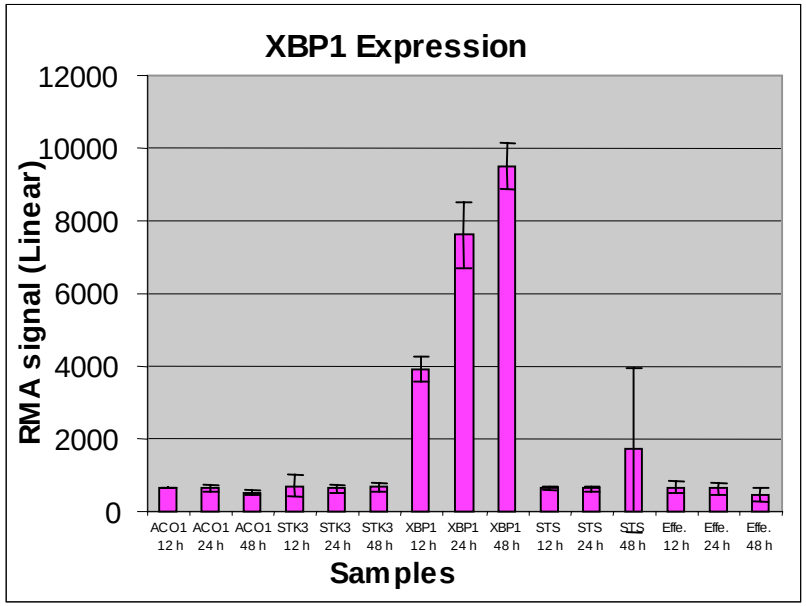
GENE	Percentage of cells undergoing apoptosis									
	12 hour		24 hour		36 hour		48 hour		60 hour	
	TUNEL	CASP3	TUNEL	CASP3	TUNEL	CASP3	TUNEL	CASP3	TUNEL	CASP3
CSTB	5	0	5	30	10	10	10	10	10	0
MGC5439	0	20	10	20	10	20	10	40	60	60
C22orf23	30	10	30	10	60	10	40	20	40	40
AF1Q	10	10	10	10	10	10	10	10	60	40
CCBP2	N/A	10	1	20	1	20	1	30	10	50
LOC134285	N/A	20	5	20	5	20	5	25	30	65
EXOC7	N/A	20	1	30	1	40	60	50	60	60
STK3	0	20	0	20	10	20	10	50	40	60
ACO1	40	10	40	10	40	20	40	30	70	50
XBP1	1	2	1	2	10	5	60	10	60	70
STS	N/A	20	N/A	20	N/A	30	N/A	50	N/A	60
Mock transfection	2	2	2	2	2	2	3	2	4	3

- **3 genes selected, STS and mock transfection samples then subjected to expression analysis on duplicate Affymetrix U133 2plus arrays at 12, 24 and 48 h (30 chips total)**

Plots of Affymetrix Expression data of 3 Transfected Genes Across All Conditions

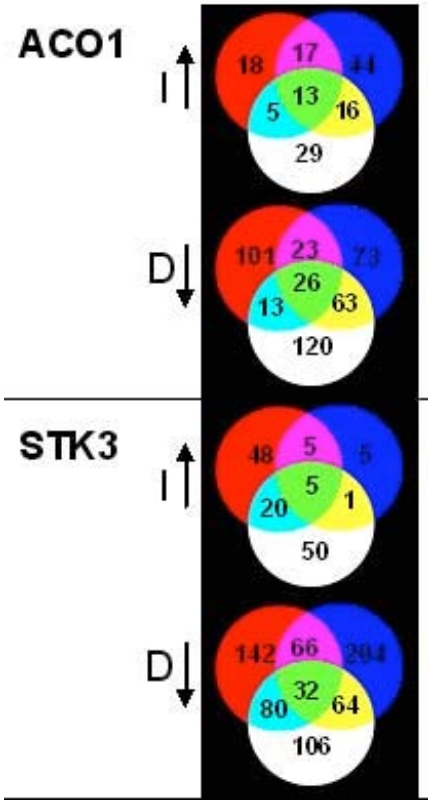






Up and Down Regulation of Gene Expression in Experimental Conditions vs. Time Matched Mock Transfection

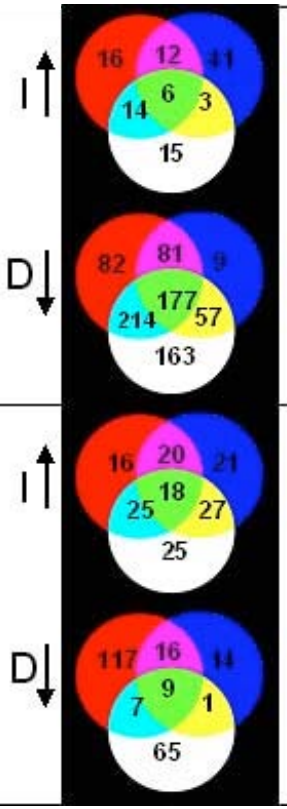
ACO1



Hour	No. genes	% I+D
12	53	24
24	90	33
48	63	22

12	163	76
24	185	67
48	222	78

XBP1



12	48	8
24	62	16
48	38	6

12	554	92
24	324	84
48	611	94

STK3

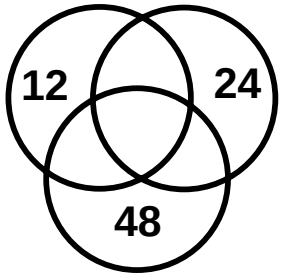
12	78	20
24	16	4
48	76	21

12	320	80
24	366	96
48	282	79

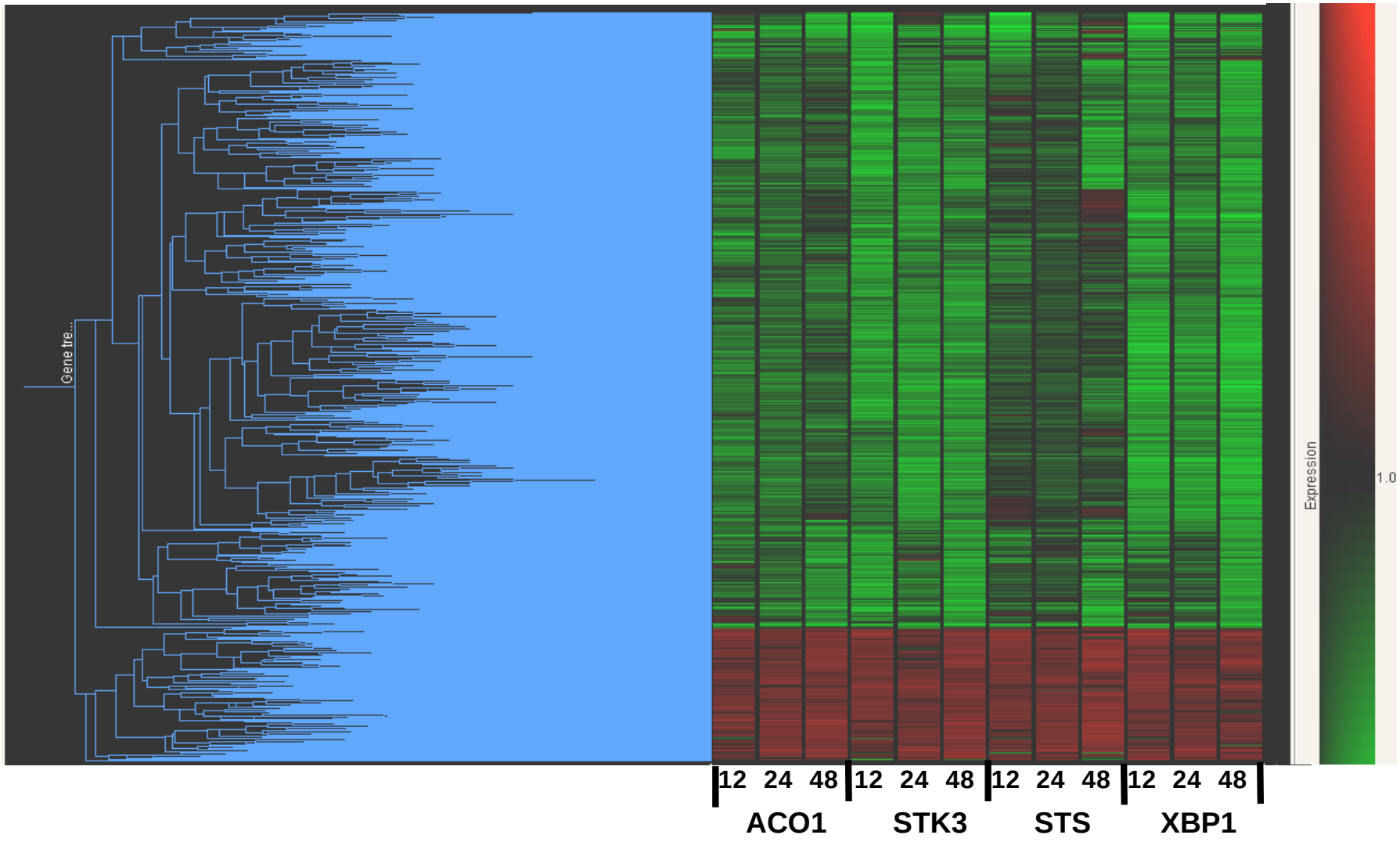
STS

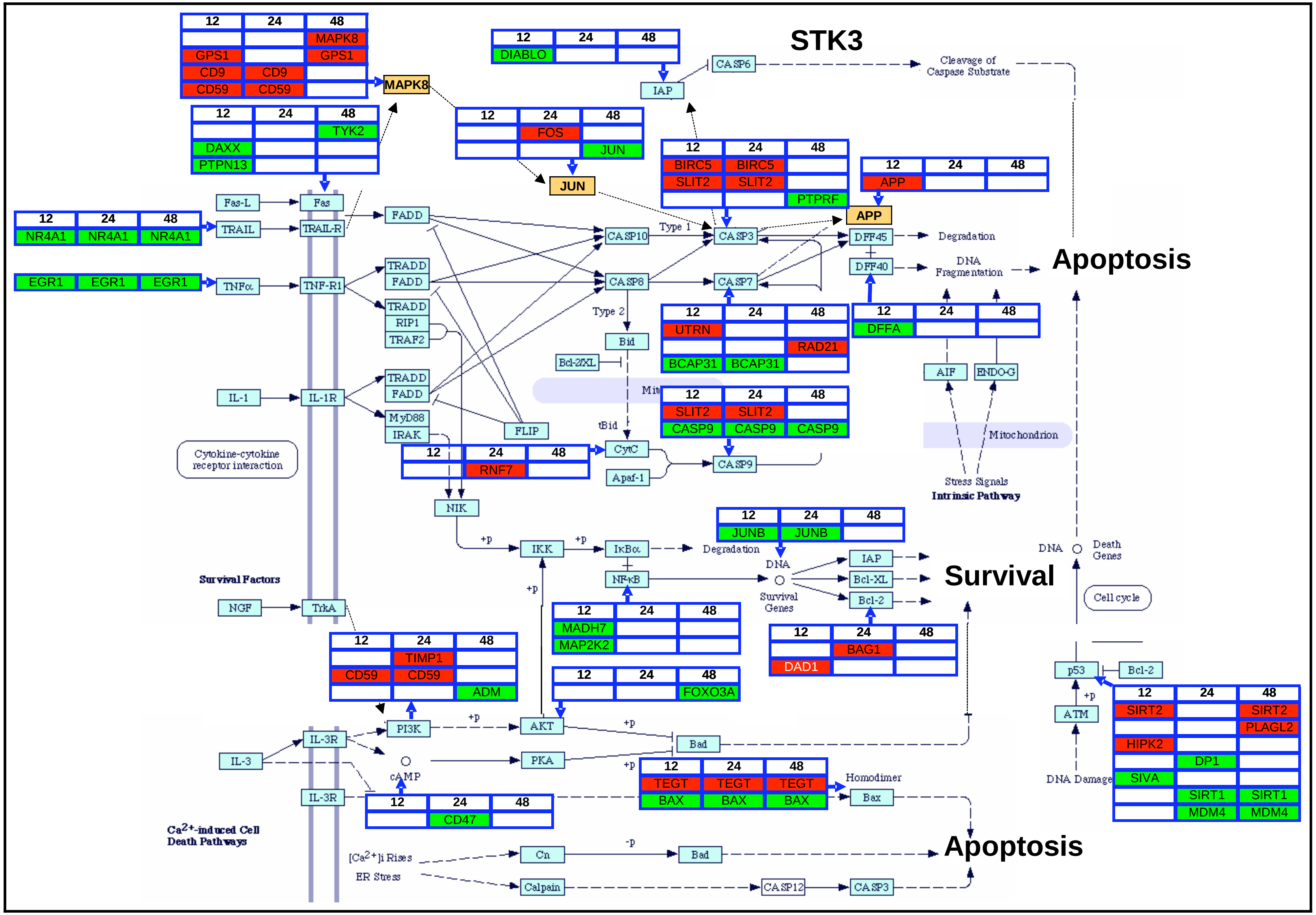
12	79	35
24	86	68
48	95	54

12	149	65
24	40	32
48	82	46

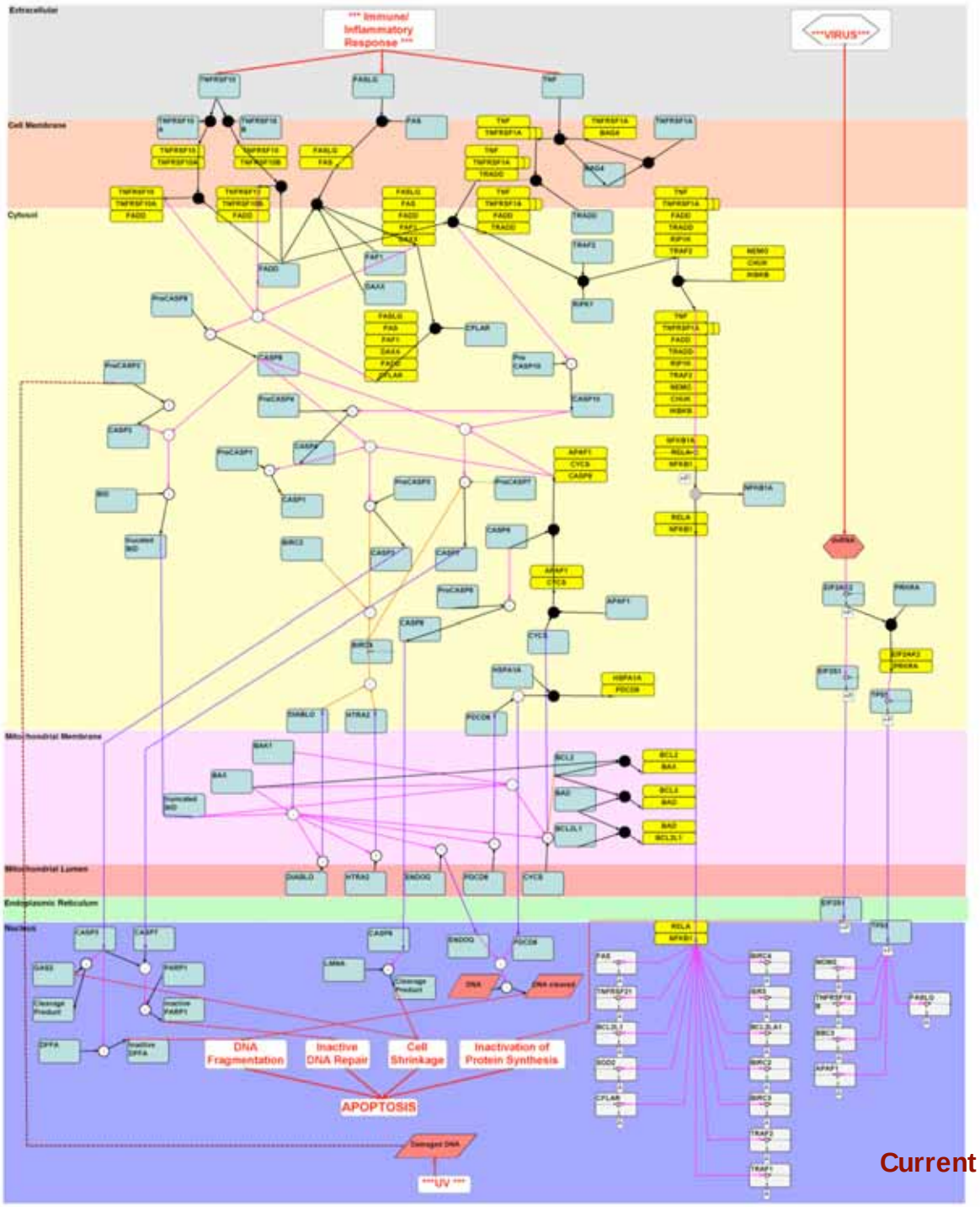


Gene Tree of Differentials





STK3
Apoptosis
Survival
Apoptosis
12 24 48
MAPK8
GPS1 GPS1
CD9 CD9
CD59 CD59
MAPK8
TYK2
DAXX
PTPN13
DIABLO
FOS
JUN
JUN
BIRC5 BIRC5
SLIT2 SLIT2
PTPRF
APP
APP
NR4A1 NR4A1 NR4A1
EGR1 EGR1 EGR1
UTRN
RAD21
BCAP31 BCAP31
DFFA
SLIT2 SLIT2
CASP9 CASP9 CASP9
RNF7
JUNB JUNB
MADH7
MAP2K2
FOXO3A
BAG1
DAD1
TIMP1
CD59 CD59
ADM
TEGT TEGT TEGT
BAX BAX BAX
CD47
SIRT2 SIRT2
PLAGL2
HIPK2
DP1
SIVA
SIRT1 SIRT1
MDM4 MDM4
Fas-L
Fas
TRAIL
TRAIL-R
FADD
TNFα
TNF-R1
TRADD
RIP1
TRAF2
IL-1
IL-1R
MyD88
IRAK
FLIP
CASP10
CASP8
CASP3
CASP7
CASP6
IAP
Cleavage of Caspase Substrate
Type 1
Type 2
Bid
Bcl-2XL
tBid
CytC
Apaf-1
CASP9
DFF45
DFF40
Degradation
DNA Fragmentation
AIF
ENDO-G
Mitochondrion
Stress Signals
Intrinsic Pathway
Cytokine-cytokine receptor interaction
NIK
IKK
IκBα
NF-κB
Degradation
DNA
Survival Genes
IAP
Bcl-XL
Bcl-2
DNA
Death Genes
Cell cycle
Survival Factors
NGF
TrkA
PI3K
AKT
Bad
PKA
IL-3
IL-3R
cAMP
Homodimer
Bax
p53
Bcl-2
ATM
DNA Damage
Ca2+-induced Cell Death Pathways
[Ca2+]i Rises
ER Stress
Cn
Bad
Calpain
CASP12
CASP3
+p
-p



Logic Mapping of Cellular Pathways

Use of standardised notion to depict biological entities and interactions, and standard layout rules

Drawn using Edinburgh Pathway Notation and Editor

Each interaction taken from literature with restriction that each must be cited in 2 papers from different laboratories

Contains:

- **61 proteins**
- **17 genes**
- **24 protein complexes**

Map currently being expanded to cover other related signalling pathways

Current Apoptosis Pathway Map

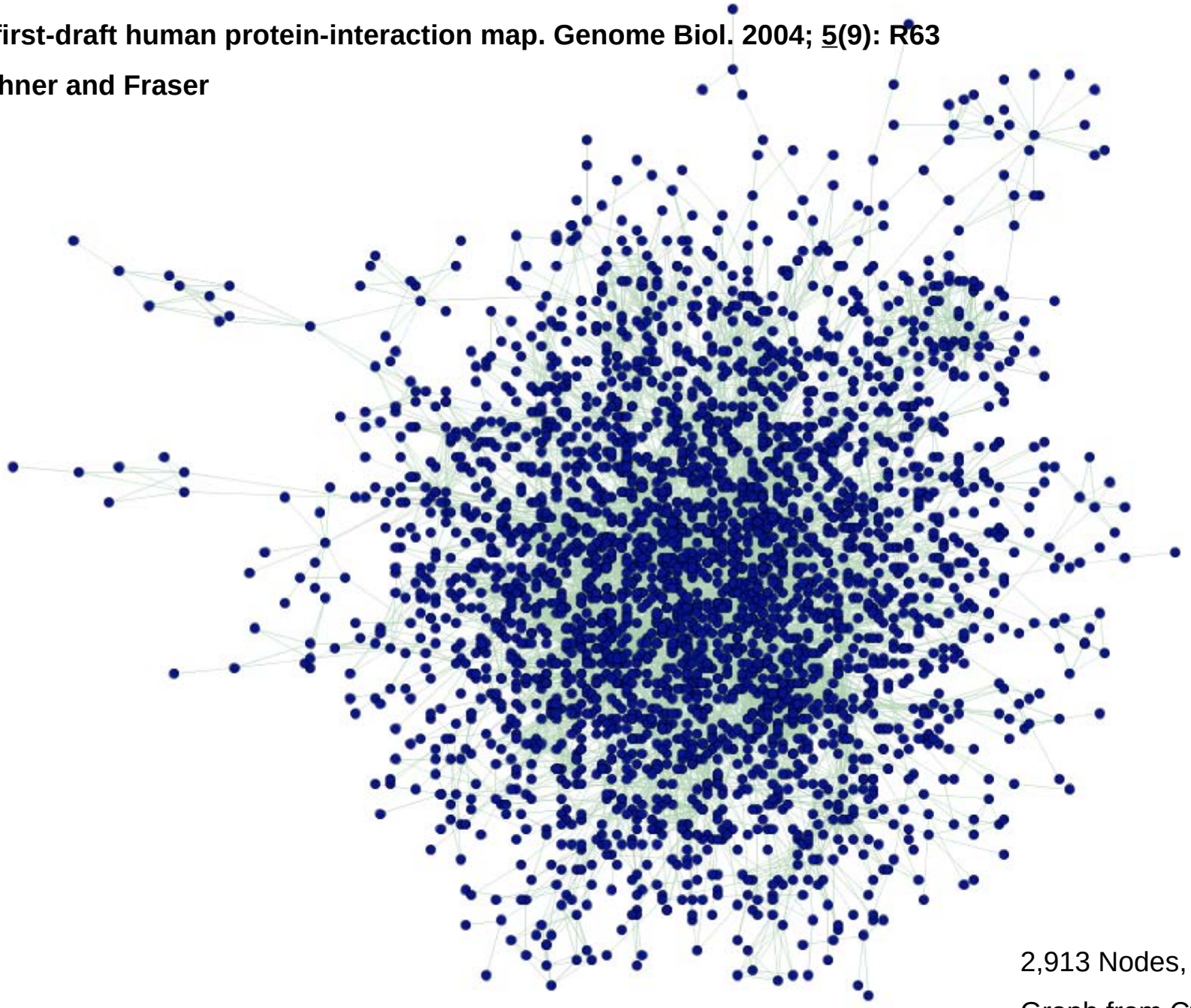
Network analysis

- Network analysis has been used to model many types of data in order to better understand complex inter-relationships (edges) between entities (nodes)

- In biology network graphs have been used in the study of sequence similarity, protein structure, evolutionary relationships, protein interactions etc.

A first-draft human protein-interaction map. Genome Biol. 2004; 5(9): R63

Lehner and Fraser



2,913 Nodes, 9,254 Edges

Graph from CytoScape

Network analysis and clustering of gene expression data

- Analysis of large expression datasets a common interest
- Data prone to noise (experimental and biological)
- Current clustering approaches are numerous, frequently based on pair-wise clustering approaches, usually slow and often ineffective at dealing with a large number of genes
- Relationship between clusters unclear
- Experimental design often non-optimal for pair-wise statistical approaches

Graph Paradigm for Gene Expression Data

- **Genes (nodes) are connected to each other in a network based on their level of co-expression (edges)**

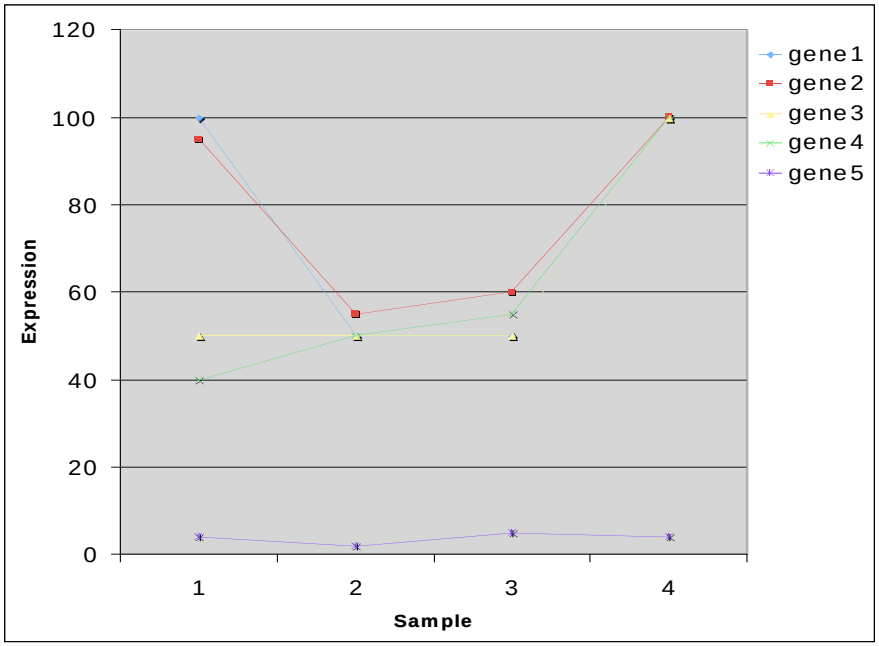
- **Co-expression measured using a correlation measure (e.g. Pearson, Spearman)**

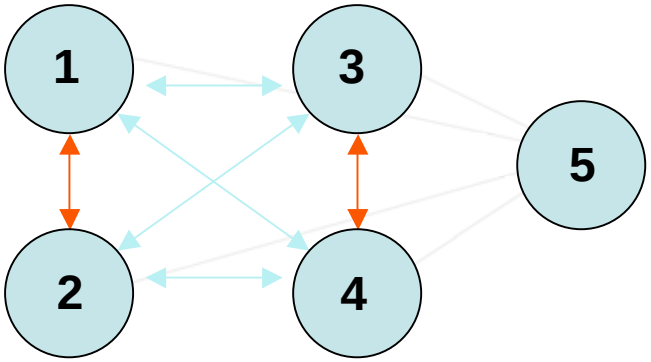
- **Development of program BioLayoutExpress:**

✓ ***data import***
✓ ***Pearson calculations***
✓ ***3D graph layout***
✓ ***link to Markov CLustering (MCL) algorithm***
✓ ***expression and annotation viewers***
✓ ***annotation statistical mining***

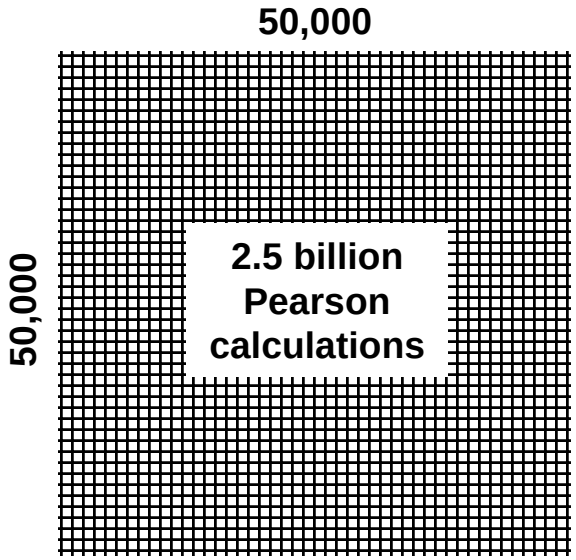
Sanger, Team 101
Anton Enright
Stijn van Dongen
Russell Grocock
Markus Brosch
EBI
Leon Goldovsky

	Tissue1	Tissue2	Tissue3	Tissue4
Gene1	100	50	50	100
Gene2	95	55	60	100
Gene3	50	50	50	100
Gene4	40	50	55	100
Gene5	4	2	5	4





	Gene1	Gene2	Gene3	Gene4	Gene5
Gene1	100%	99%	58%	38%	23%
Gene2	99%	100%	64%	46%	31%
Gene3	58%	64%	100%	97%	13%
Gene4	38%	46%	97%	100%	16%
Gene5	23%	31%	13%	16%	100%



Network Analysis of hCMV Infection of Human Macrophages

Experimental design and workup of samples by Christian Sinzger, Tübingen, Germany

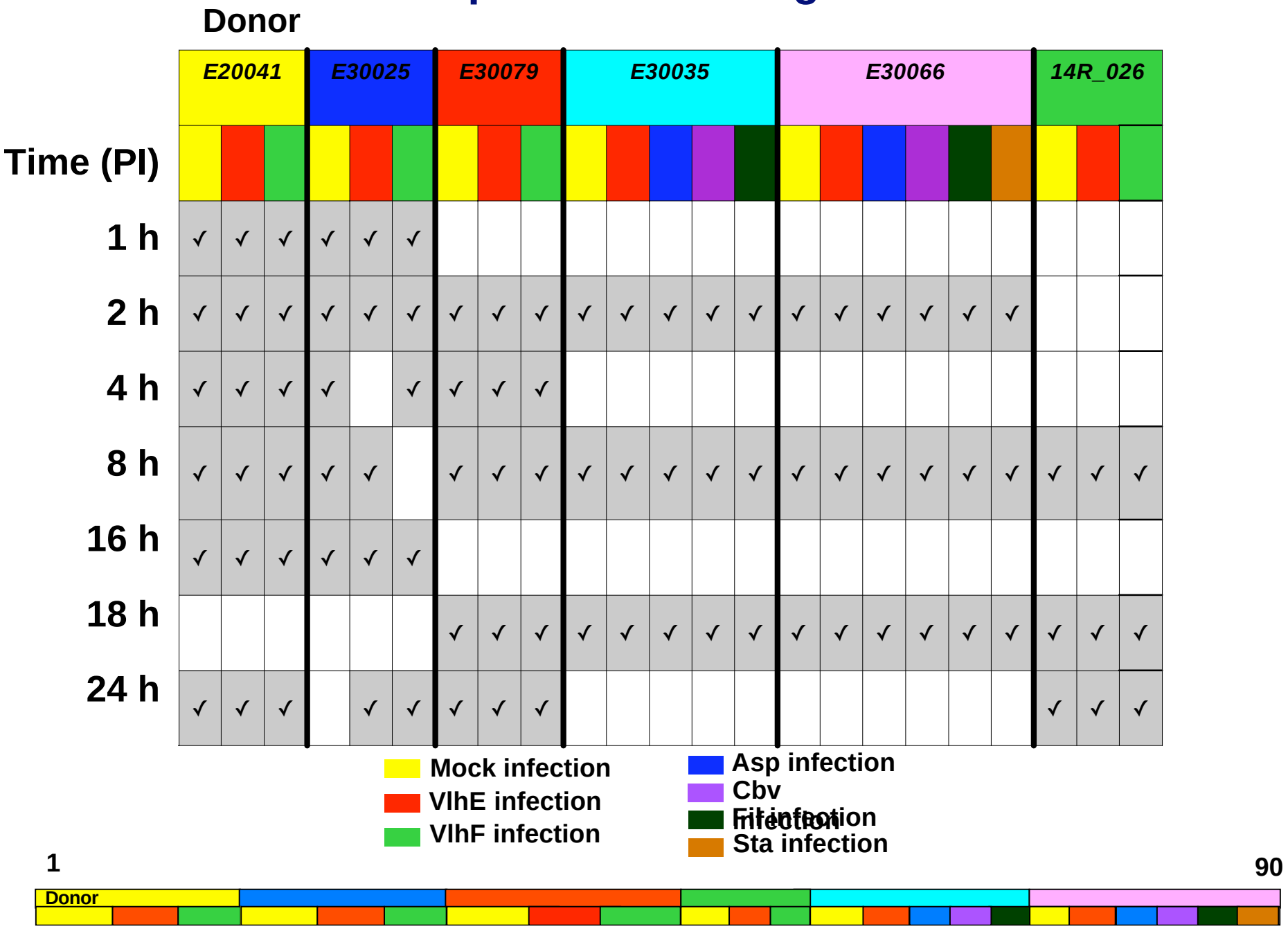
Questions addressed:

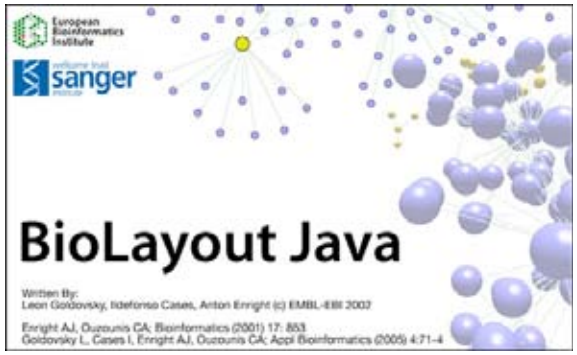
- **What is the transcriptional response of human macrophages to infection by hCMV?**
- **What is the difference in response to a productive (VlhE) and non-productive (VlhF) infection?**
- **How specific is the response to viral infection?**

Analysis used 90 Affymetrix U133A chips

Data QC and statistical analysis by Thorsten Forster

Experimental design





Expression data
(normalised and annotated)

↓

Gene to gene
Pearson correlation calculated
for every probe set on the array

↓

Pearson correlations >0.7 saved

↓

Pearson correlation file >0.7
filtered based on user defined threshold (0.7-1.0)

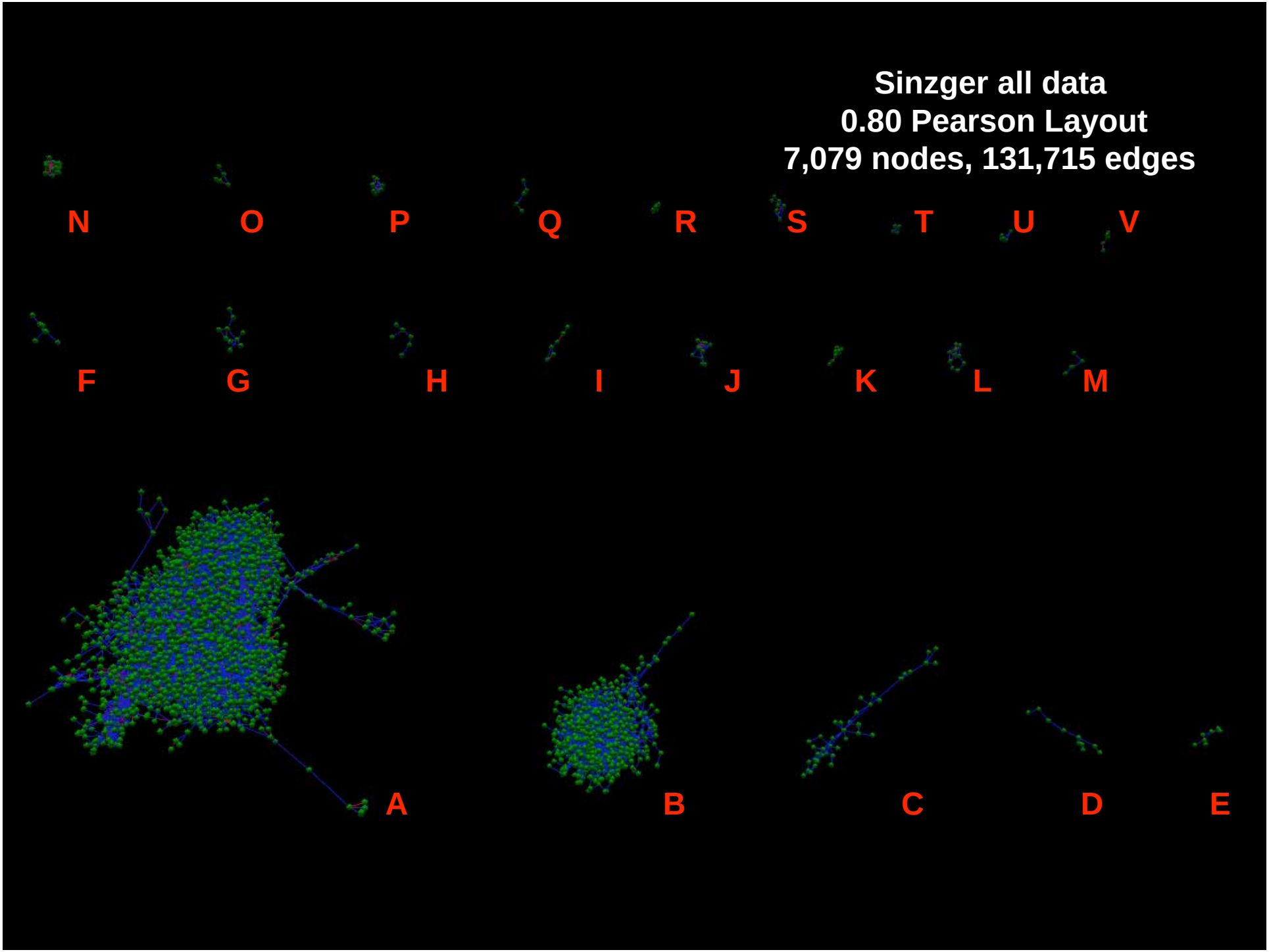
↓

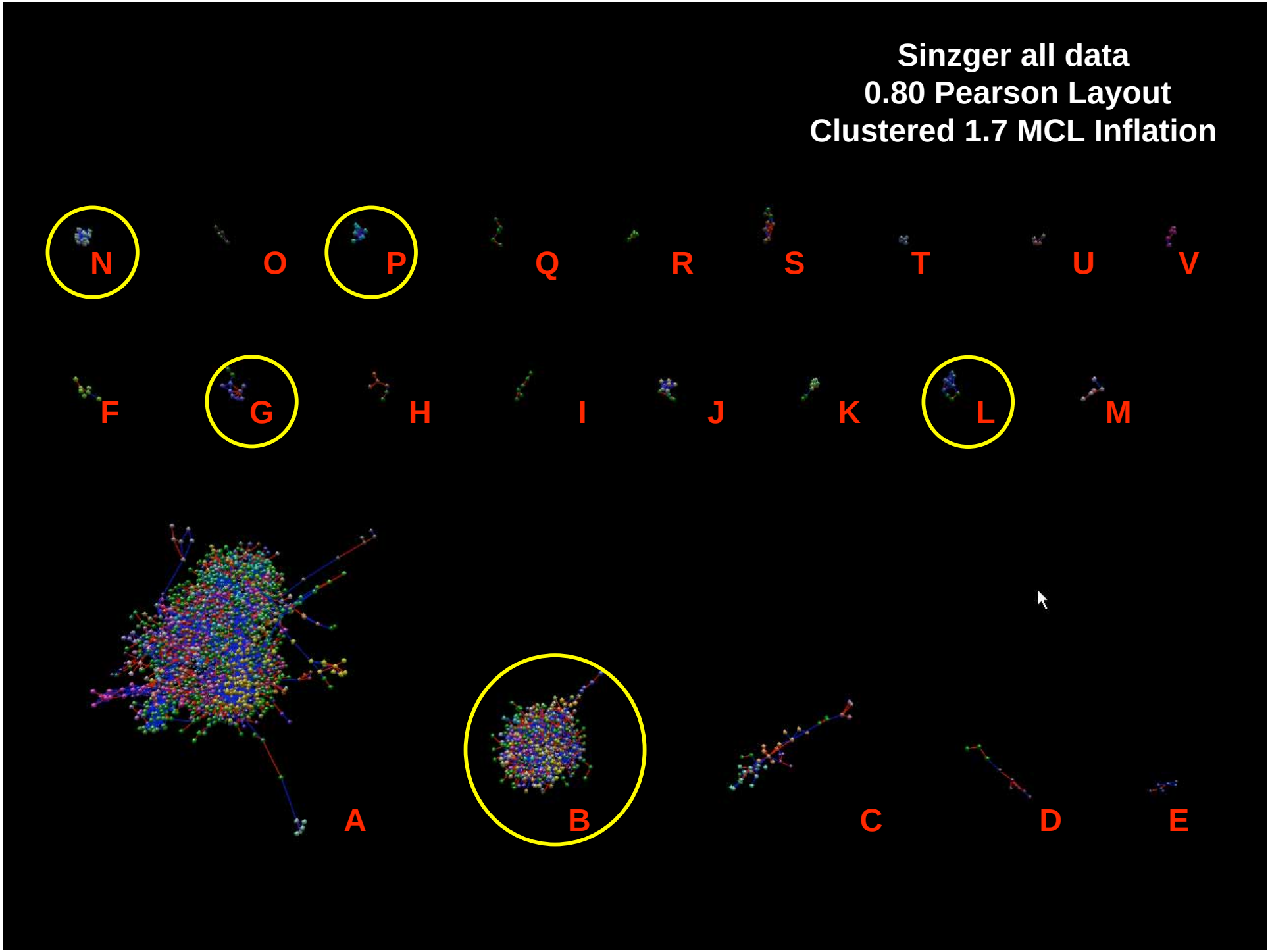
Edges drawn between nodes (genes) based on
correlations > than selected threshold

↓

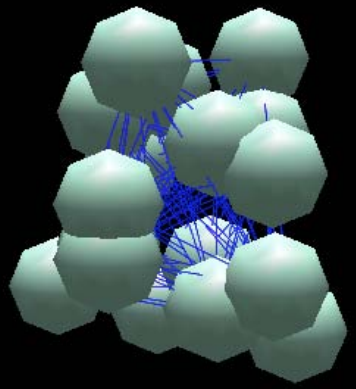
Network graphs laid
out in tiled arrangement and clustered

Singletons and graphs with
<4 members removed

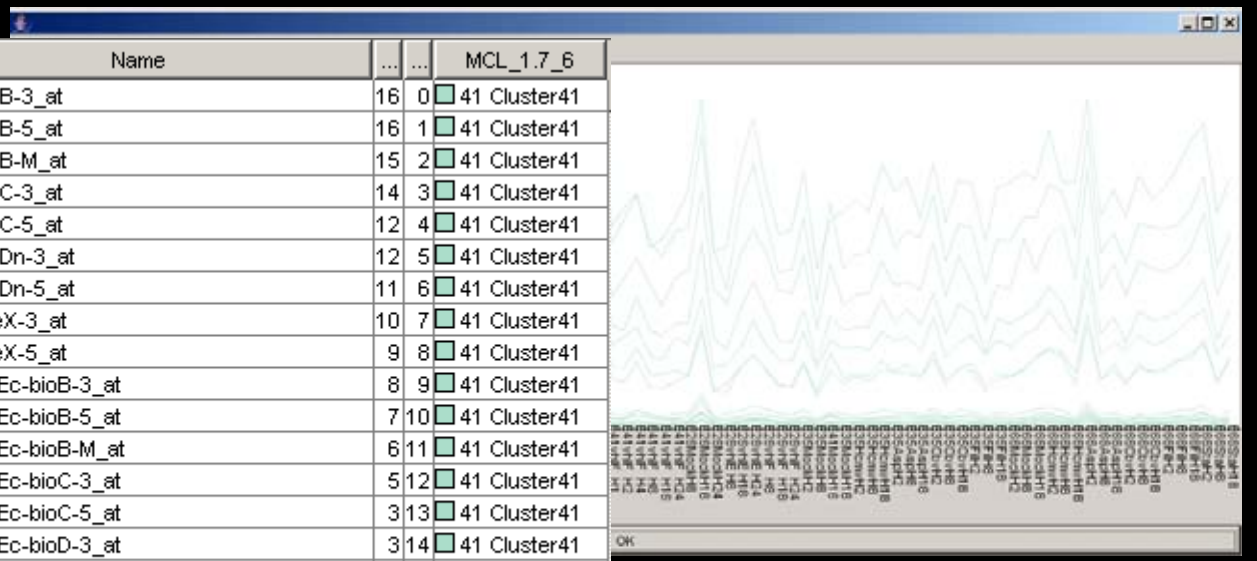

Sinzger all data
0.80 Pearson Layout
7,079 nodes, 131,715 edges
N
O
P
Q
R
S
T
U
V
F
G
H
I
J
K
L
M
A
B
C
D
E


Sinzger all data
0.80 Pearson Layout
Clustered 1.7 MCL Inflation
N
O
P
Q
R
S
T
U
V
F
G
H
I
J
K
L
M
A
B
C
D
E

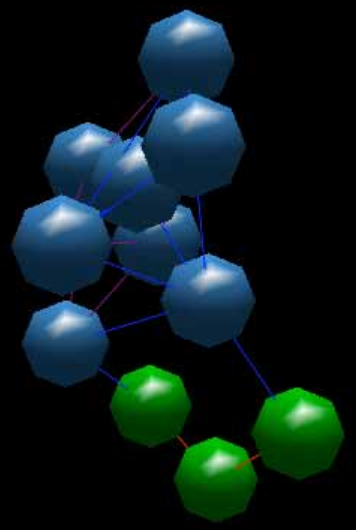
N

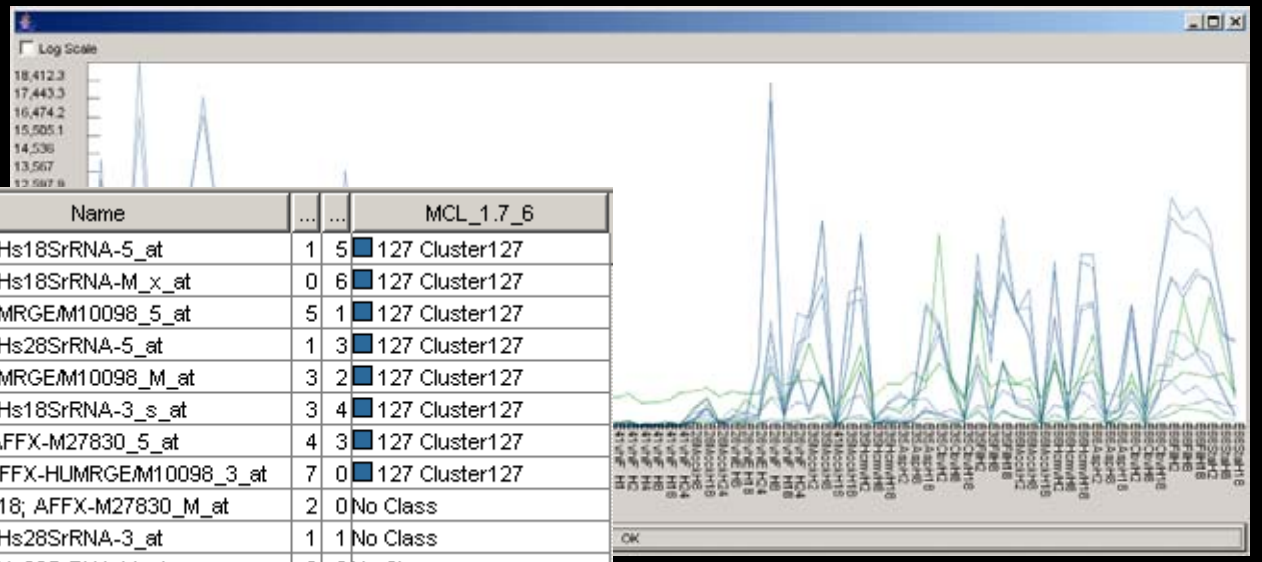


Name	...	...	MCL_1.7_6
AFFX-BioB-3_at	16	0	41 Cluster41
AFFX-BioB-5_at	16	1	41 Cluster41
AFFX-BioB-M_at	15	2	41 Cluster41
AFFX-BioC-3_at	14	3	41 Cluster41
AFFX-BioC-5_at	12	4	41 Cluster41
AFFX-BioDn-3_at	12	5	41 Cluster41
AFFX-BioDn-5_at	11	6	41 Cluster41
AFFX-CreX-3_at	10	7	41 Cluster41
AFFX-CreX-5_at	9	8	41 Cluster41
AFFX-r2-Ec-bioB-3_at	8	9	41 Cluster41
AFFX-r2-Ec-bioB-5_at	7	10	41 Cluster41
AFFX-r2-Ec-bioB-M_at	6	11	41 Cluster41
AFFX-r2-Ec-bioC-3_at	5	12	41 Cluster41
AFFX-r2-Ec-bioC-5_at	3	13	41 Cluster41
AFFX-r2-Ec-bioD-3_at	3	14	41 Cluster41
AFFX-r2-Ec-bioD-5_at	2	15	41 Cluster41
AFFX-r2-P1-cre-3_at	1	14	41 Cluster41
AFFX-r2-P1-cre-5_at	0	16	41 Cluster41



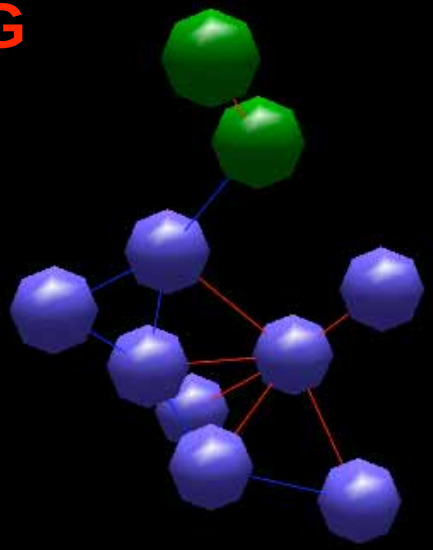
L

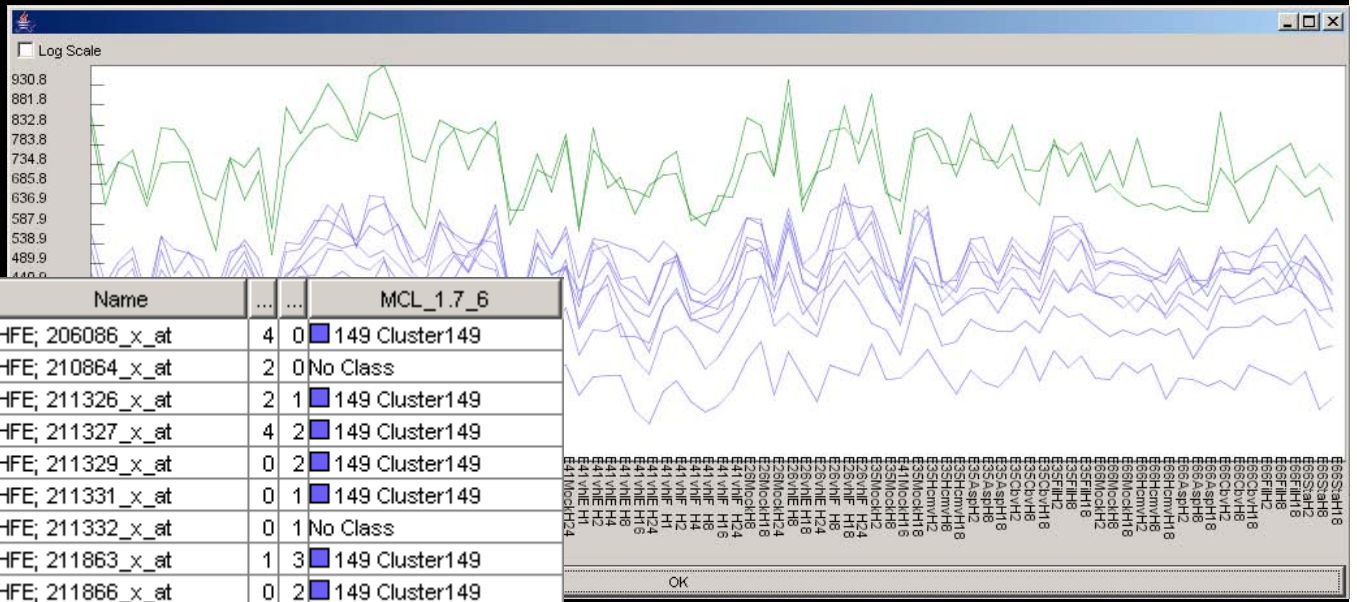




Name	...	...	MCL_1.7_6
AFFX-r2-Hs18SrRNA-5_at	1	5	127 Cluster127
AFFX-r2-Hs18SrRNA-M_x_at	0	6	127 Cluster127
AFFX-HUMRGE/M10098_5_at	5	1	127 Cluster127
AFFX-r2-Hs28SrRNA-5_at	1	3	127 Cluster127
AFFX-HUMRGE/M10098_M_at	3	2	127 Cluster127
AFFX-r2-Hs18SrRNA-3_s_at	3	4	127 Cluster127
SOX18; AFFX-M27830_5_at	4	3	127 Cluster127
SRP68; AFFX-HUMRGE/M10098_3_at	7	0	127 Cluster127
LOC440118; AFFX-M27830_M_at	2	0	No Class
AFFX-r2-Hs28SrRNA-3_at	1	1	No Class
AFFX-r2-Hs28SrRNA-M_at	0	2	No Class

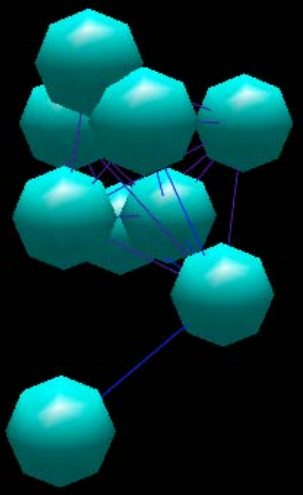
G

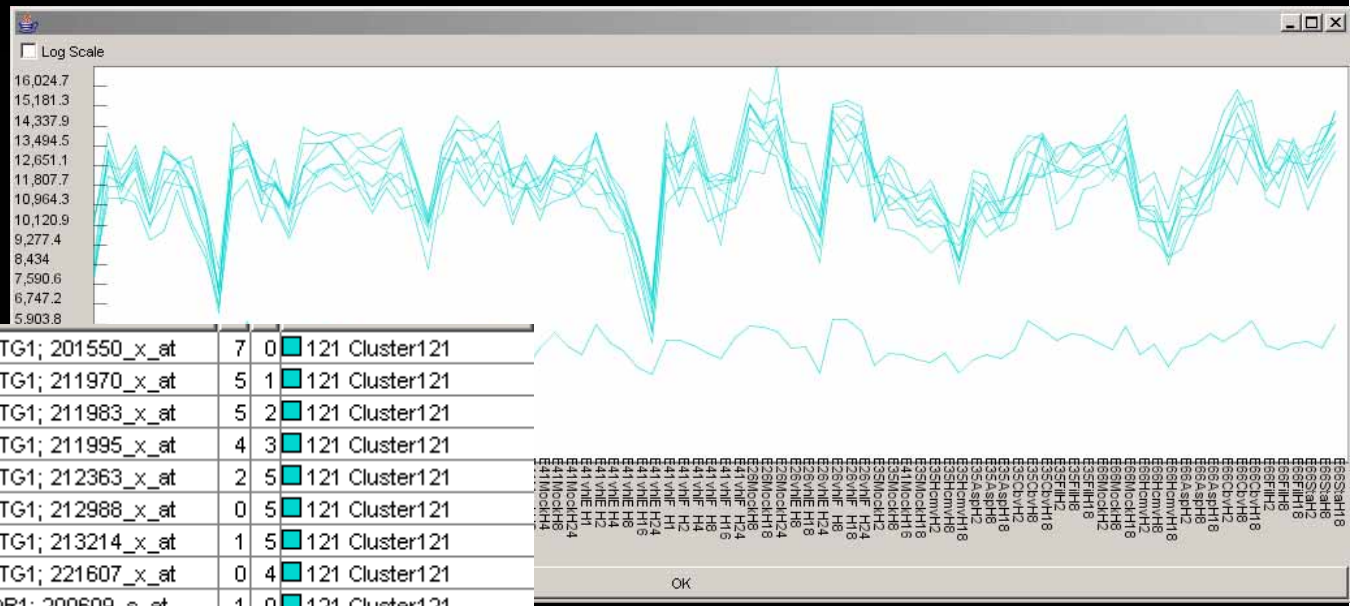




Name	...	...	MCL_1.7_6
HFE; 206086_x_at	4	0	149 Cluster149
HFE; 210864_x_at	2	0	No Class
HFE; 211326_x_at	2	1	149 Cluster149
HFE; 211327_x_at	4	2	149 Cluster149
HFE; 211329_x_at	0	2	149 Cluster149
HFE; 211331_x_at	0	1	149 Cluster149
HFE; 211332_x_at	0	1	No Class
HFE; 211863_x_at	1	3	149 Cluster149
HFE; 211866_x_at	0	2	149 Cluster149
HFE; 214647_s_at	0	1	149 Cluster149

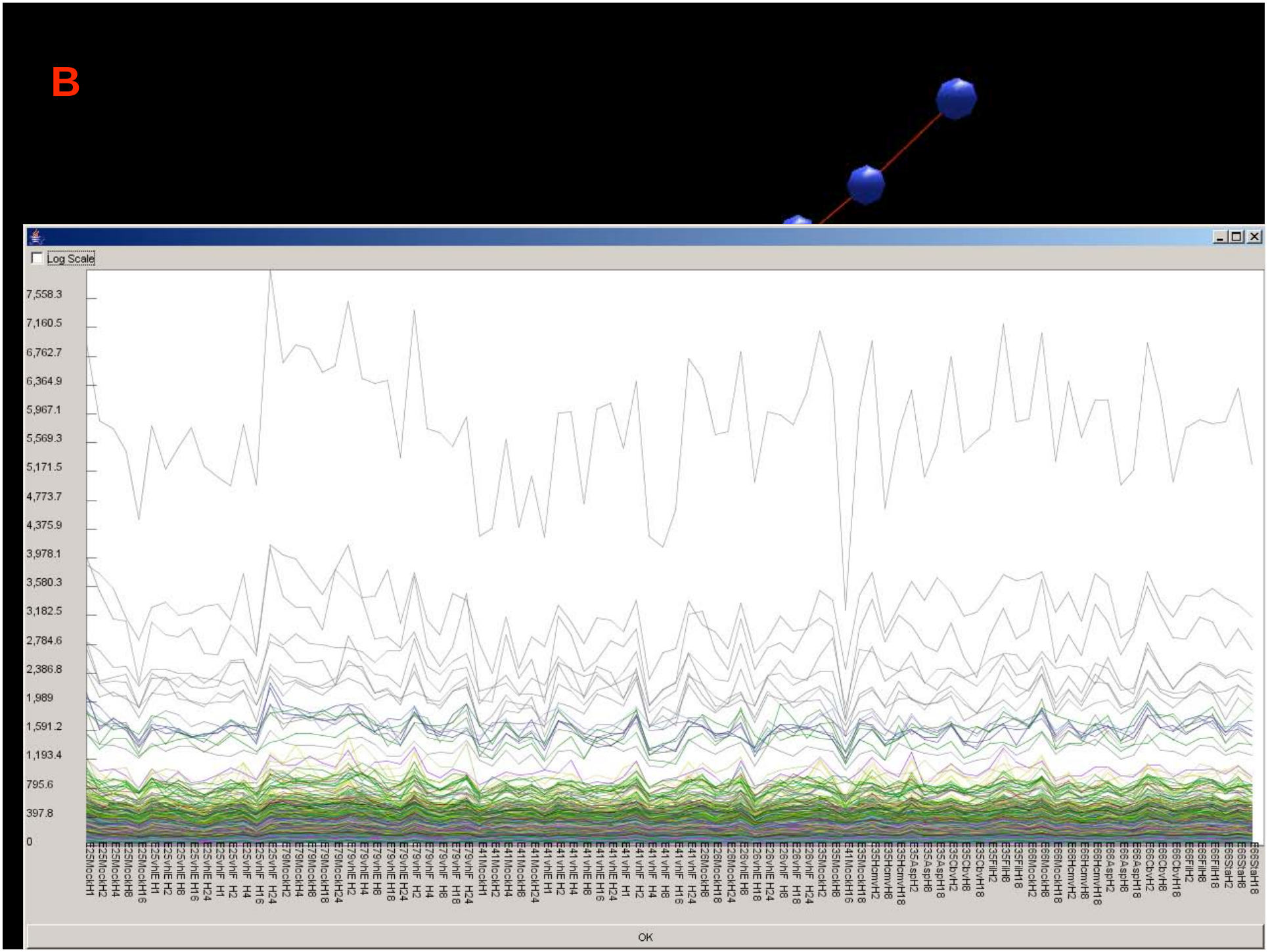
P

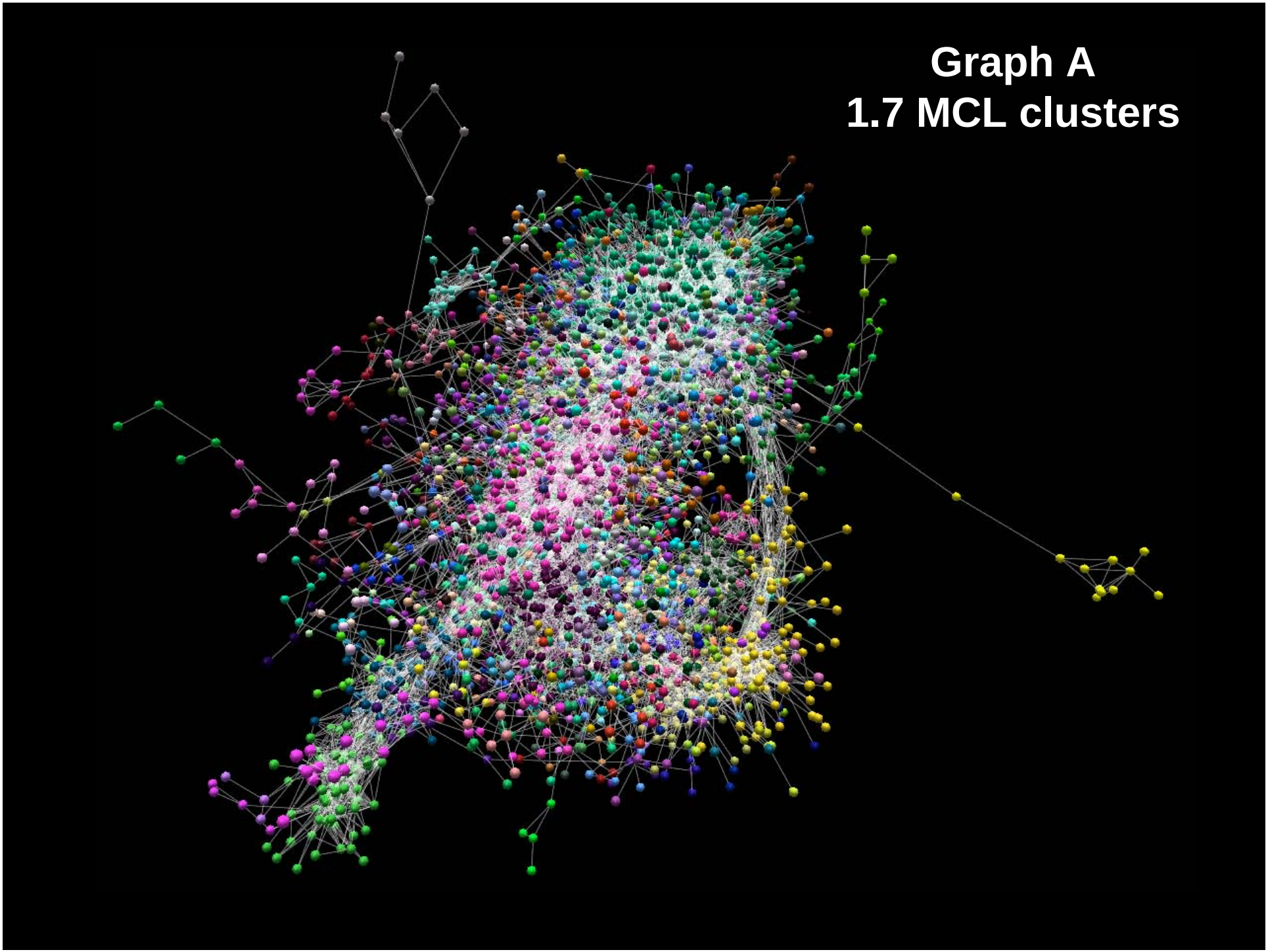


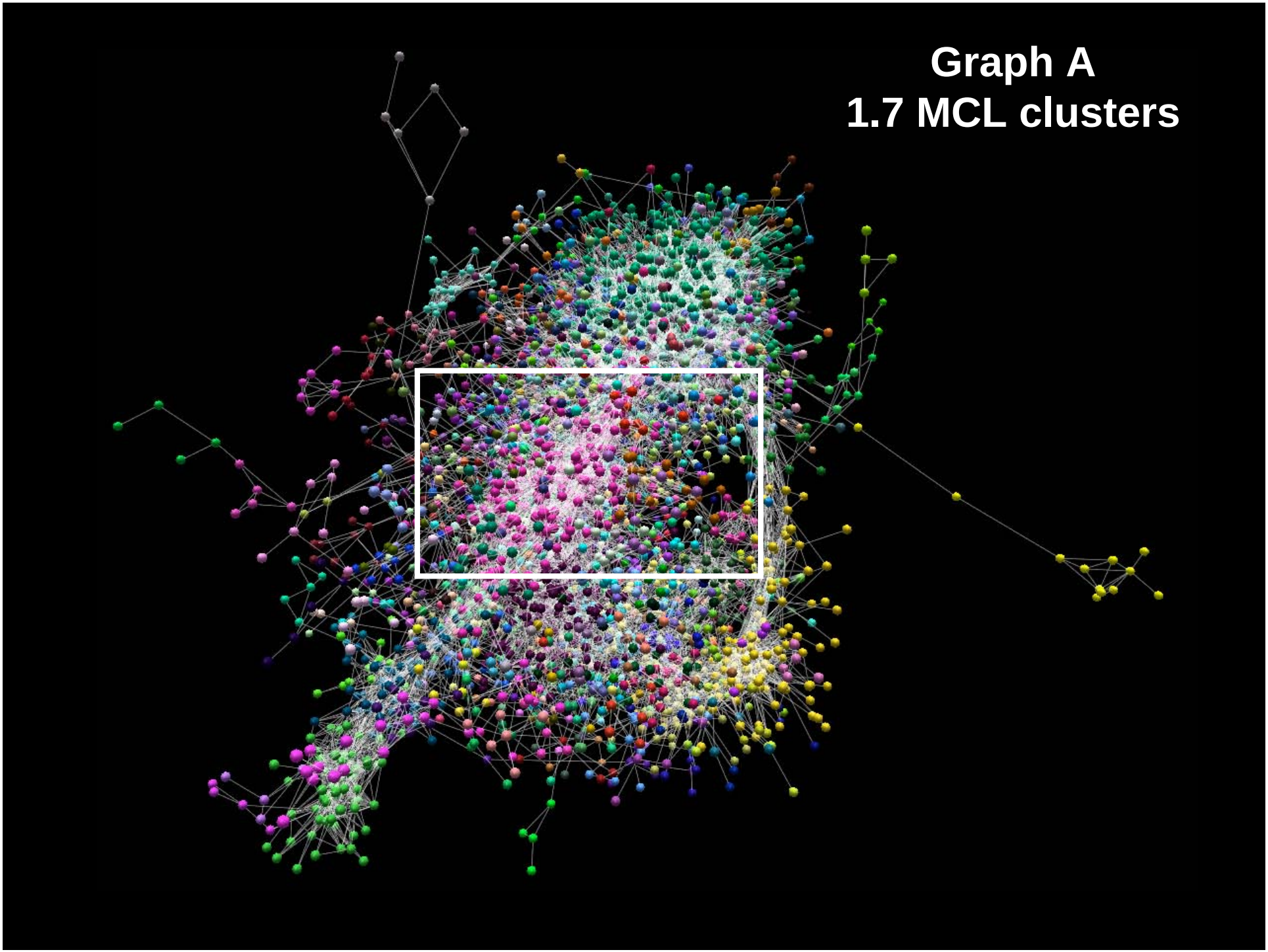


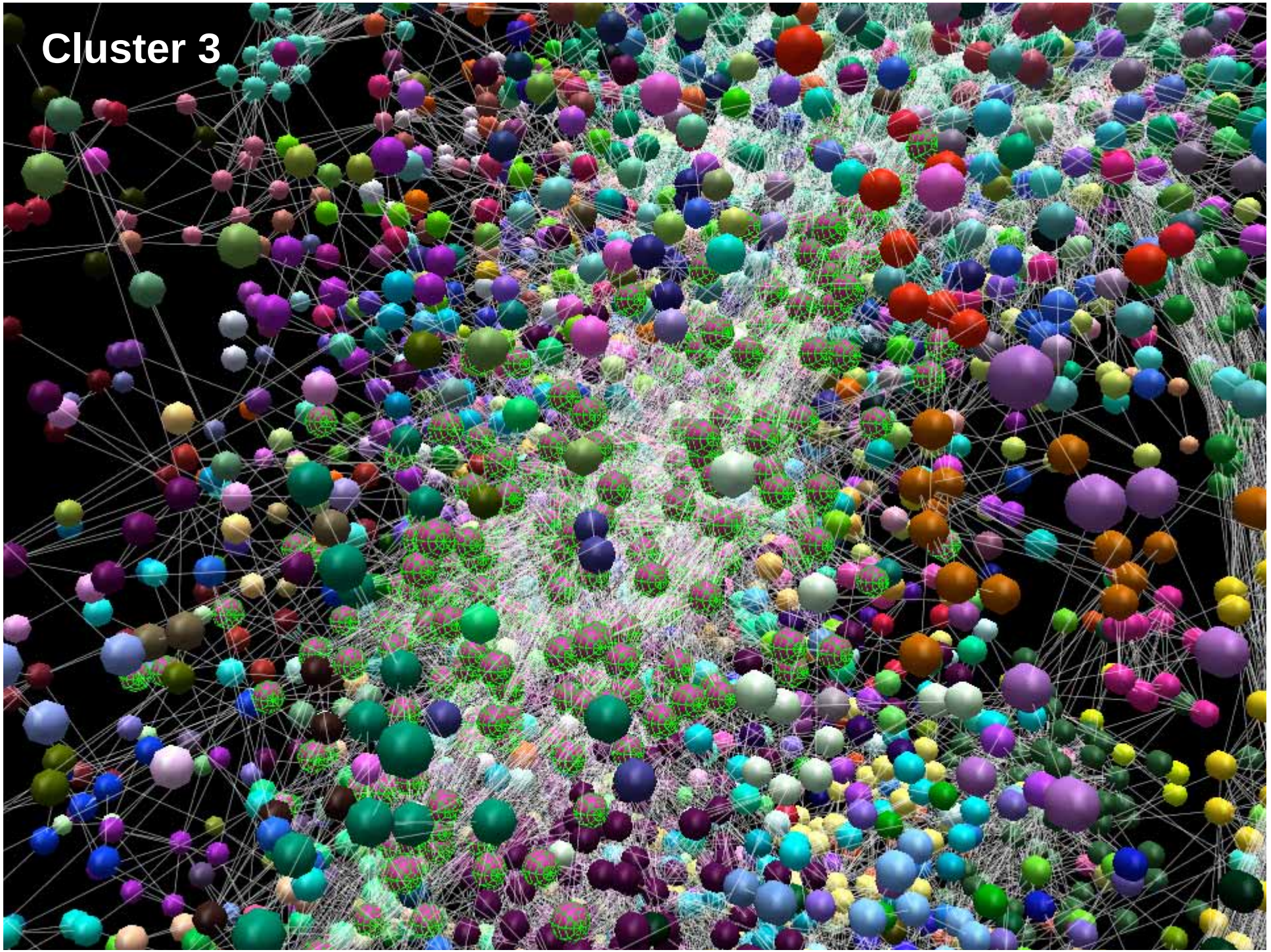
ACTG1; 201550_x_at	7	0	121 Cluster121
ACTG1; 211970_x_at	5	1	121 Cluster121
ACTG1; 211983_x_at	5	2	121 Cluster121
ACTG1; 211995_x_at	4	3	121 Cluster121
ACTG1; 212363_x_at	2	5	121 Cluster121
ACTG1; 212988_x_at	0	5	121 Cluster121
ACTG1; 213214_x_at	1	5	121 Cluster121
ACTG1; 221607_x_at	0	4	121 Cluster121
WDR1; 200609_s_at	1	0	121 Cluster121

B

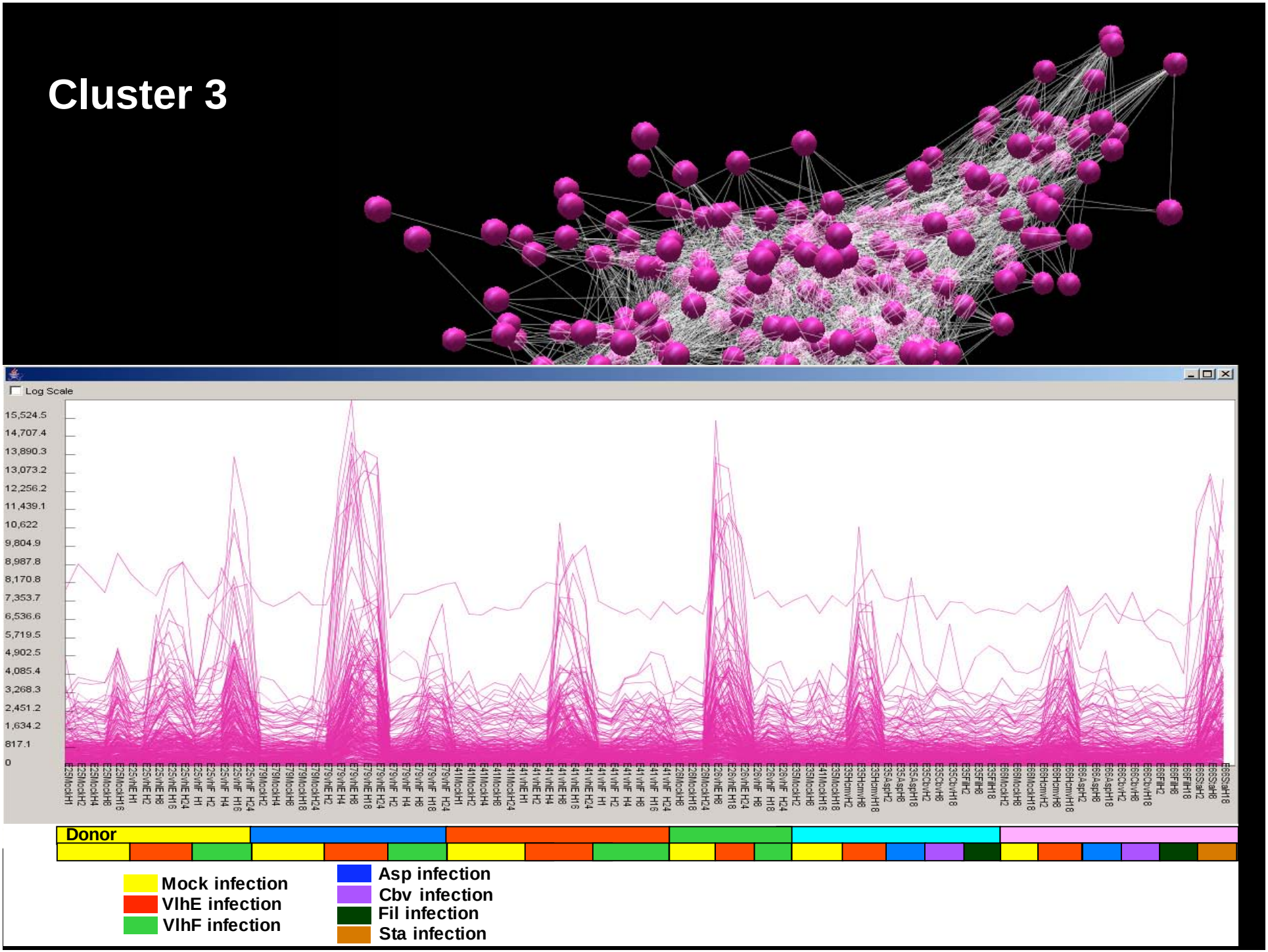

Log Scale
7,558.3
7,160.5
6,762.7
6,364.9
5,967.1
5,569.3
5,171.5
4,773.7
4,375.9
3,978.1
3,580.3
3,182.5
2,784.6
2,386.8
1,989
1,591.2
1,193.4
795.6
397.8
0
E25MockH1
E25MockH2
E25MockH4
E25MockH8
E25MockH16
E25vhE H1
E25vhE H2
E25vhE H8
E25vhE H16
E25vhE H24
E25vhF H1
E25vhF H2
E25vhF H4
E25vhF H16
E25vhF H24
E79MockH2
E79MockH4
E79MockH8
E79MockH18
E79MockH24
E79vhE H2
E79vhE H4
E79vhE H8
E79vhE H16
E79vhE H24
E79vhF H2
E79vhF H4
E79vhF H8
E79vhF H18
E79vhF H24
E41MockH1
E41MockH2
E41MockH4
E41MockH8
E41MockH24
E41vhE H1
E41vhE H2
E41vhE H4
E41vhE H8
E41vhE H16
E41vhE H24
E41vhF H1
E41vhF H2
E41vhF H4
E41vhF H8
E41vhF H16
E41vhF H24
E26MockH8
E26MockH18
E26MockH24
E26vhE H8
E26vhE H18
E26vhE H24
E26vhF H8
E26vhF H18
E26vhF H24
E35MockH2
E35MockH8
E35MockH16
E35MockH18
E35HcmvH2
E35HcmvH8
E35HcmvH18
E35AspH2
E35AspH8
E35AspH18
E35CbvH8
E35CbvH18
E35FlH2
E35FlH8
E35FlH18
E66MockH2
E66MockH8
E66MockH18
E66HcmvH2
E66HcmvH8
E66HcmvH18
E66AspH2
E66AspH8
E66AspH18
E66CbvH2
E66CbvH8
E66CbvH18
E66FlH2
E66FlH8
E66FlH18
E66StaH2
E66StaH8
E66StaH18
OK


Graph A
1.7 MCL clusters

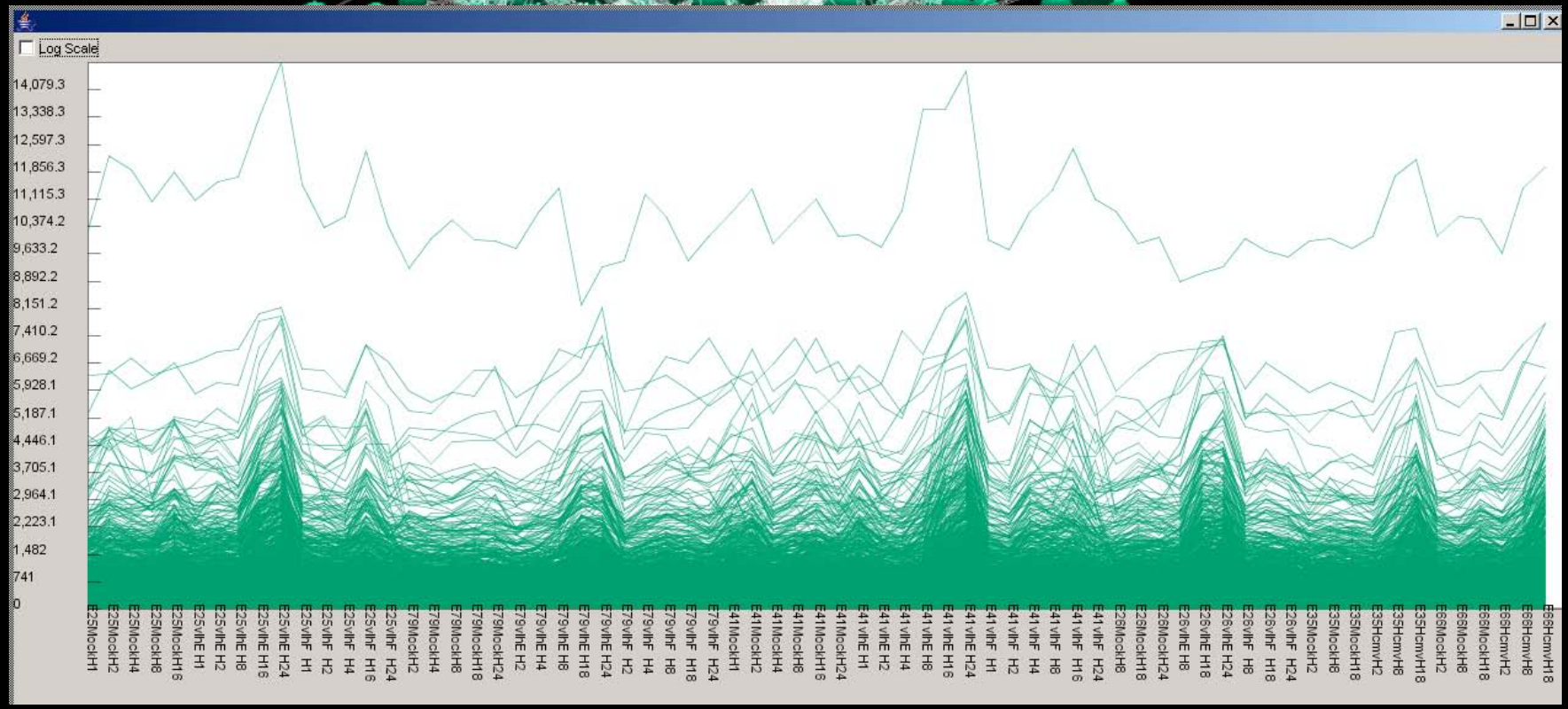

Graph A
1.7 MCL clusters


Cluster 3

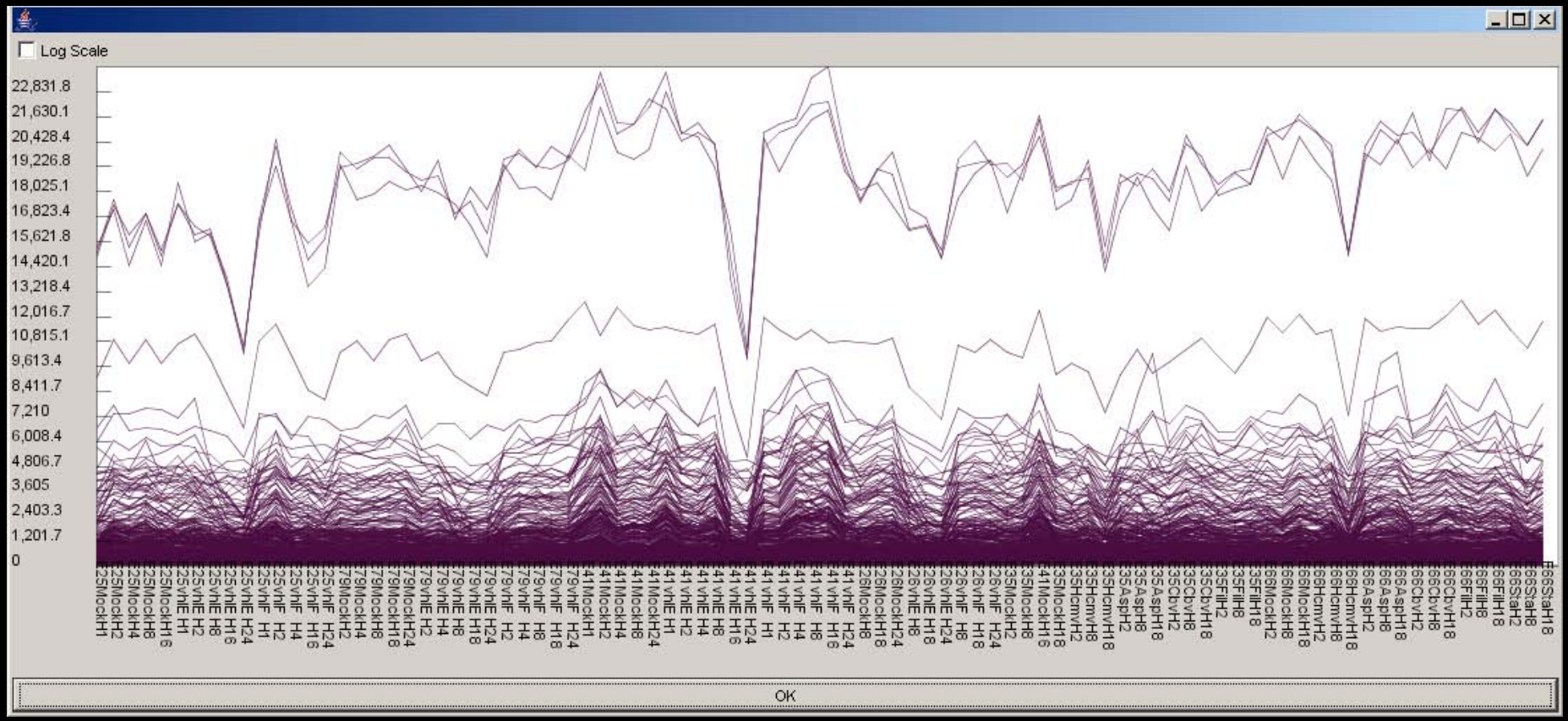
Cluster 3

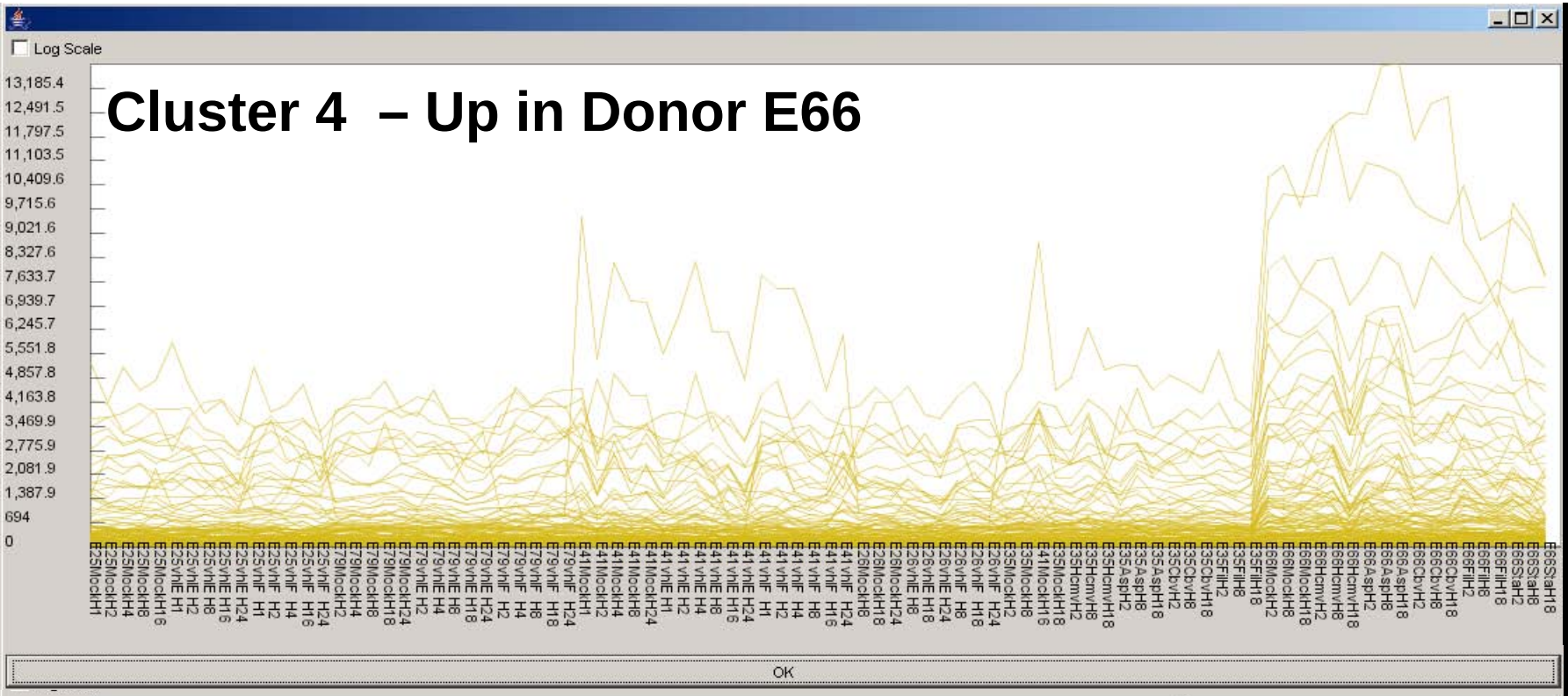

Log Scale
15,524.5
14,707.4
13,890.3
13,073.2
12,256.2
11,439.1
10,622
9,804.9
8,987.8
8,170.8
7,353.7
6,536.6
5,719.5
4,902.5
4,085.4
3,268.3
2,451.2
1,634.2
817.1
0
Donor
Mock infection
VIhE infection
VIhF infection
Asp infection
Cbv infection
Fil infection
Sta infection

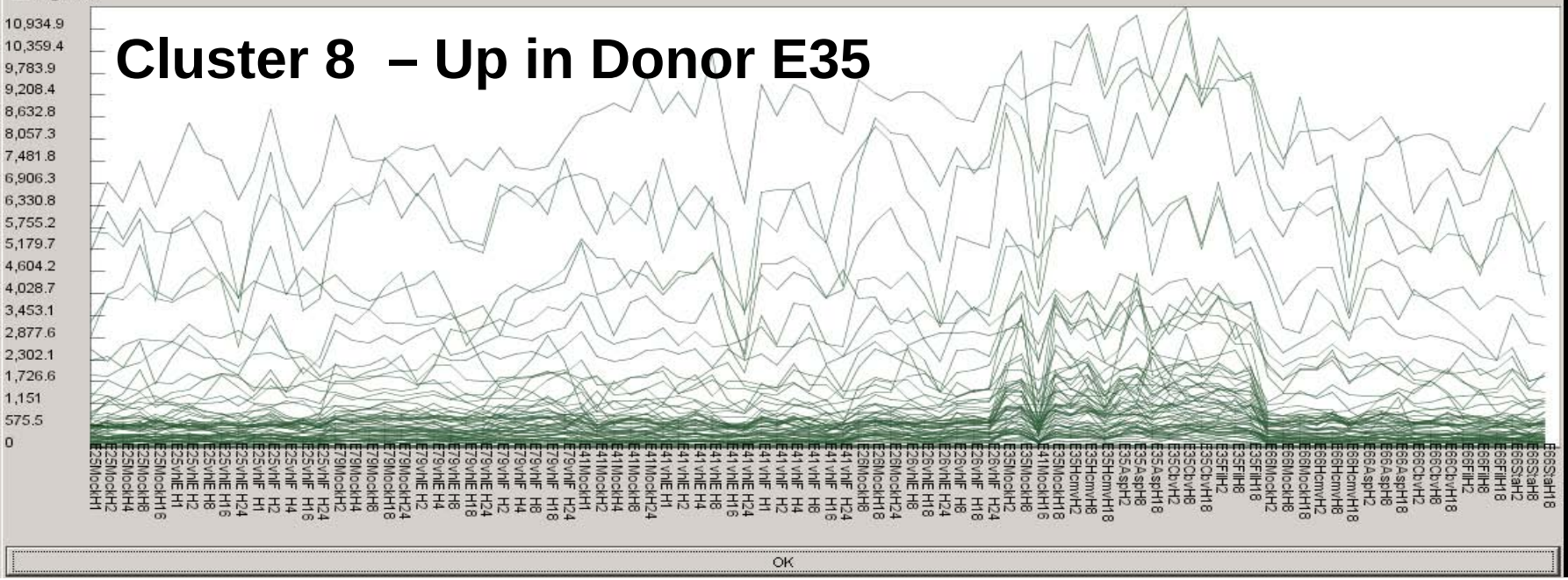
Cluster 1


Log Scale
14,079.3
13,338.3
12,597.3
11,856.3
11,115.3
10,374.2
9,633.2
8,892.2
8,151.2
7,410.2
6,669.2
5,928.1
5,187.1
4,446.1
3,705.1
2,964.1
2,223.1
1,482
741
0
E25MockH1
E25MockH2
E25MockH4
E25MockH8
E25MockH16
E25vlhE H1
E25vlhE H2
E25vlhE H8
E25vlhE H16
E25vlhE H24
E25vlhF H1
E25vlhF H2
E25vlhF H4
E25vlhF H16
E25vlhF H24
E79MockH2
E79MockH4
E79MockH8
E79MockH18
E79MockH24
E79vlhE H2
E79vlhE H4
E79vlhE H8
E79vlhE H18
E79vlhE H24
E79vlhF H2
E79vlhF H4
E79vlhF H8
E79vlhF H18
E79vlhF H24
E41MockH1
E41MockH2
E41MockH4
E41MockH8
E41MockH16
E41MockH24
E41vlhE H1
E41vlhE H2
E41vlhE H4
E41vlhE H8
E41vlhE H16
E41vlhE H24
E41vlhF H1
E41vlhF H2
E41vlhF H4
E41vlhF H8
E41vlhF H16
E41vlhF H24
E28MockH8
E28MockH18
E28MockH24
E28vlhE H8
E28vlhE H18
E28vlhE H24
E28vlhF H8
E28vlhF H18
E28vlhF H24
E35MockH2
E35MockH8
E35MockH18
E35HcmvH2
E35HcmvH8
E35HcmvH18
E66MockH2
E66MockH8
E66MockH18
E66HcmvH2
E66HcmvH8
E66HcmvH18

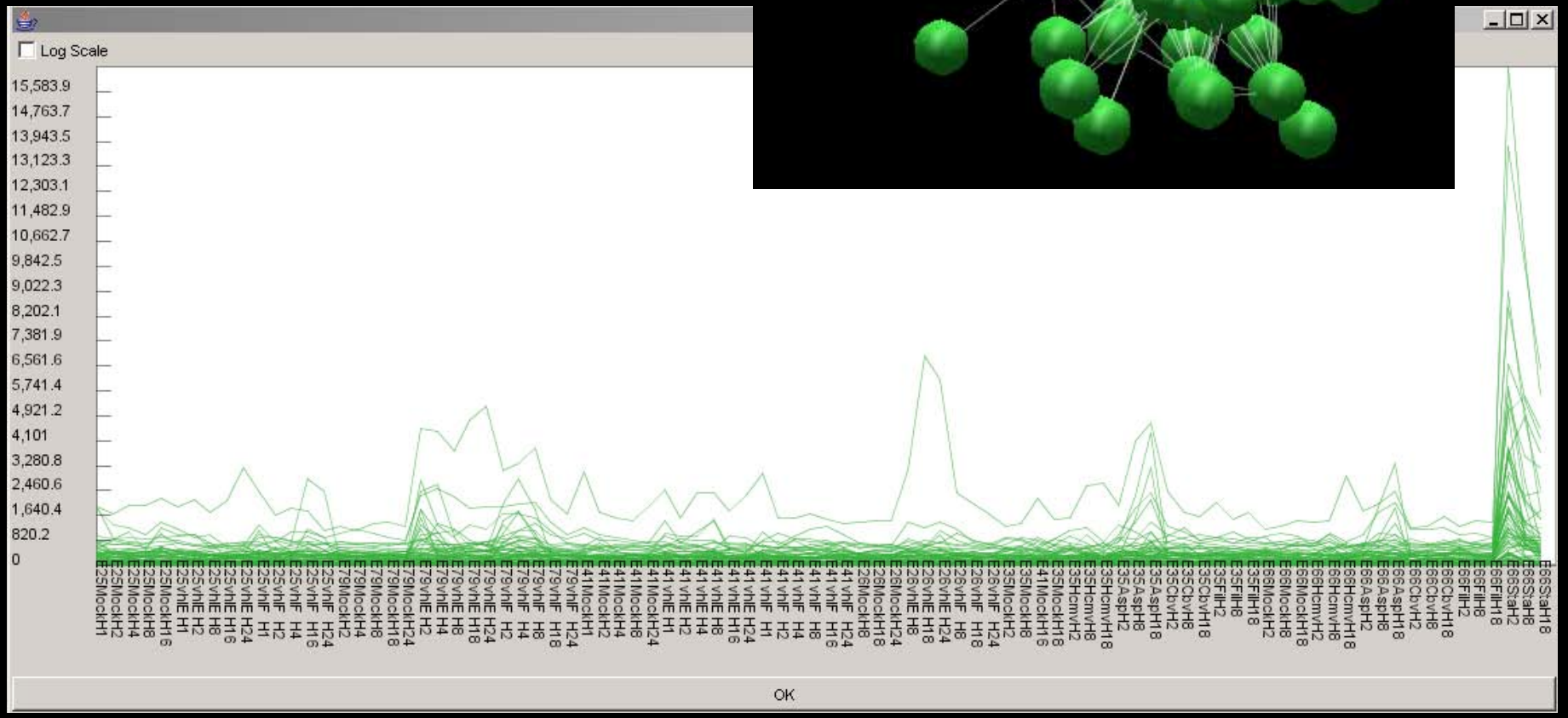
Cluster 2 – Down in VlhE (VlhF)



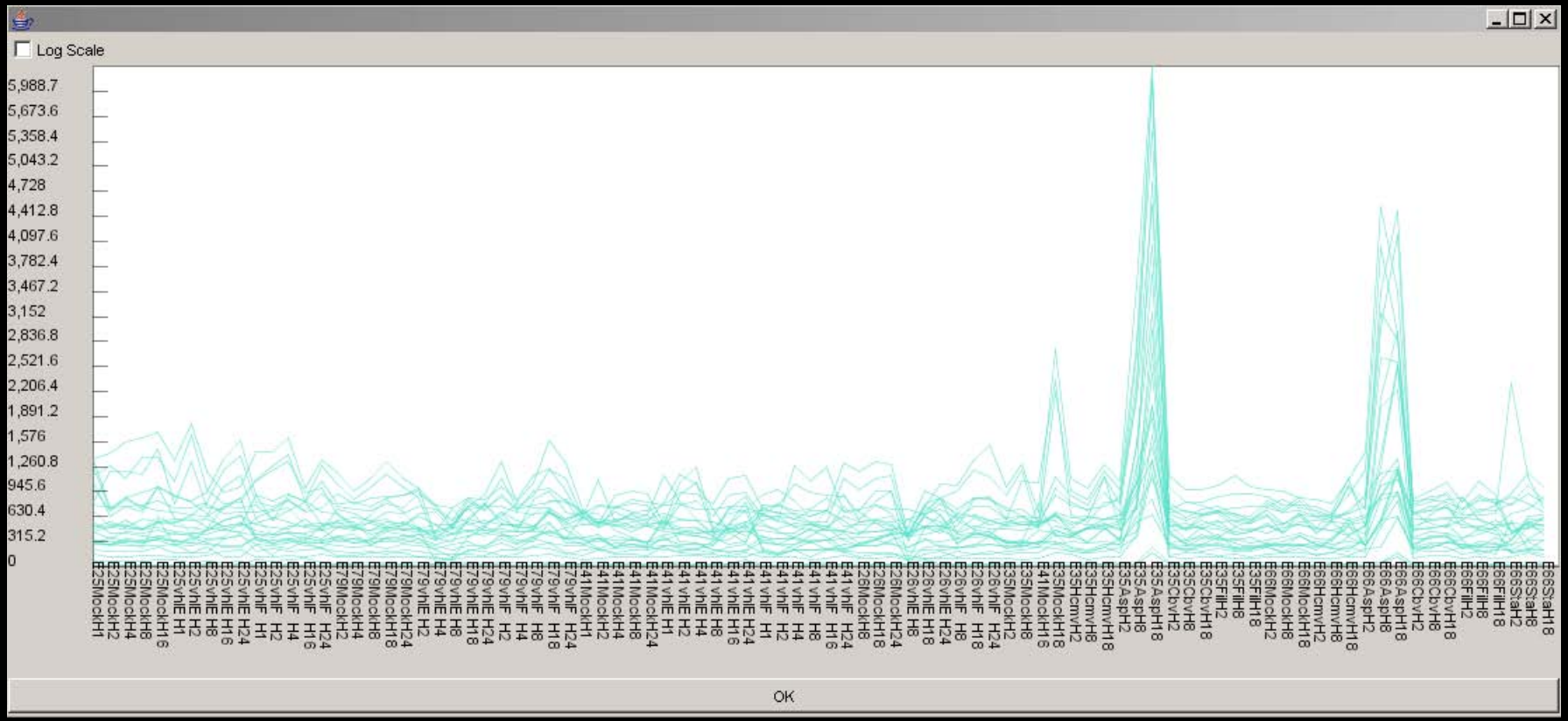

Log Scale
13,185.4
12,491.5
11,797.5
11,103.5
10,409.6
9,715.6
9,021.6
8,327.6
7,633.7
6,939.7
6,245.7
5,551.8
4,857.8
4,163.8
3,469.9
2,775.9
2,081.9
1,387.9
694
0
Cluster 4 – Up in Donor E66
OK

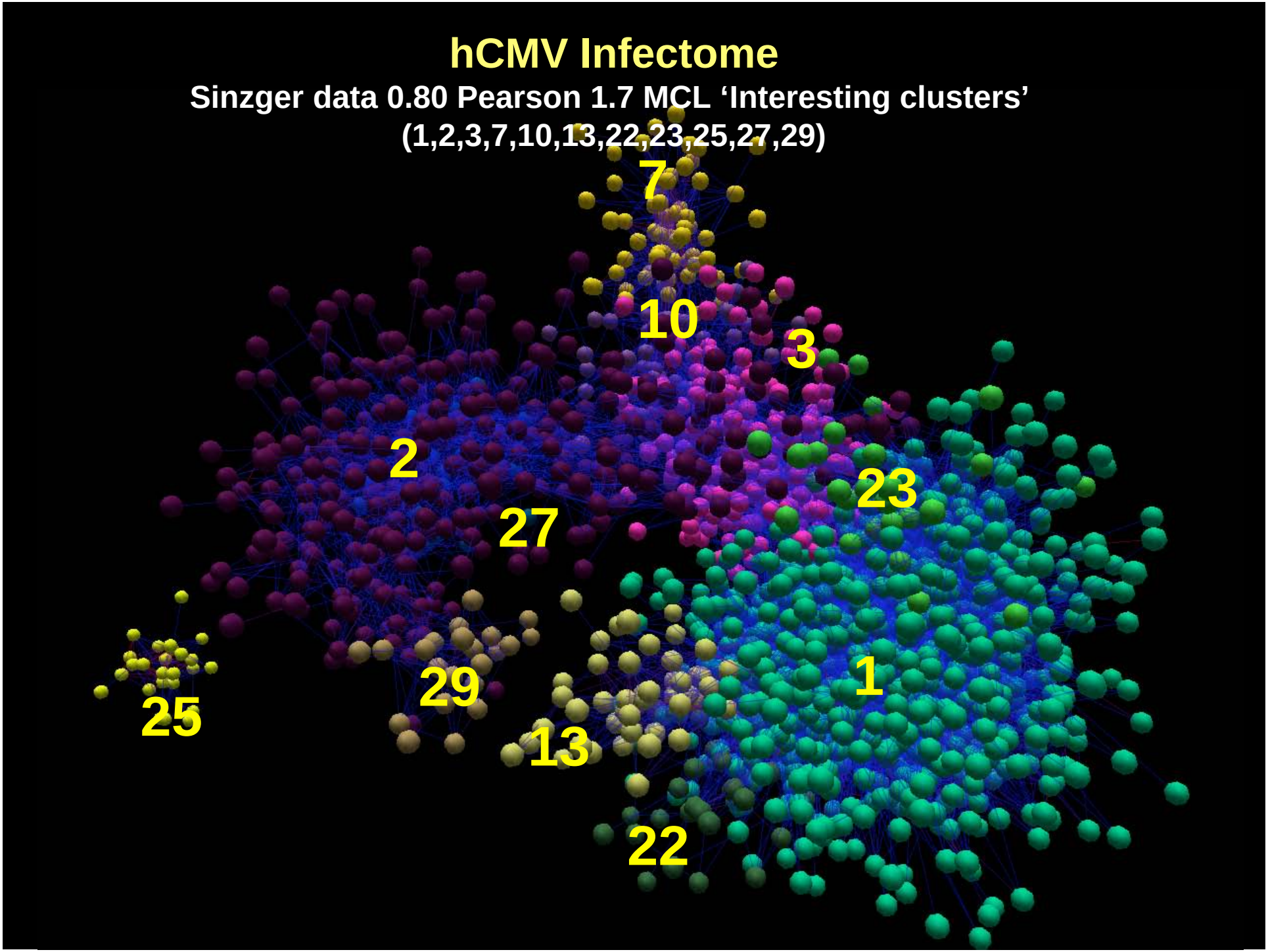

10,934.9
10,359.4
9,783.9
9,208.4
8,632.8
8,057.3
7,481.8
6,906.3
6,330.8
5,755.2
5,179.7
4,604.2
4,028.7
3,453.1
2,877.6
2,302.1
1,726.6
1,151
575.5
0
Cluster 8 – Up in Donor E35
OK

Cluster 7 – Up in Sta early



Cluster 25 – Up in Asp (8 h)

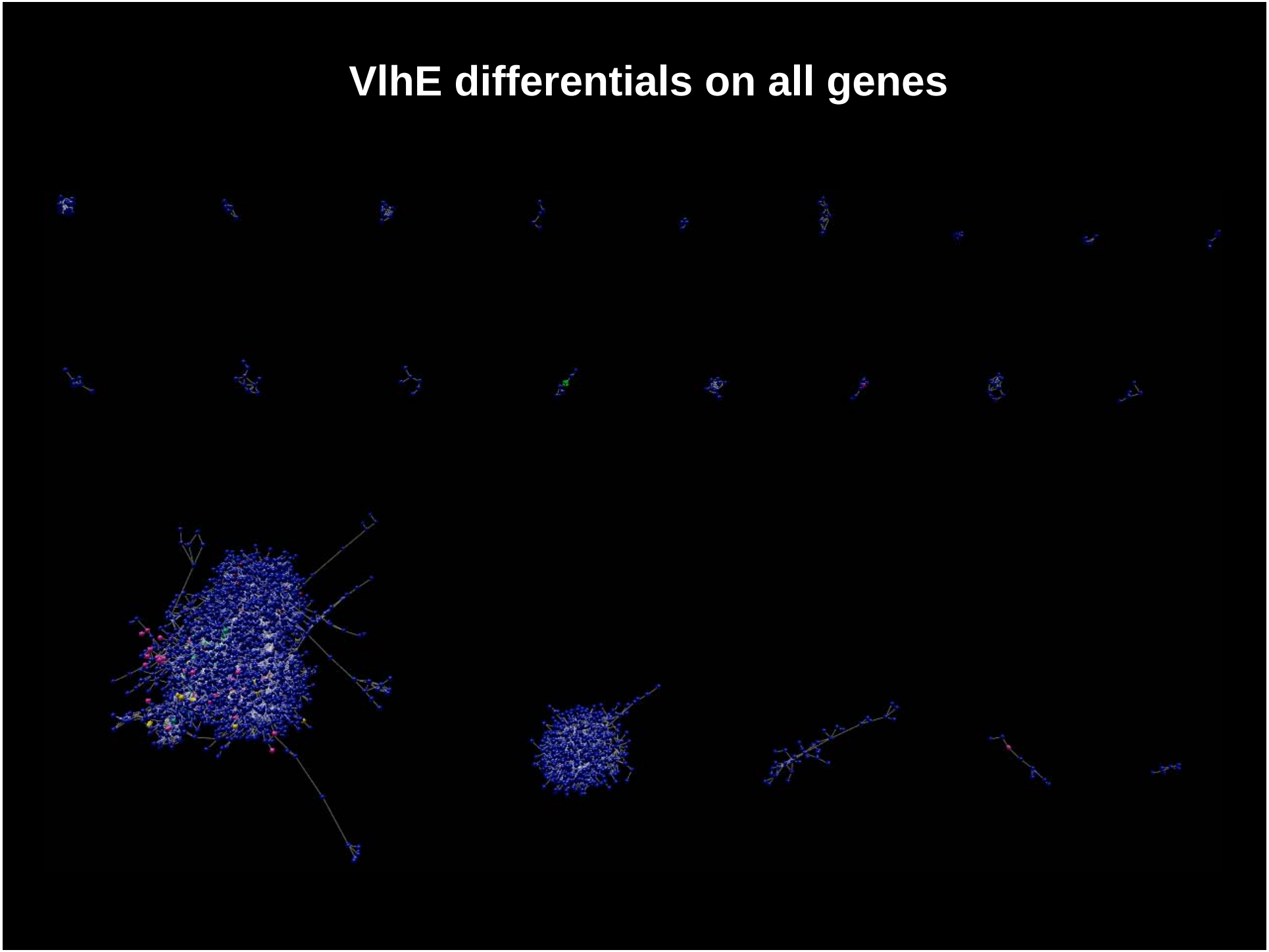



hCMV Infectome
Sinzger data 0.80 Pearson 1.7 MCL 'Interesting clusters'
(1,2,3,7,10,13,22,23,25,27,29)
7
10
3
2
23
27
29
1
25
13
22

Graphical Display of Statistical Hits and Pathway Genes

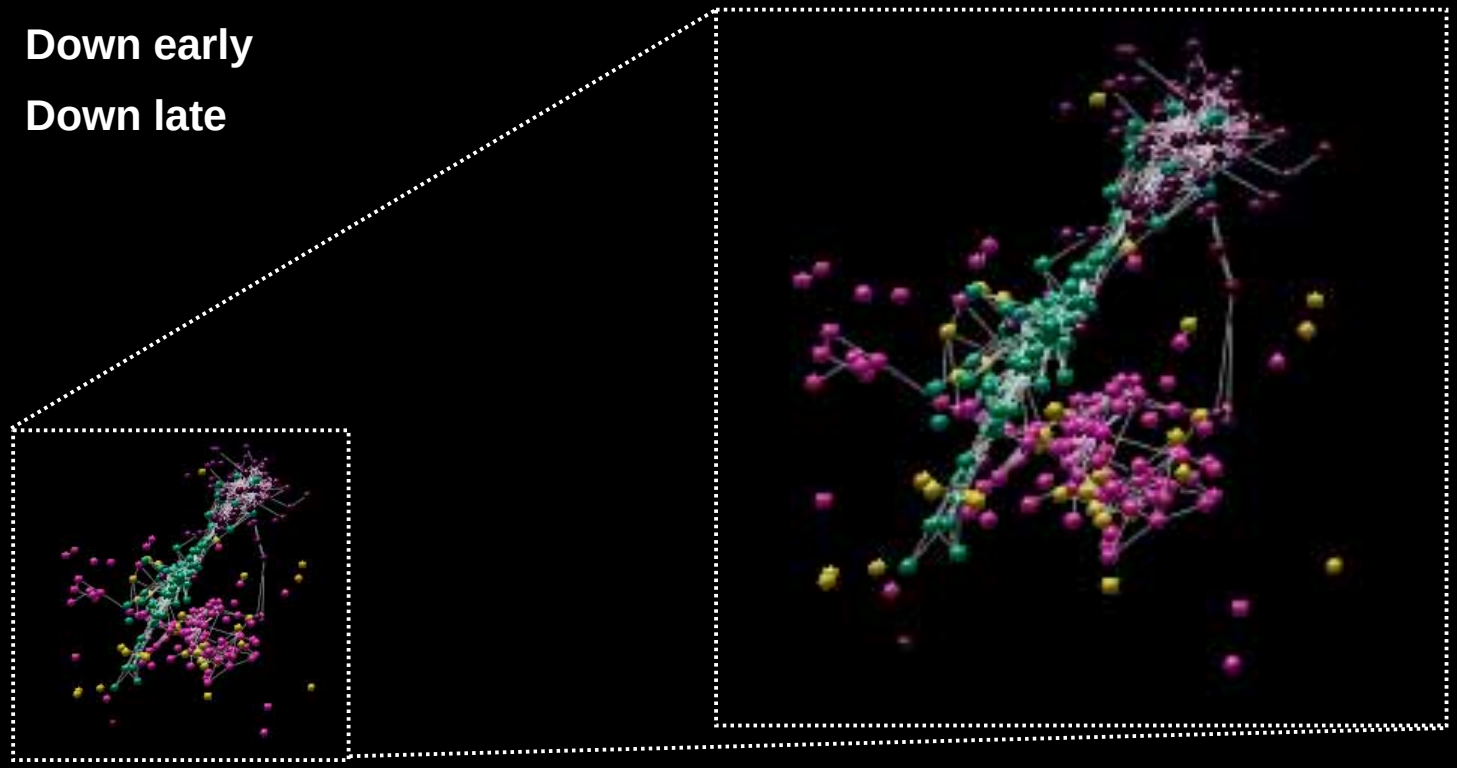
- Statistical analysis of differential expression VlhE and VlhF vs. mock control
- interferon pathway and apoptosis genes

VlhE differentials on all genes

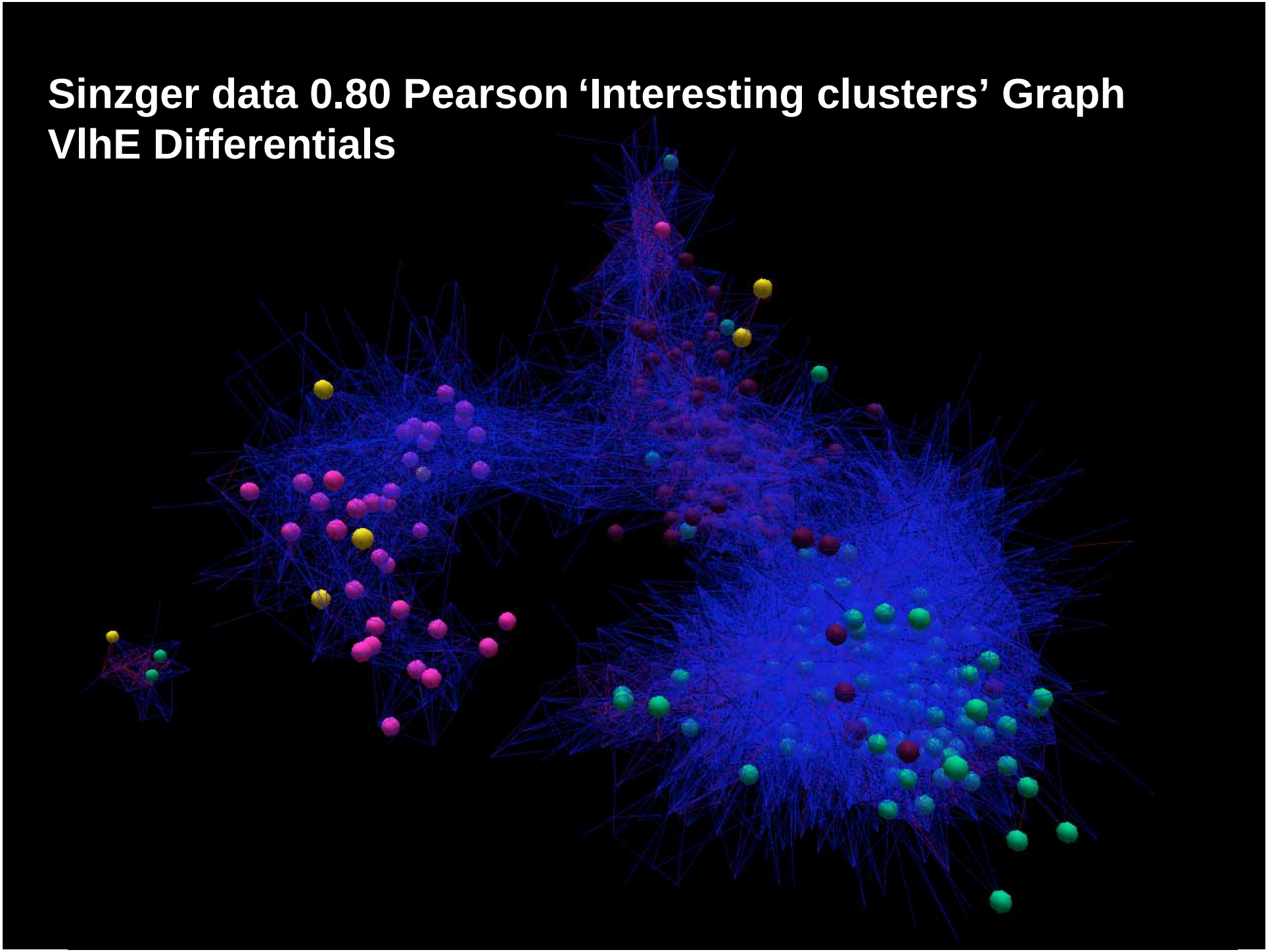


VlhE differentials

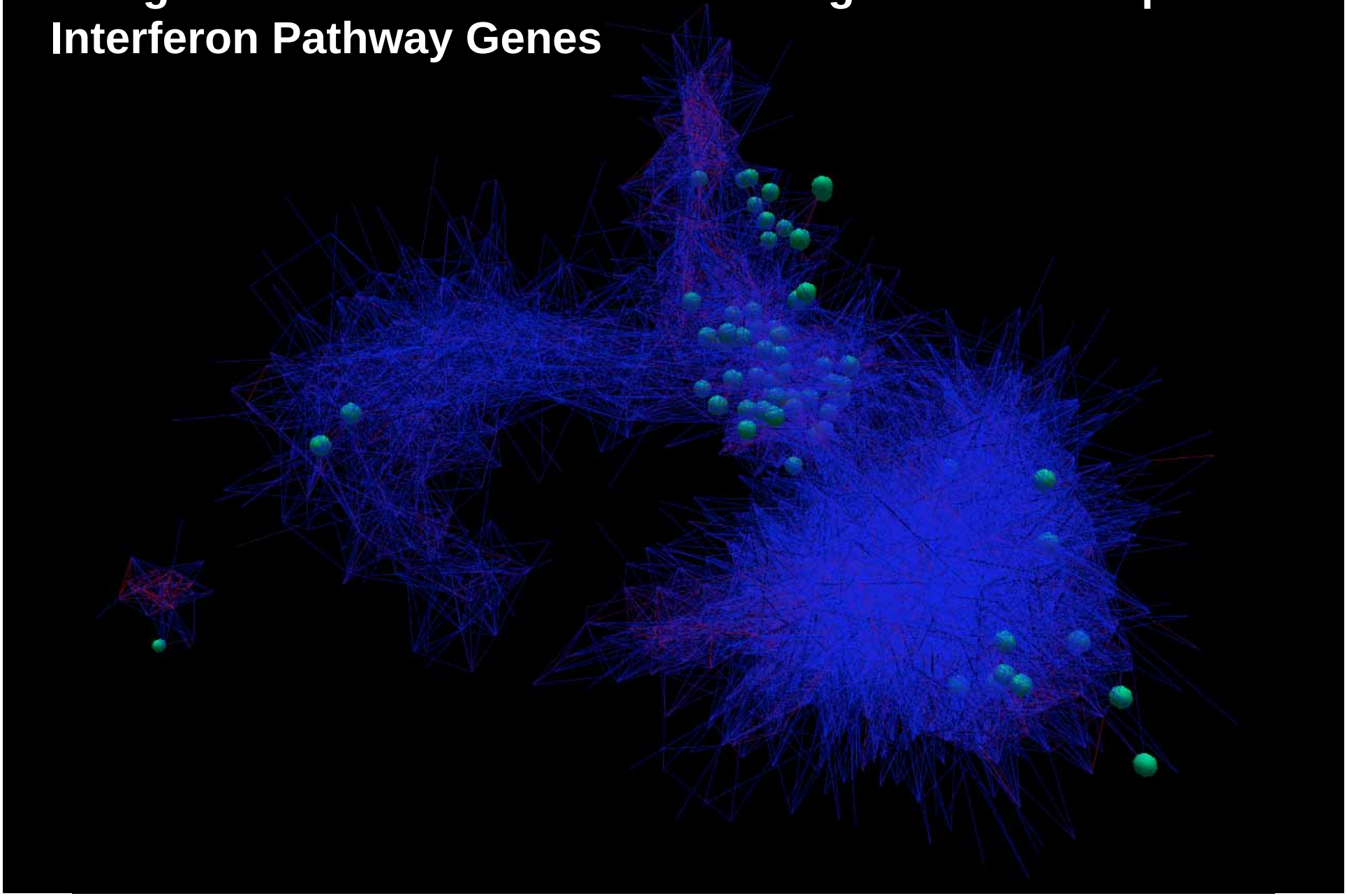
- Up early
- Up late
- Down early
- Down late

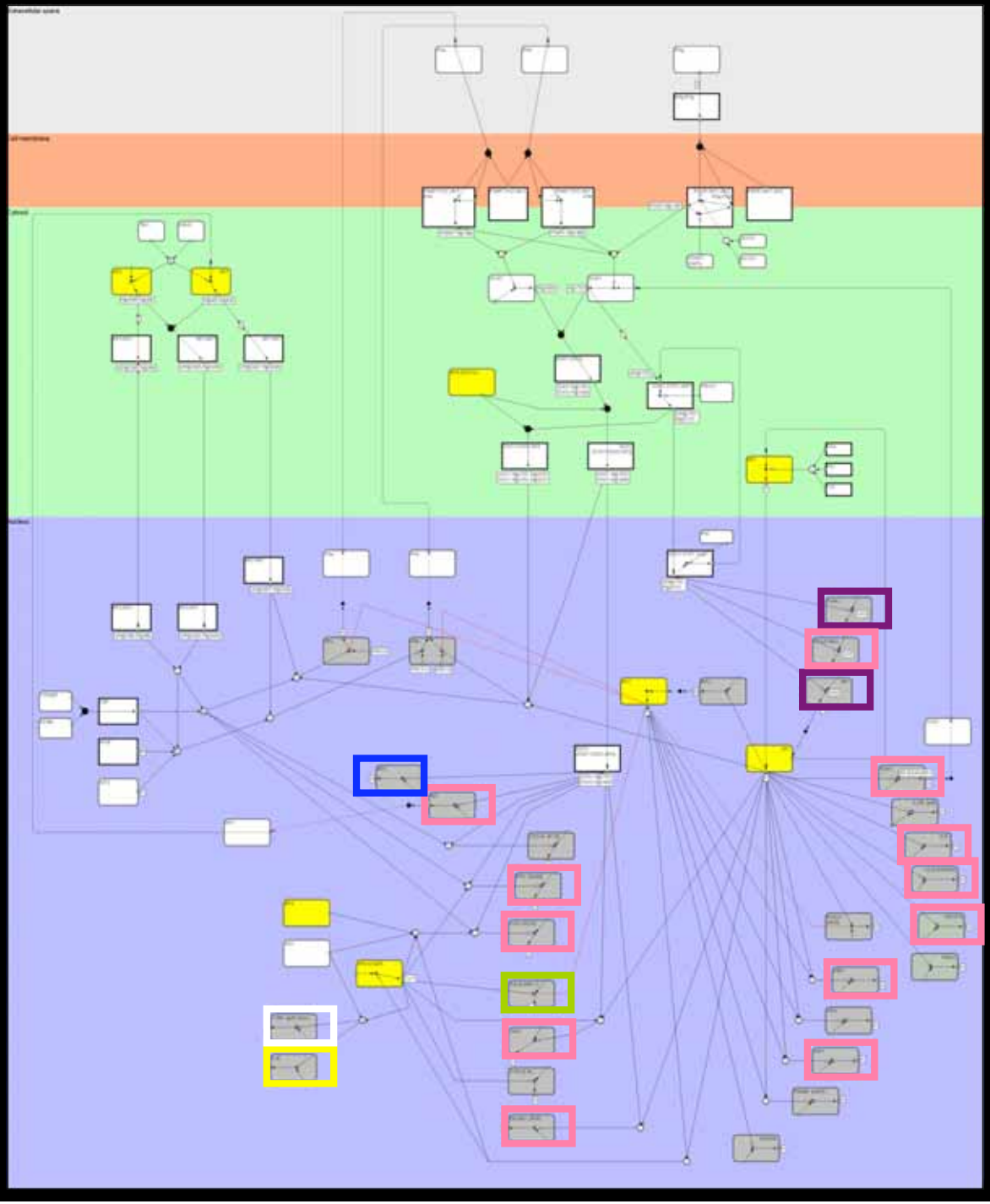


Sinzger data 0.80 Pearson ‘Interesting clusters’ Graph VlhE Differentials



Sinzger data 0.80 Pearson ‘Interesting clusters’ Graph Interferon Pathway Genes

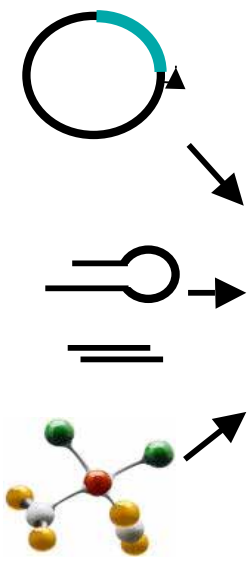




Network to IFN Pathway

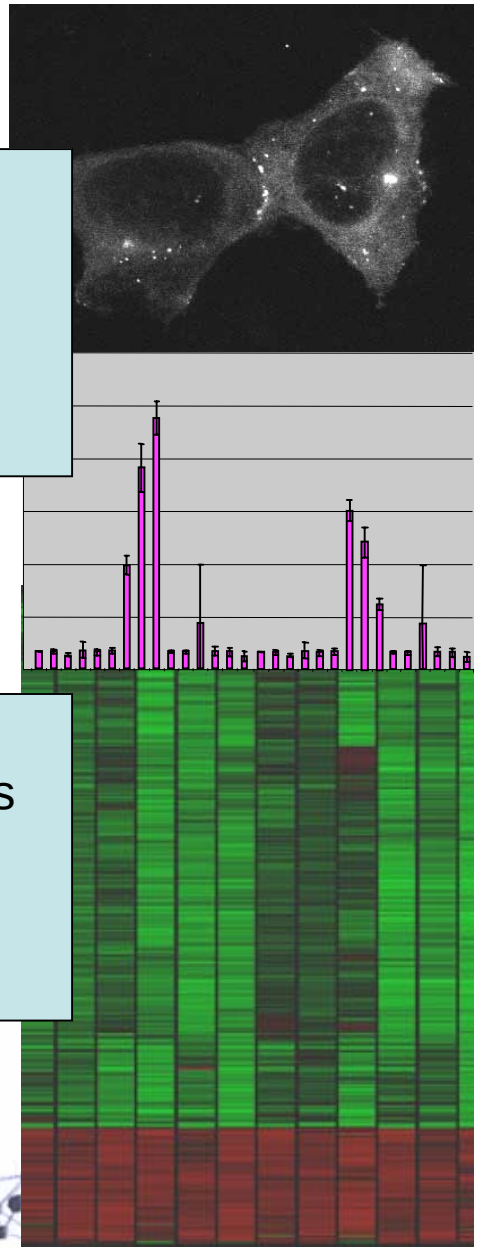
Cluster 3
Cluster 2
Cluster 7
Cluster 43
Cluster 77
No Class

A Systems Biology Approach



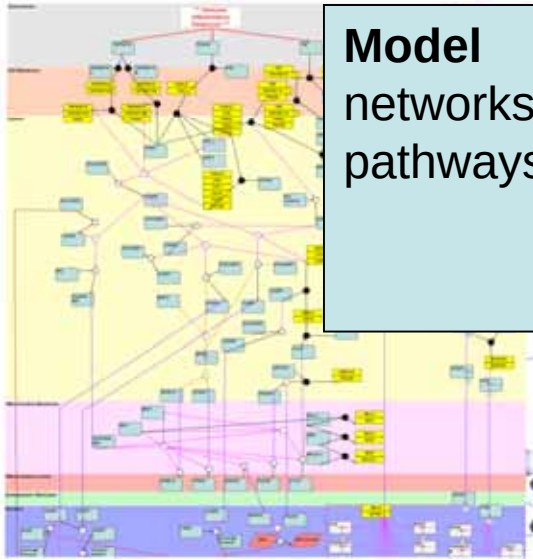
Manipulate
Knock-in
RNAi
compound
GM

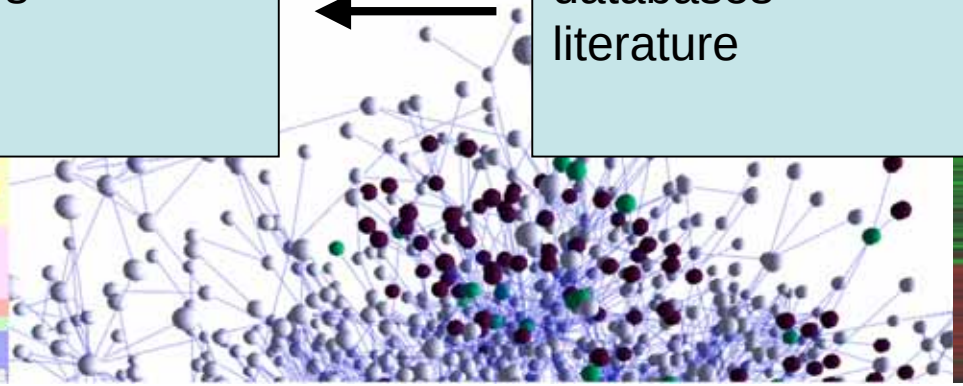
Measure
apoptosis
markers
microarray



Mine
data analysis tools
databases
literature

Model
networks
pathways





Acknowledgements

Ella Palmer, RFCGR, UK

Andy Miller, Genetic Therapies Centre, Imperial College, UK

Christian Sinzger, University of Tubingen, DE

Anton Enright, Sanger Institute, Cambridge, UK

Peter Ghazal, GTI, Edinburgh, UK

Interested in learning more?
Wellcome Trust Advanced Course in
‘Functional Genomics and Systems Biology’
Sanger Institute, Cambridge UK, June 2007