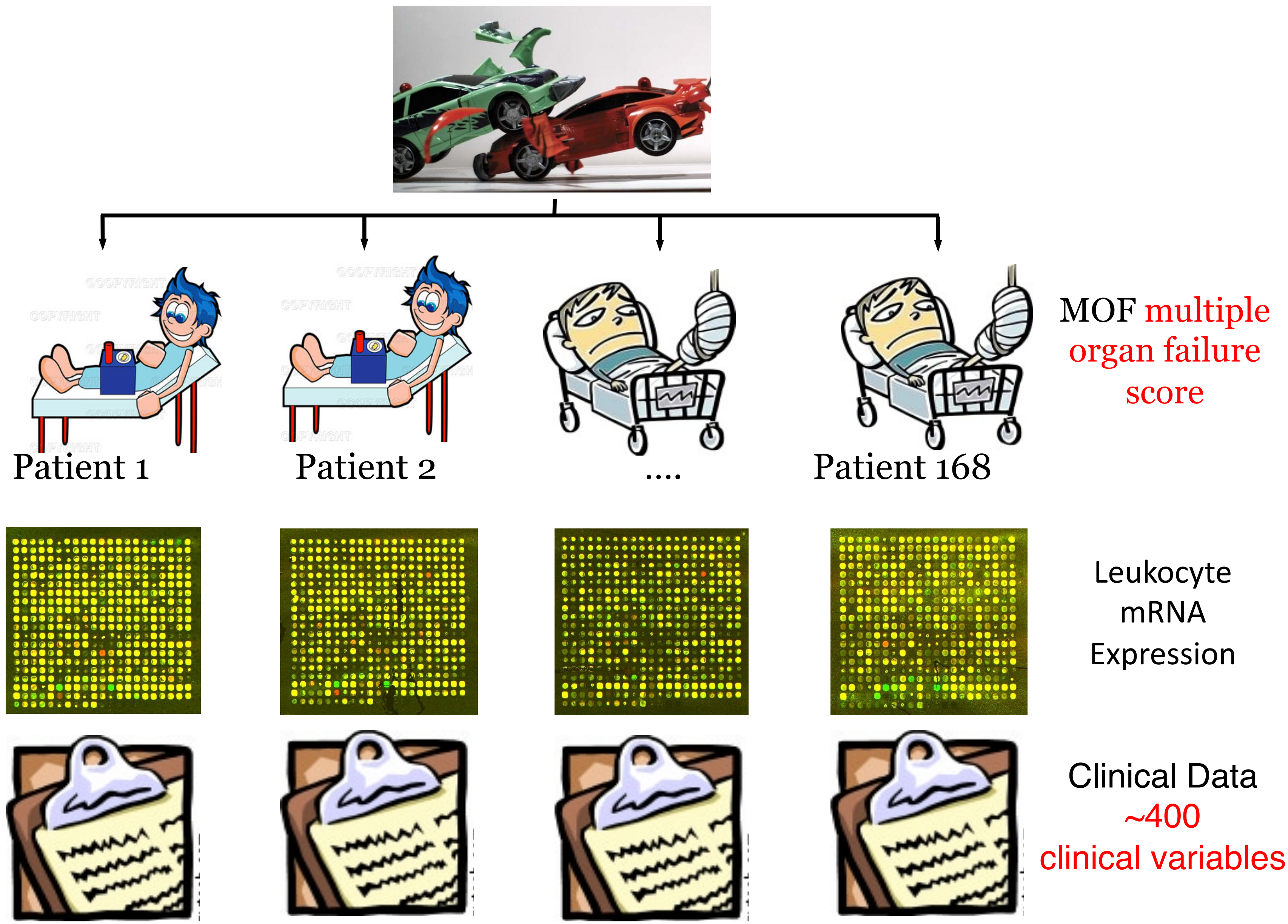
Dissecting Inflammatory Complications in Critically Injured Patients by Within-Patient Gene Expression Changes

John D. Storey
Lewis-Sigler Institute for Integrative Genomics
Department of Molecular Biology
Program in Applied and Computational Mathematics
Princeton University

Why Trauma?

- Top killer of individuals 1-44 years old in the U.S.
- Among most expensive healthcare costs in developed countries
- Injuries frequently lead to inflammation, sepsis, and multiple organ failure (MOF)
- History of failed clinical trials and poorly understood biology

Inflammation and the Host Response to Injury



NIH "Glue Grant"

- Almost $100 million; 10th largest NIH grant
- 10 year effort, 2001-2011

- 1487 critically injured patients: Longitudinal clinical data (> 393 variables)
- 168 patients: longitudinal leukocyte gene expression
- 111 patients: longitudinal cell separated gene expression (monocytes, neutrophils, T-cells)

The Lancet, Volume 365, Issue 9458, Pages 454 - 455, 5 February 2005 < Previ
doi:10.1016/S0140-6736(05)17878-7 (?) Cite or Link Using DOI

Microarrays and molecular research: noise discovery?

John PA Ioannidis [a] [b] [c]

An array of problems

Despite the huge amount of published microarray data in cancer, little is being converted into clinical practice. Validating initial data is proving to be a key challenge, reports SIMON FRANTZ.

Microarrays in the clinic

Guy W Tillinghast

...rtium has evaluated methods
...large-scale gene expression data.

CANCER

Why Most Gene Expression Signatures of Tumors Have Not Been Useful in the Clinic

Serge Koscielny

Published 13 January 2010; Volume 2 Issue 14 14ps2

Expectations, validity, and reality in gene expression profiling

Kyoungmi Kim[a,b,*], Stanislav O. Zakharkin[c,1], David B. Allison[d]

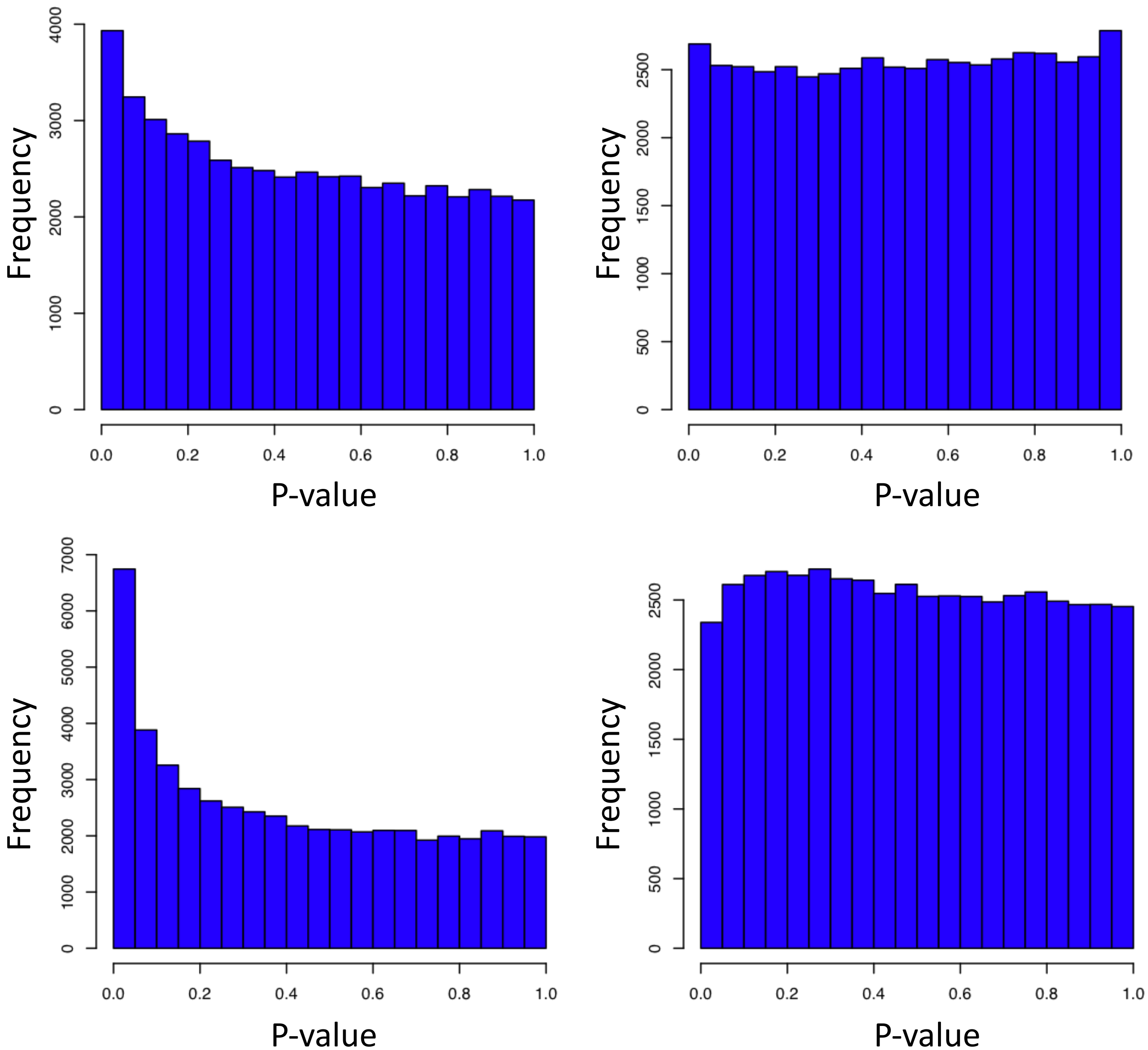
Data at Initial Time Point

Characteristic	Phase I (n=42)	Phase II (n=37)	Phase III (n=56)	Phase IV (n=31)
Age, mean (SD), y	32.3 (9.5)	33.5 (11.8)	32.7 (11.3)	39.0 (11.2)
Male sex, No. (%)	24 (57)	24 (65)	40 (71)	19 (61)
Racial/Ethinic Group , No. (%)				
White, non-Hispanic	33 (79)	30 (81)	48 (86)	23 (75)
Black, non-Hispanic	4 (10)	1 (3)	2 (4)	2 (6)
Hispanic	3 (7)	2 (5)	2 (4)	4 (13)
Other or missing information	2 (5)	4 (11)	4 (5)	2 (6)
Date of Microarray, First, Last	5/25/04 – 9/02/04	9/27/04- 3/24/05	8/24/05- 2/14/06	5/25/06- 7/01/06
MOF Score	4.2 (2.3)	3.0 (1.8)	3.9 (2.0)	4.4 (2.0)

Four "Replicated" Studies

Phase 1

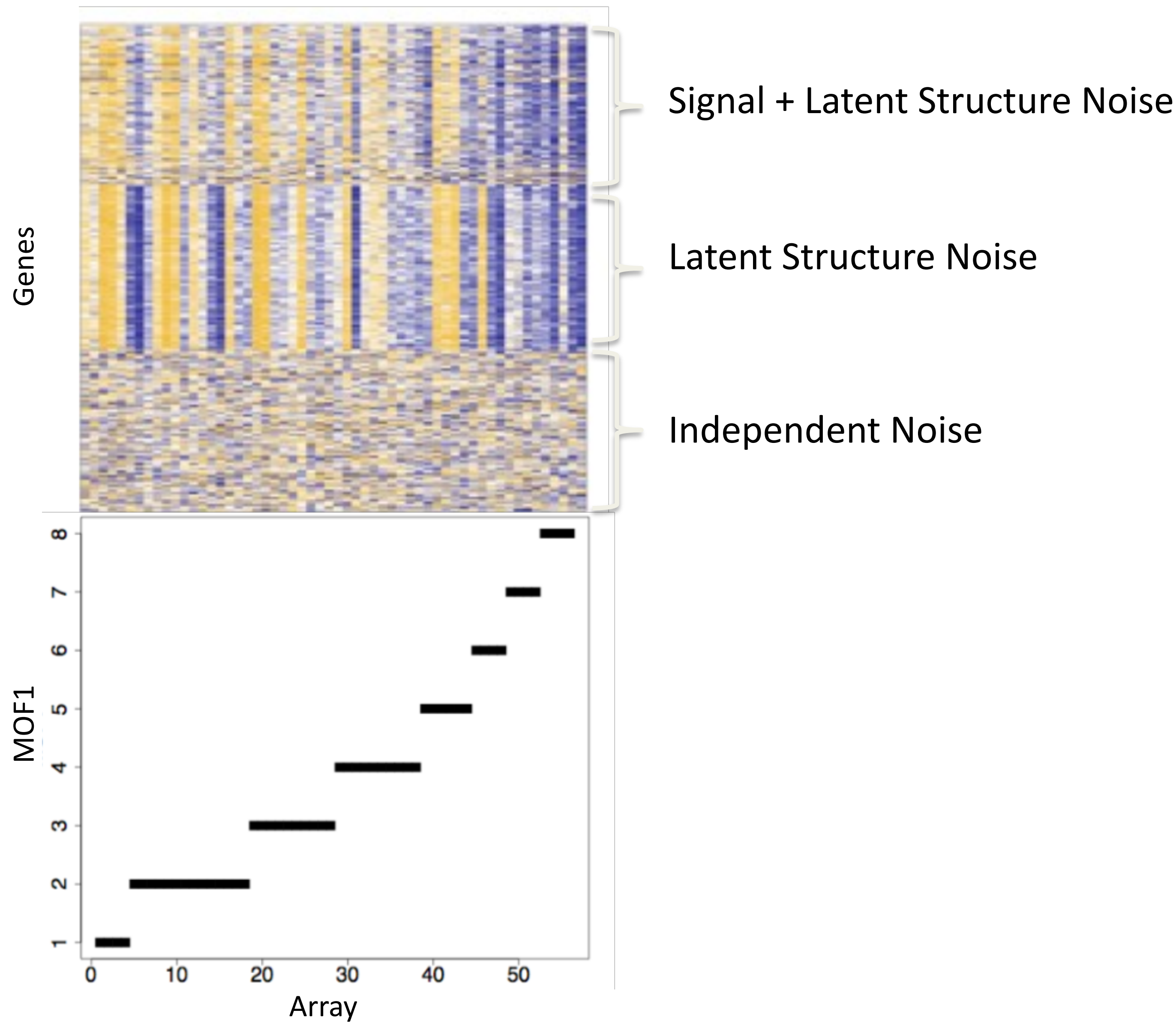
Phase 2



Phase 3

Phase 4

Latent Structure in the "Noise"



2007

OPEN ACCESS Freely available online PLoS GENETICS

Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis

Jeffrey T. Leek[1], **John D. Storey**[1,2*]

1 Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, 2 Department of Genome Sciences, University of Washington, Seattle, Washington, United States of America

2008

PNAS

A general framework for multiple testing dependence

Jeffrey T. Leek[a] and **John D. Storey**[b,1]

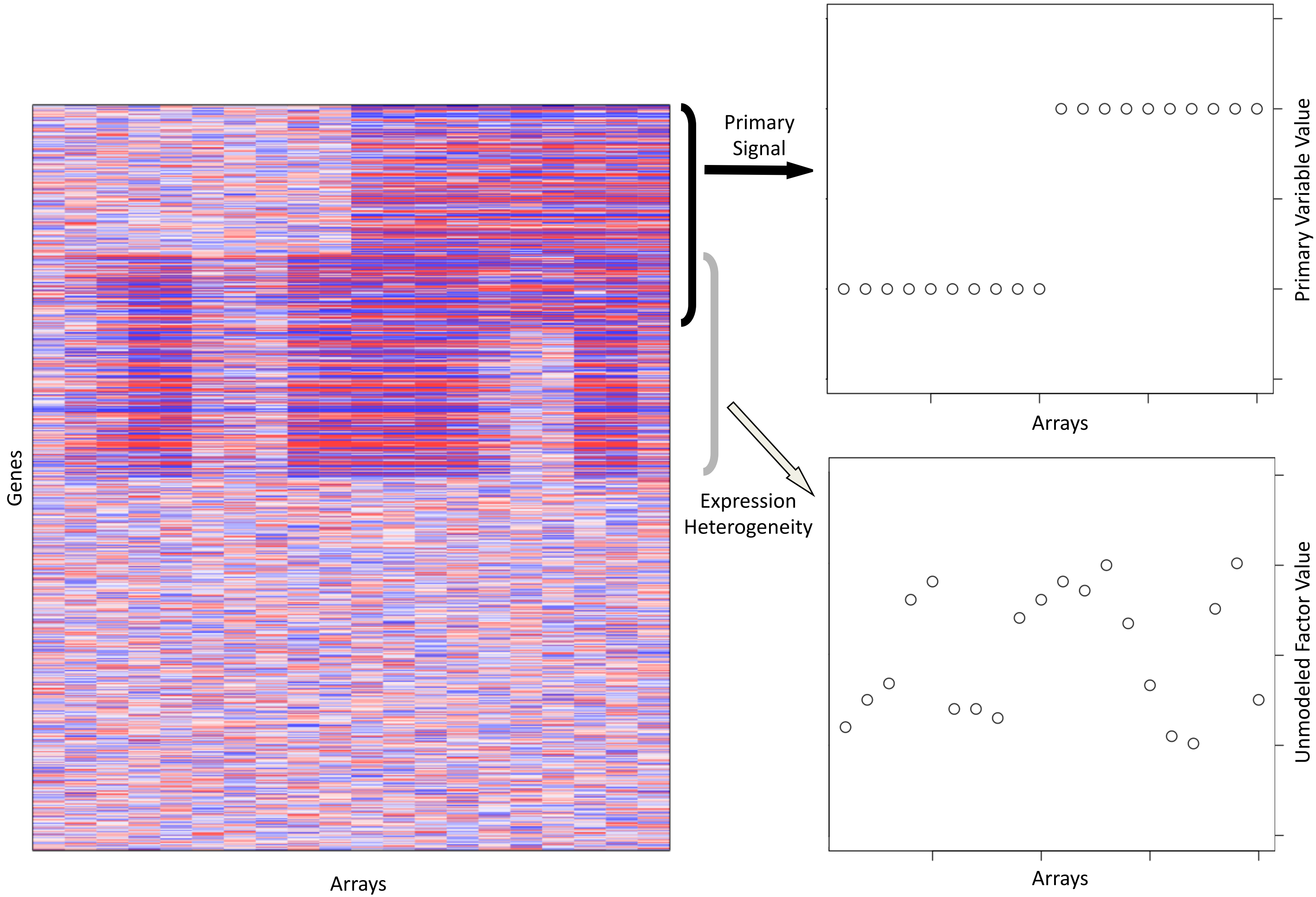
[a]Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD 21287; and [b]Lewis-Sigler Institute and Department of Molecular Biology, Princeton University, Princeton, NJ 08544

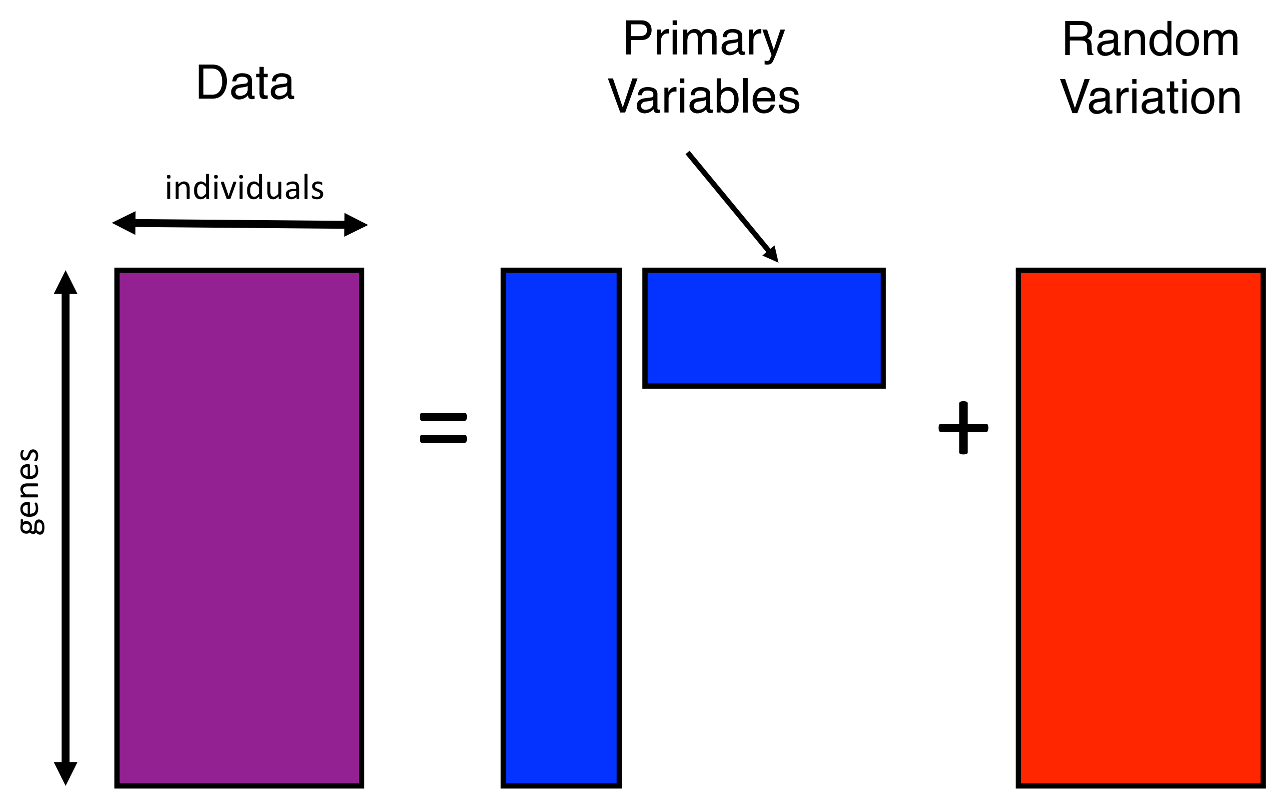
Communicated by Burton H. Singer, Princeton University, Princeton, NJ, September 4, 2008 (received for review May 8, 2008)

We develop a general framework for performing large-scale significance testing in the presence of arbitrarily strong dependence. We derive a low-dimensional set of random vectors, called a dependence kernel, that fully captures the dependence structure in an observed high-dimensional dataset. This result shows a surprising

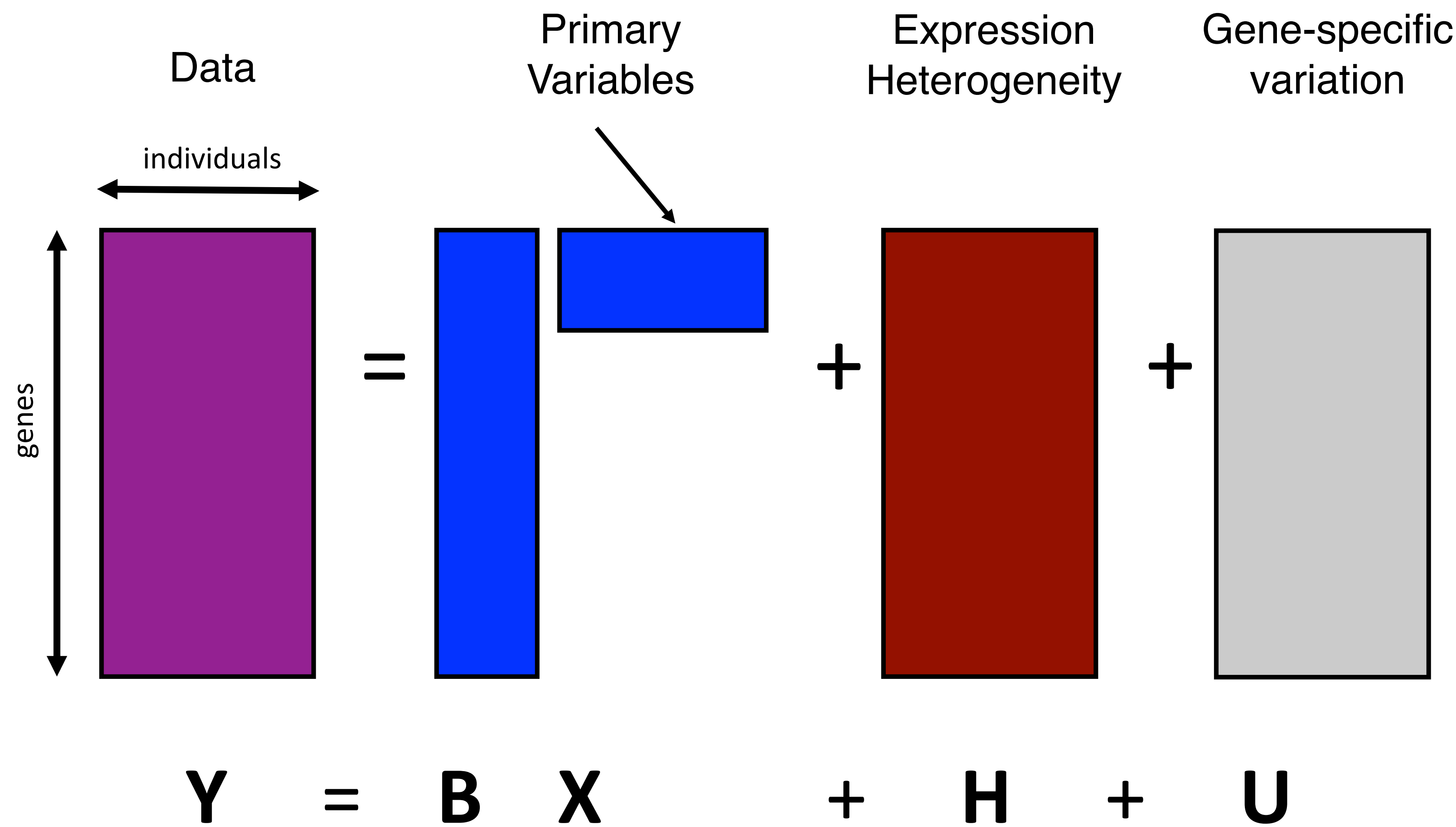
among multiple tests; no assumptions about a restricted dependence structure are required. By exploiting the dimensionality of the problem, we are able to account for dependence on each specific dataset, rather than relying on a population-level solution. We introduce a model that, when fit, makes the tests indepen-

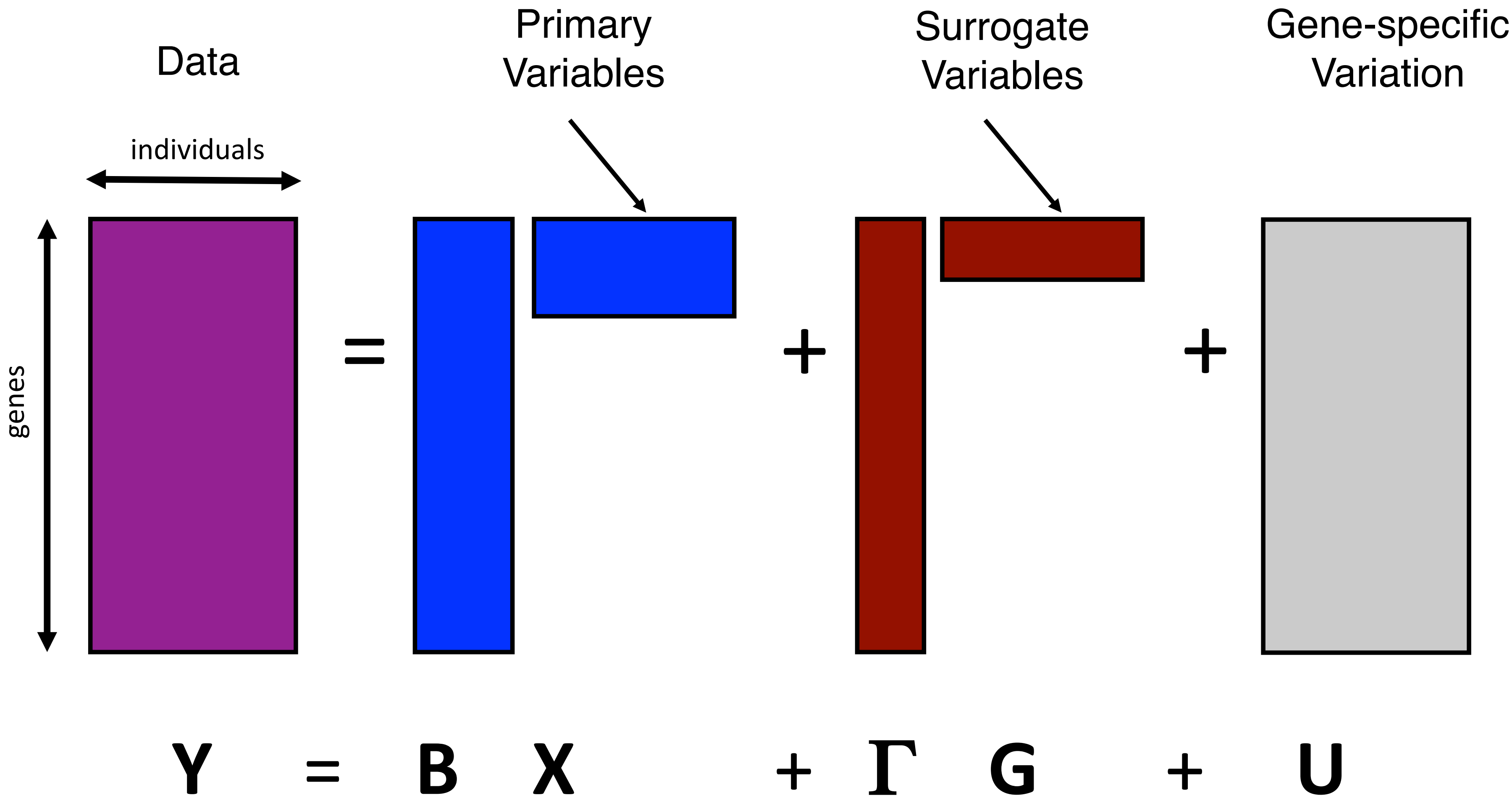
Surrogate Variable Analysis (SVA)



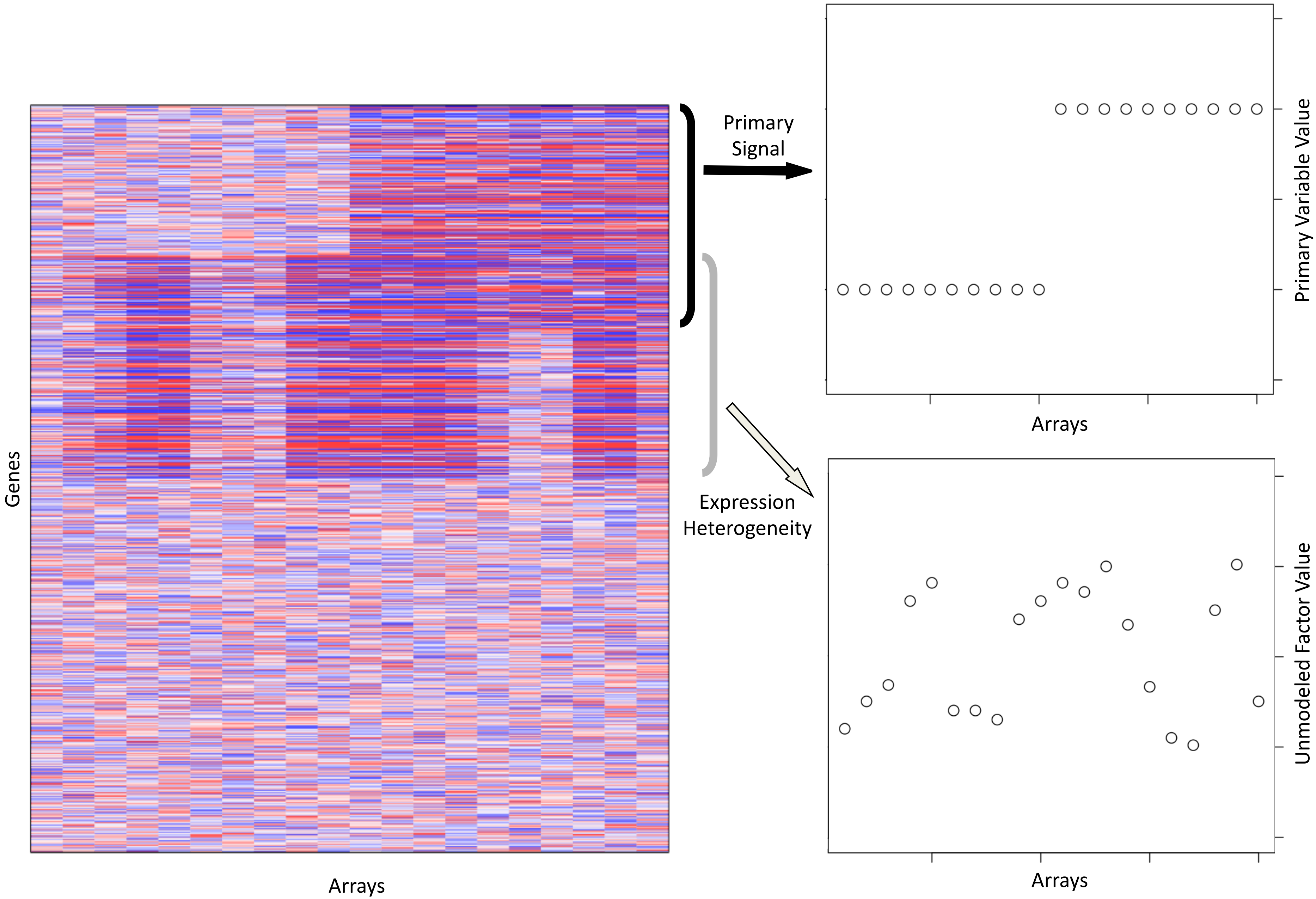

Data
Primary
Variables
Random
Variation
individuals
genes
=
+

$$Y = B\ X + E$$

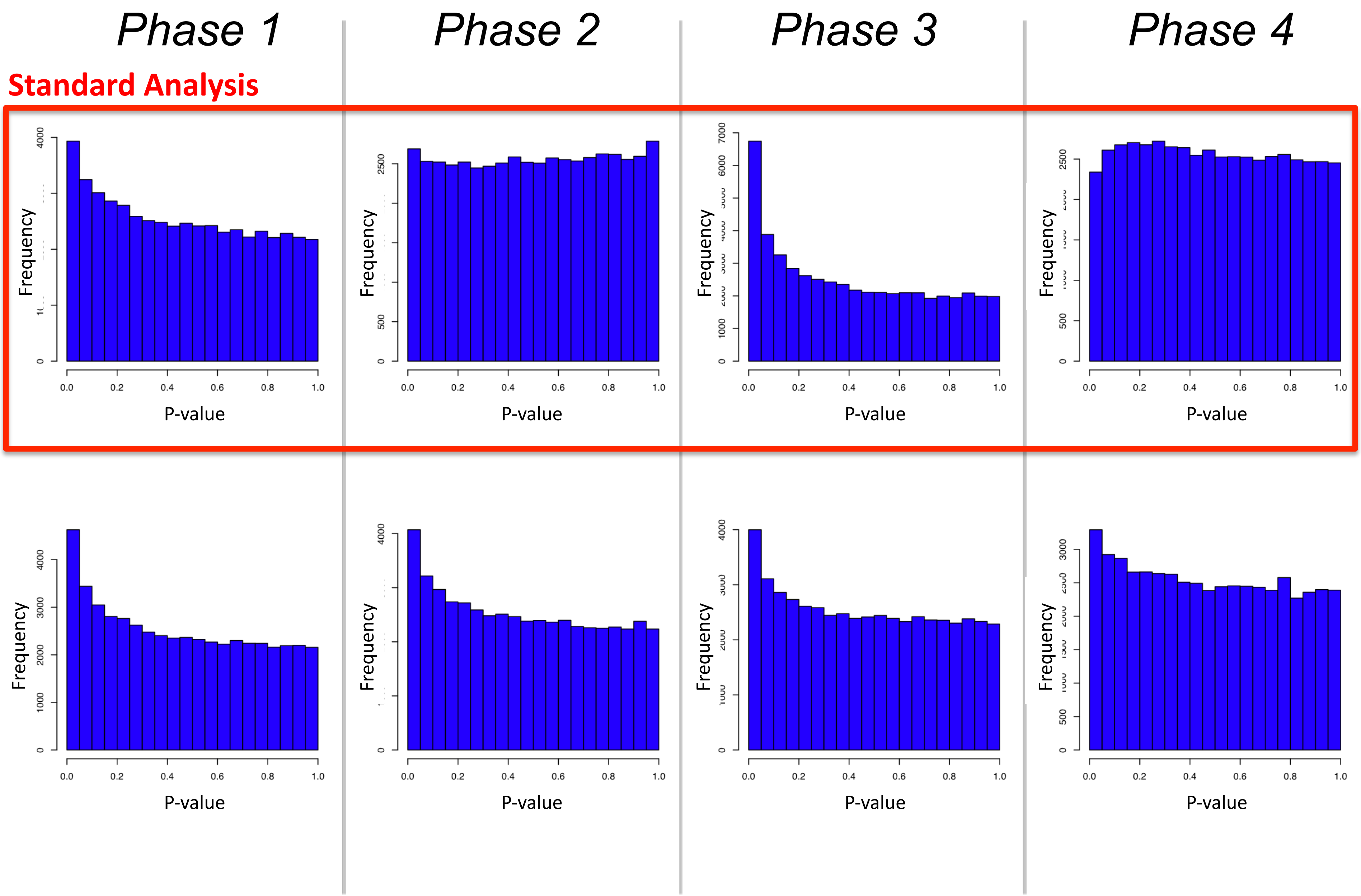

Data
Primary Variables
Expression Heterogeneity
Gene-specific variation
individuals
genes
Y = B X + H + U


Data
Primary Variables
Surrogate Variables
Gene-specific Variation
individuals
genes
Y = B X + Γ G + U

Surrogate Variable Analysis (SVA)

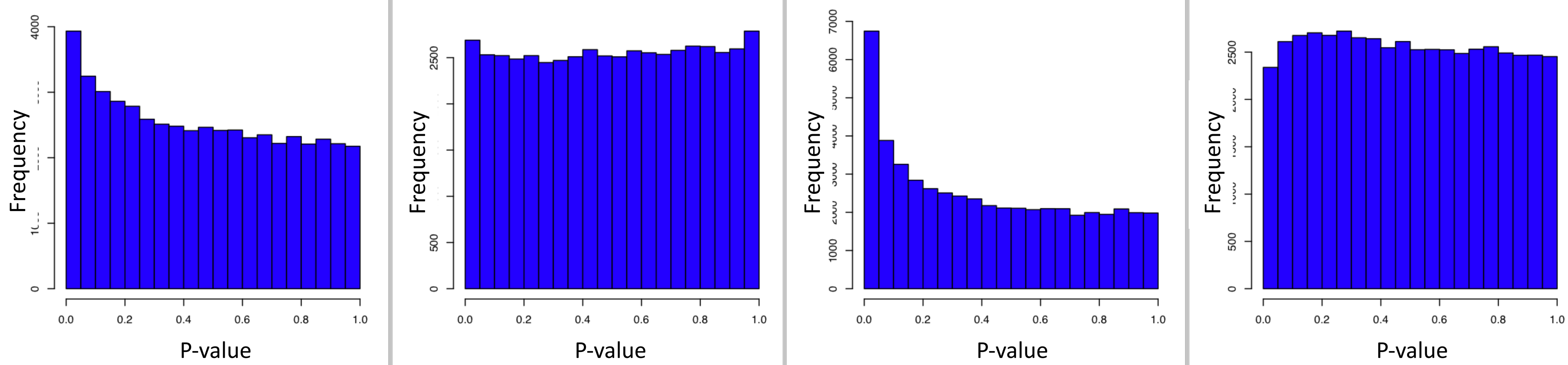


Before and After Modeling Latent Structure

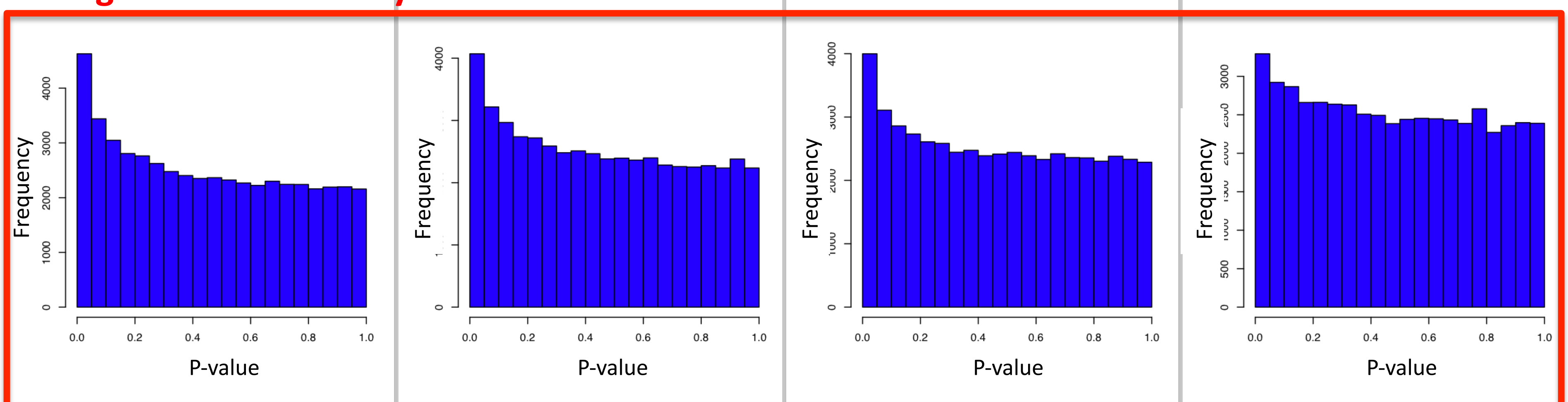


Before and After Modeling Latent Structure

Phase 1 | *Phase 2* | *Phase 3* | *Phase 4*



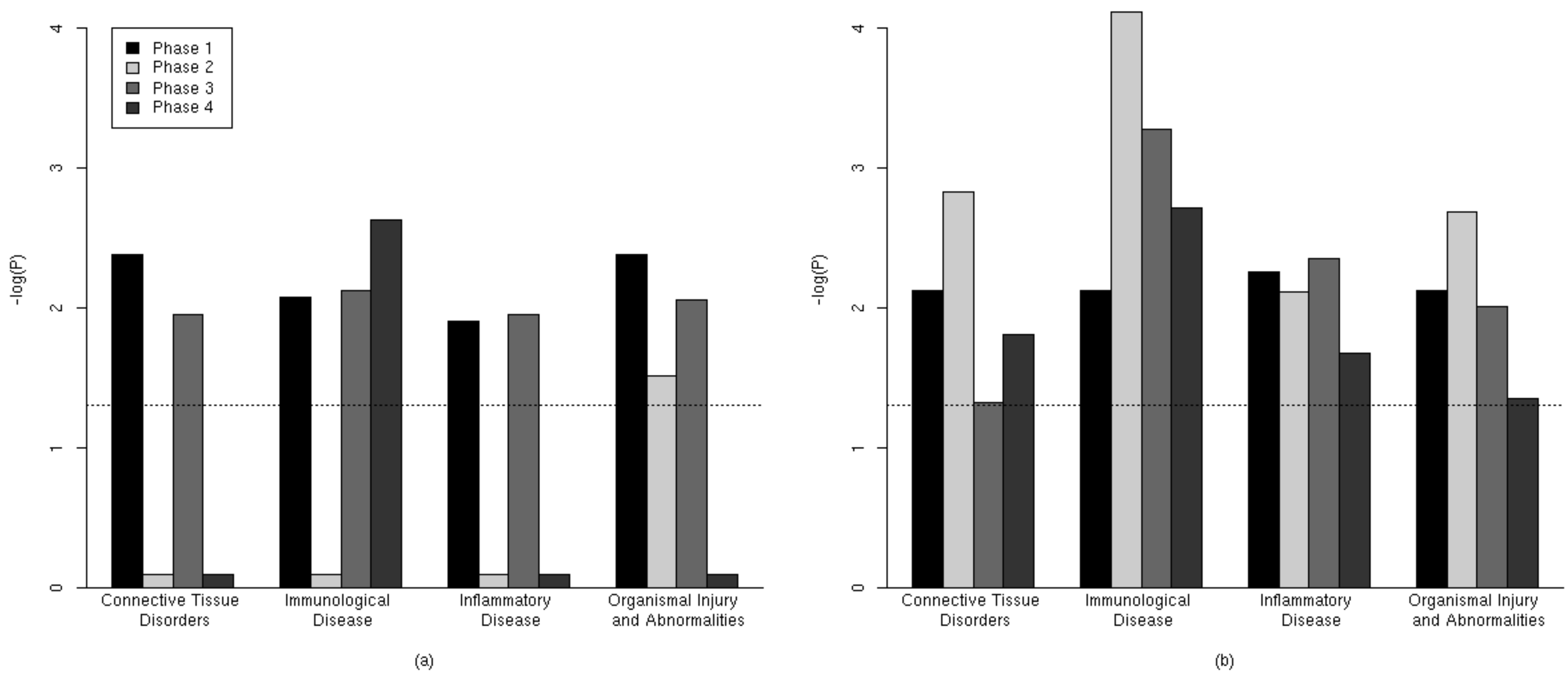
Surrogate Variable Analysis



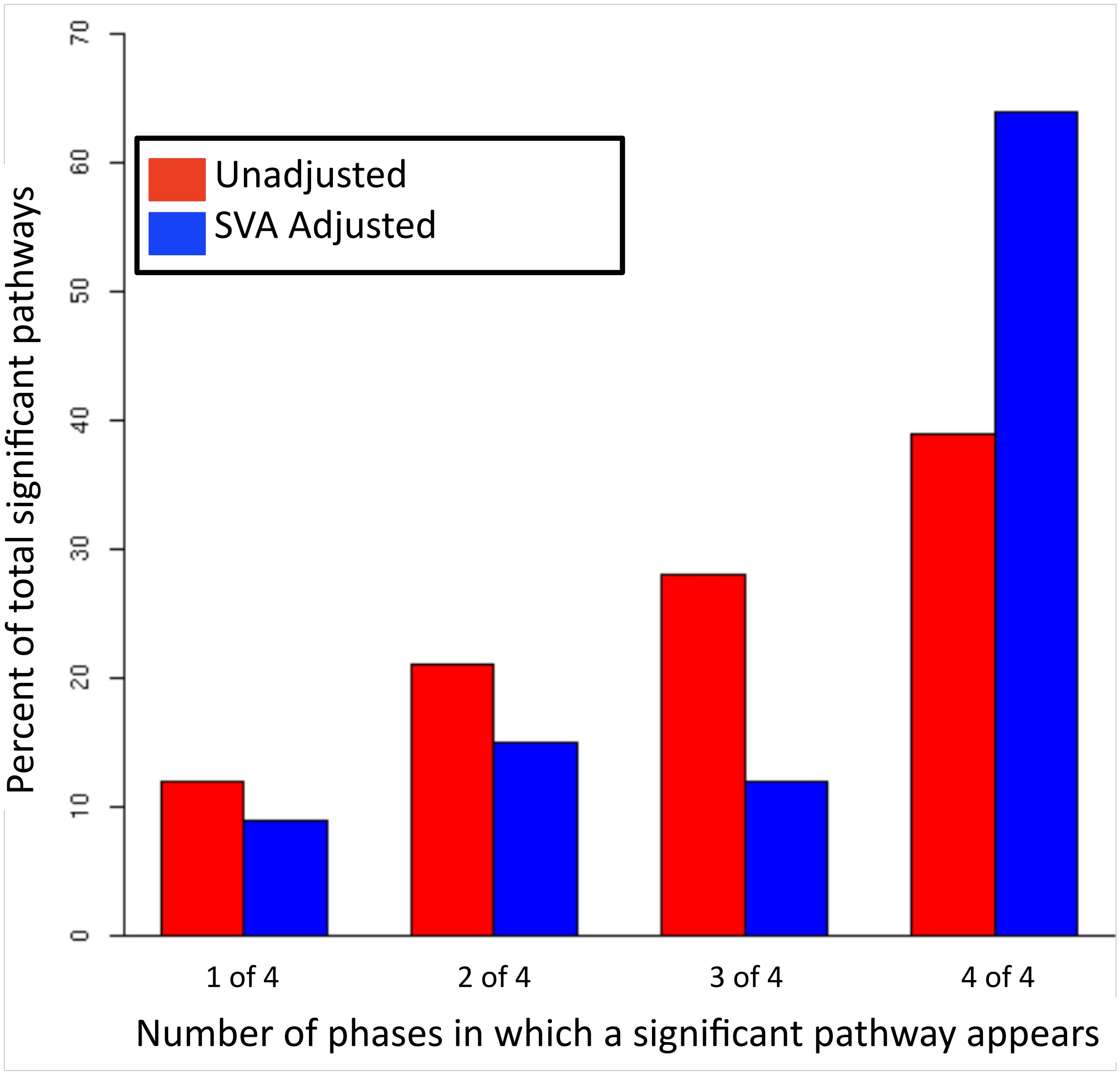
Ingenuity Pathways Analysis

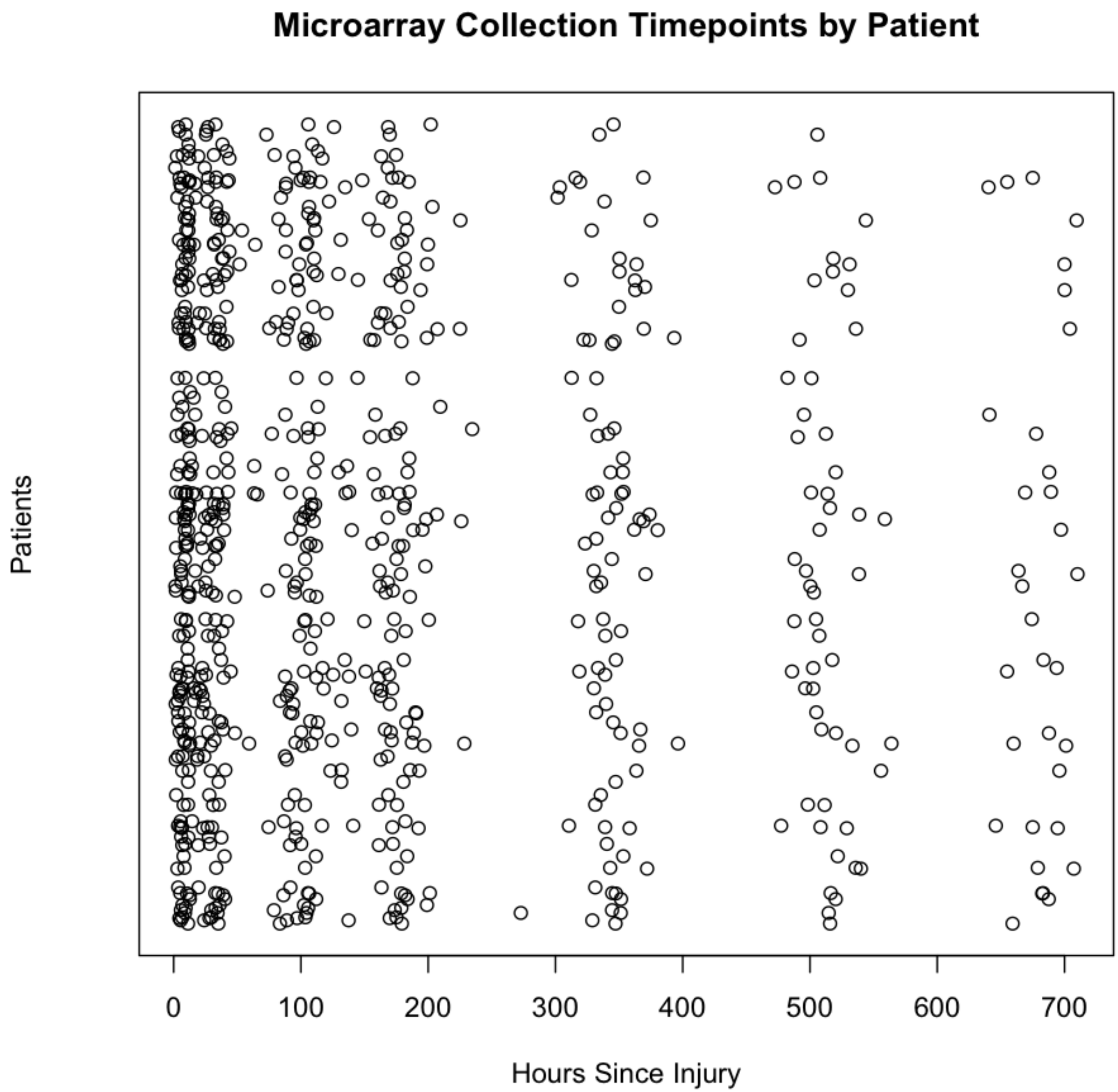
Unadjusted Analysis

Surrogate Variable Analysis

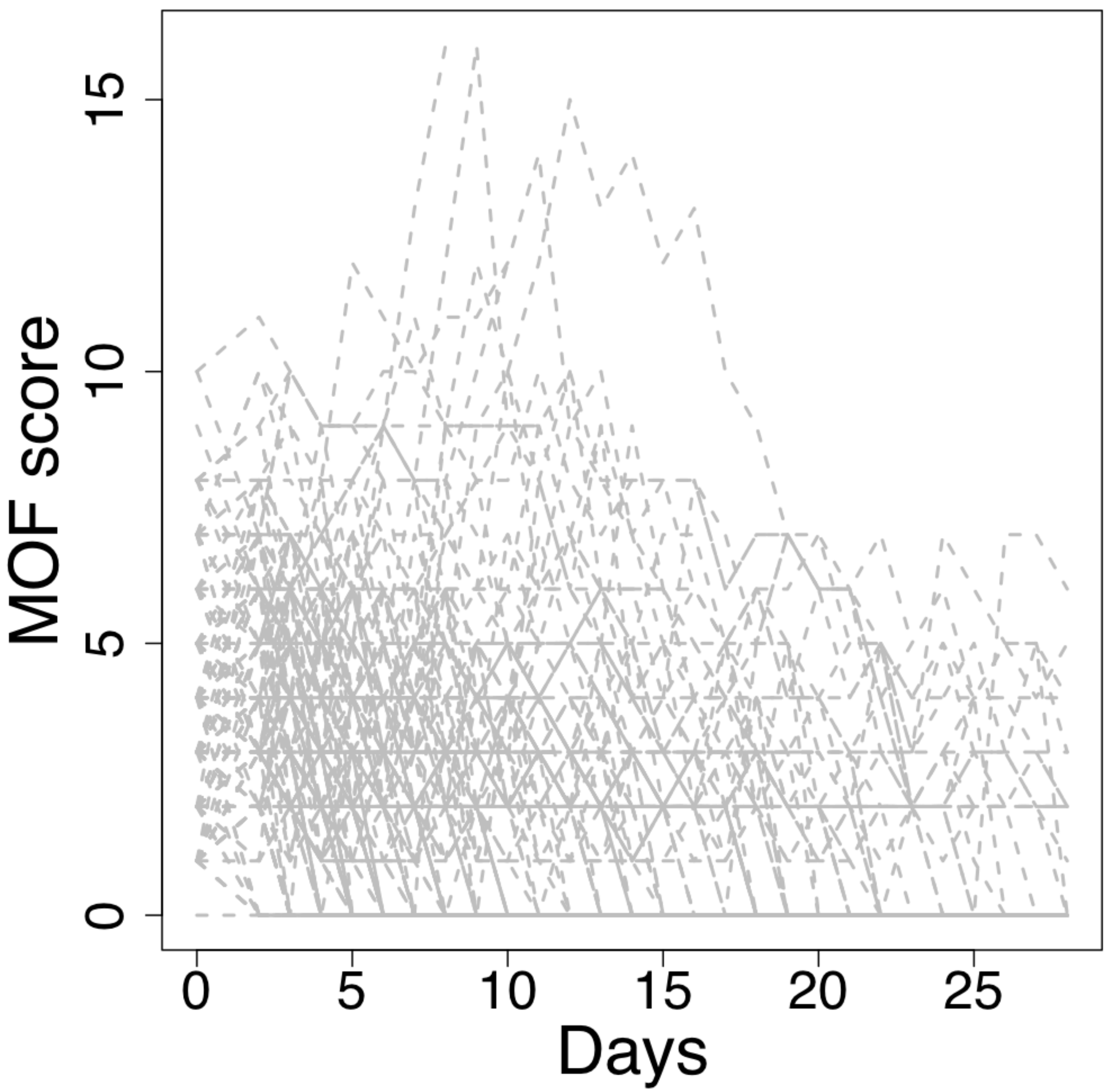


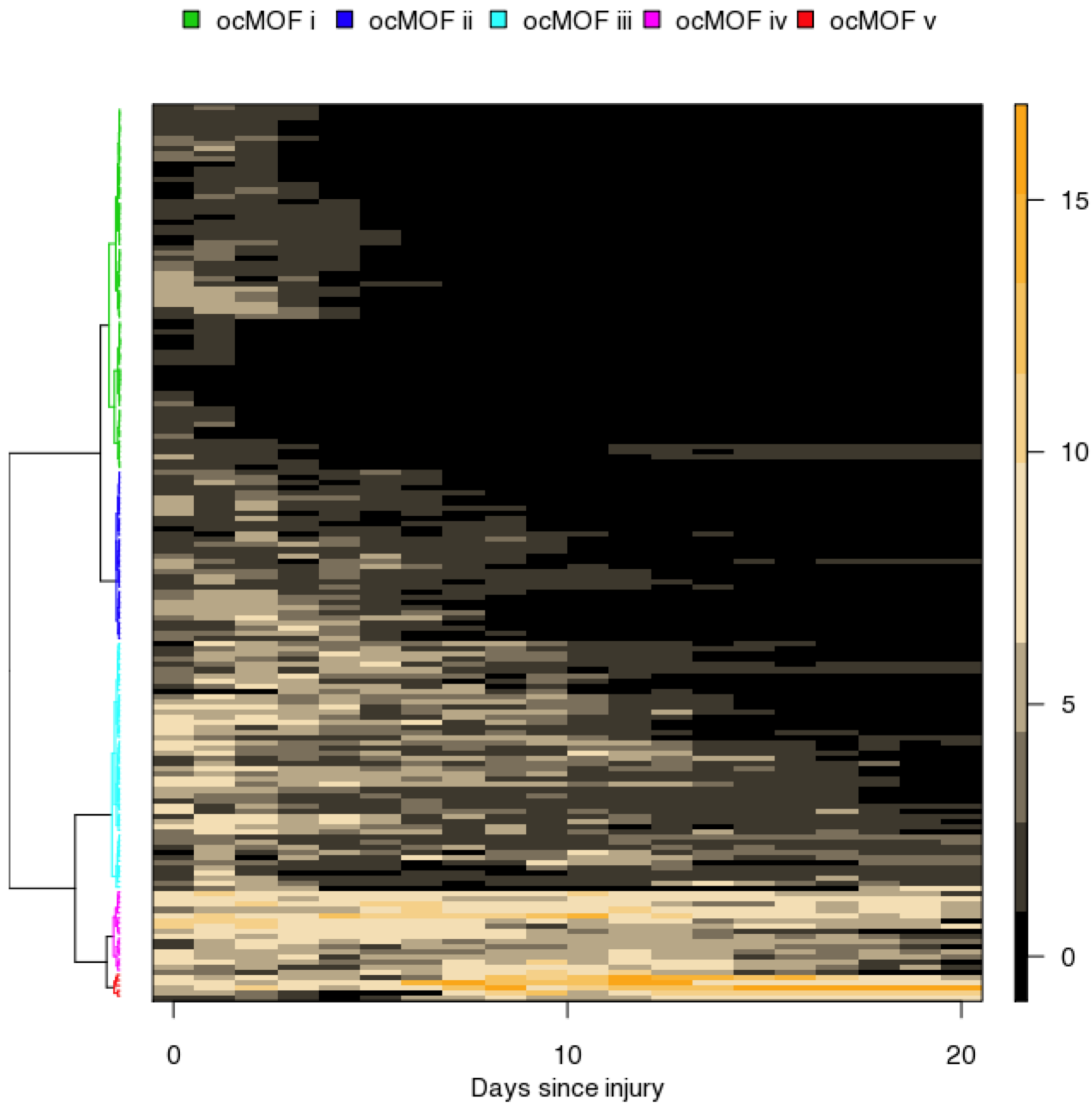
Functional Enrichment Across Phases



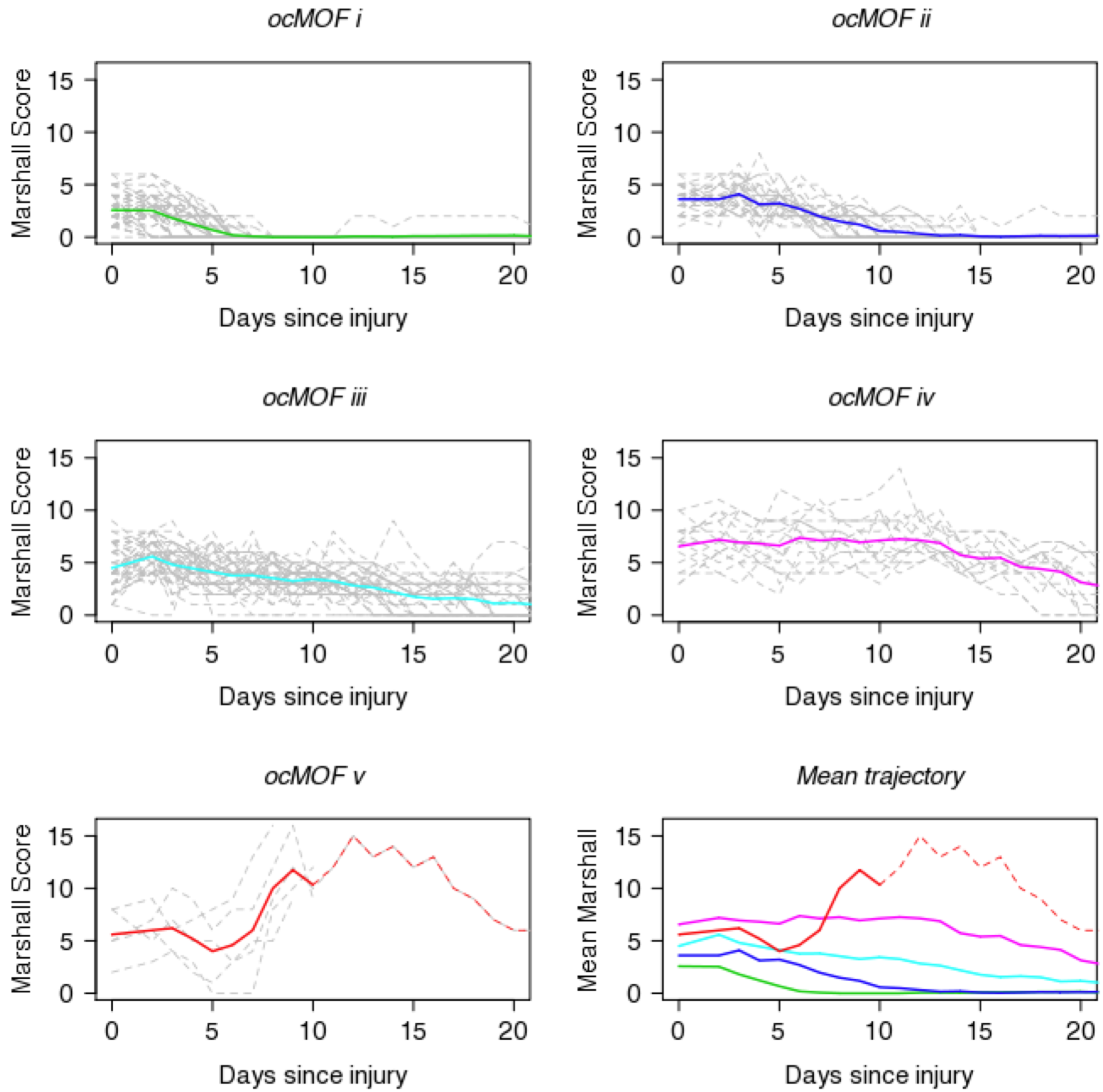

Microarray Collection Timepoints by Patient
Patients
Hours Since Injury
0
100
200
300
400
500
600
700

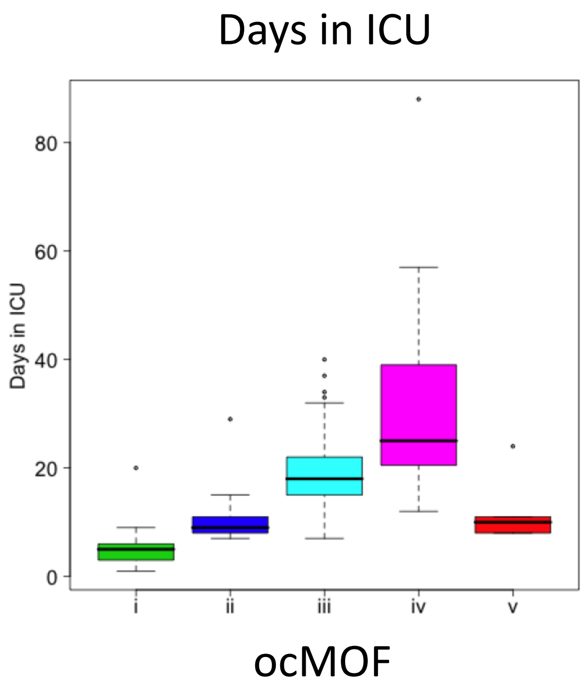
Marshall MOF Trajectories

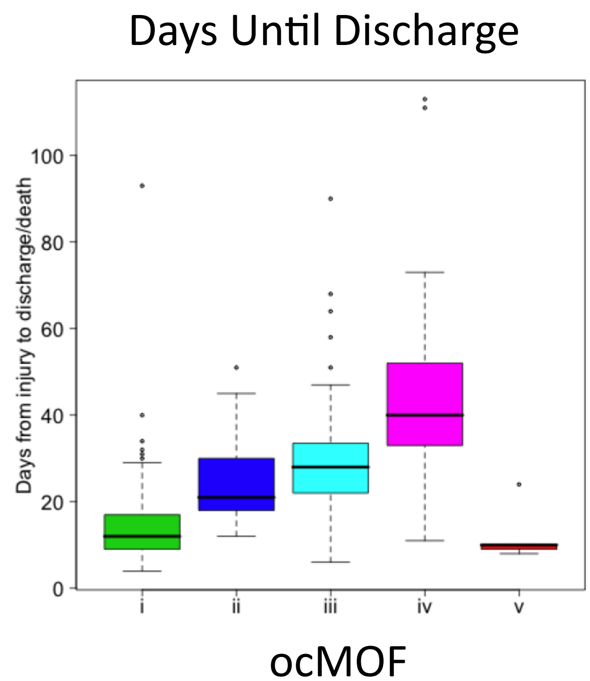


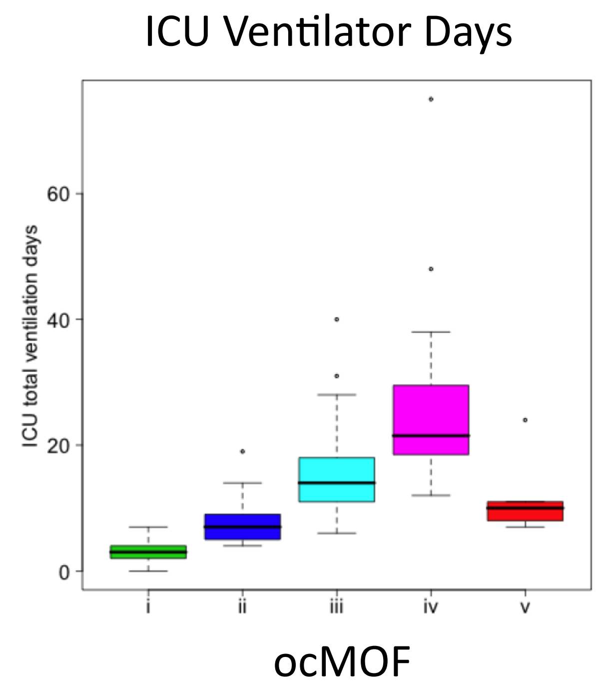


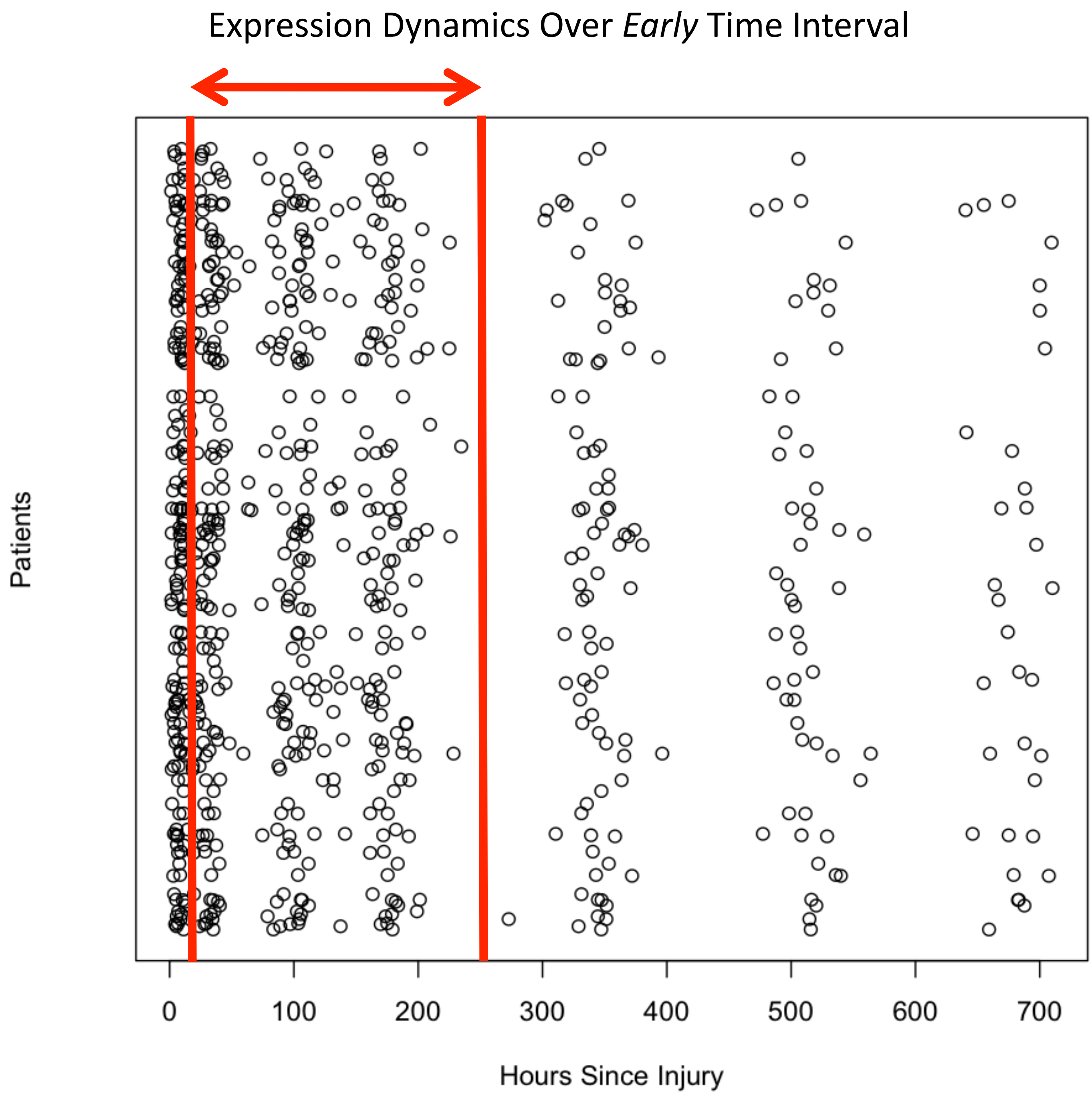
ocMOF i
ocMOF ii
ocMOF iii
ocMOF iv
ocMOF v
0
5
10
15
0
10
20
Days since injury


ocMOF i
ocMOF ii
ocMOF iii
ocMOF iv
ocMOF v
Mean trajectory
Marshall Score
Mean Marshall
Days since injury
0
5
10
15
20

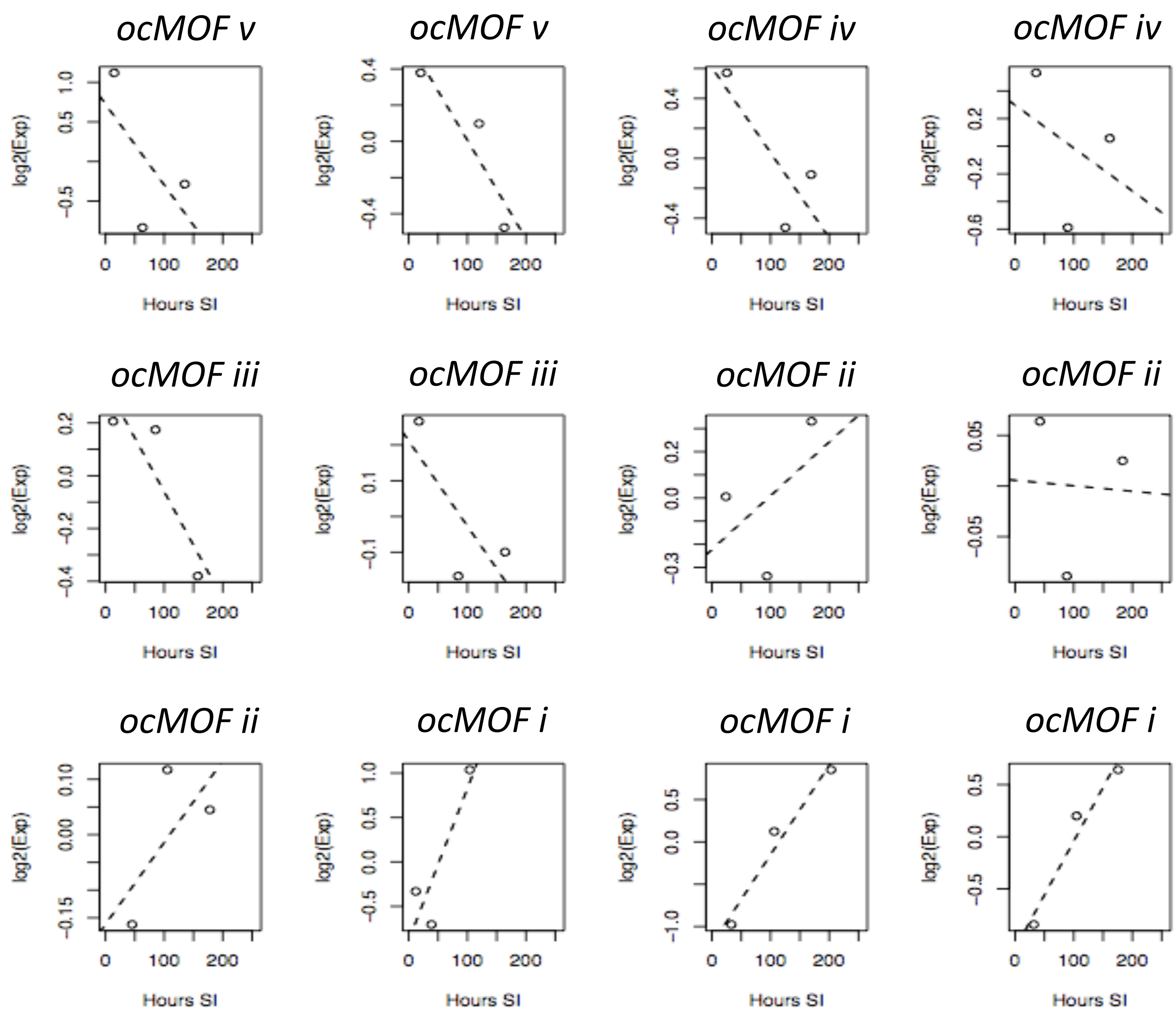

Days in ICU
Days in ICU
0
20
40
60
80
i
ii
iii
iv
v
ocMOF

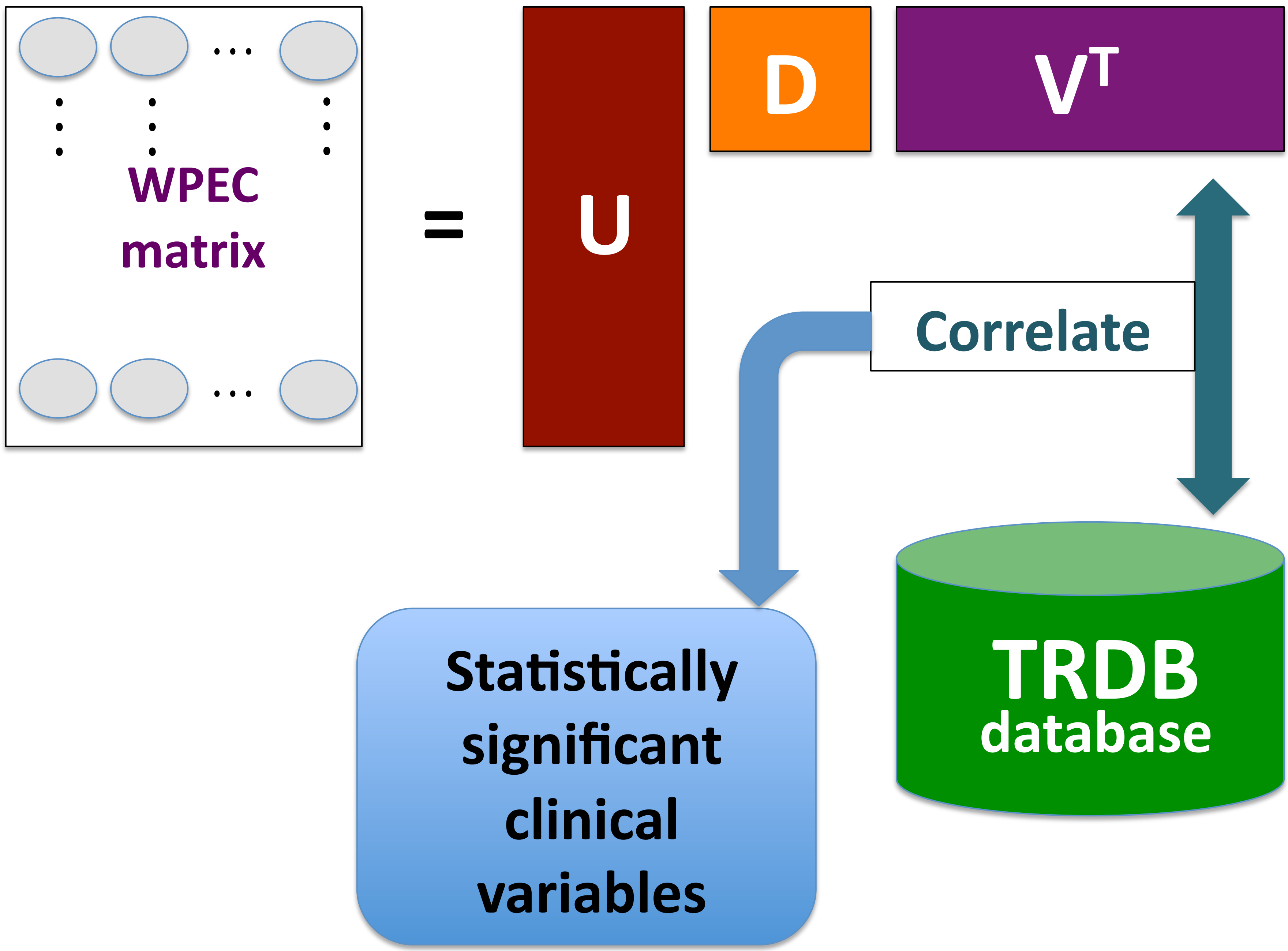

Days Until Discharge
Days from injury to discharge/death
0
20
40
60
80
100
i
ii
iii
iv
v
ocMOF

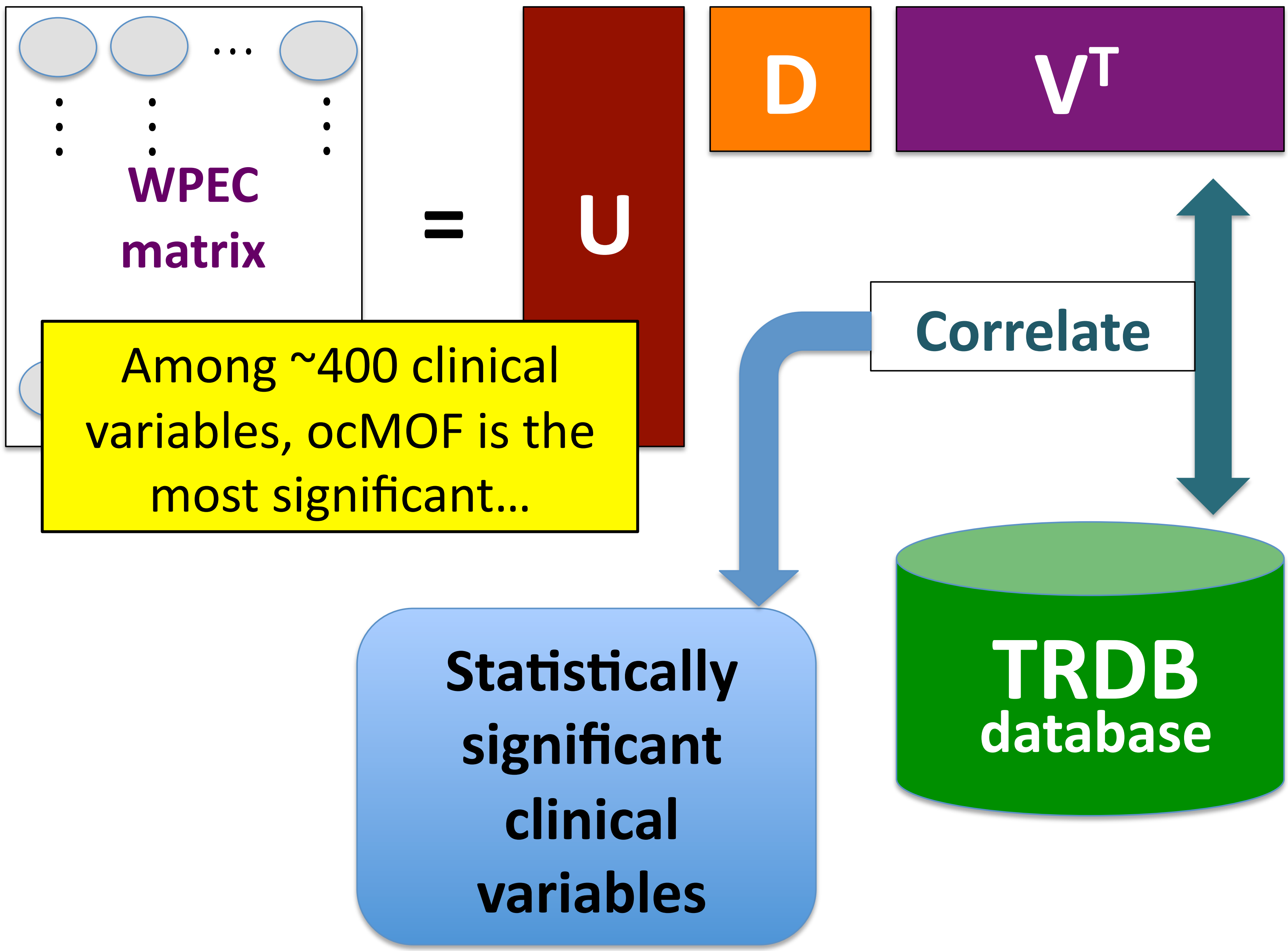

ICU Ventilator Days
ICU total ventilation days
0
20
40
60
i
ii
iii
iv
v
ocMOF

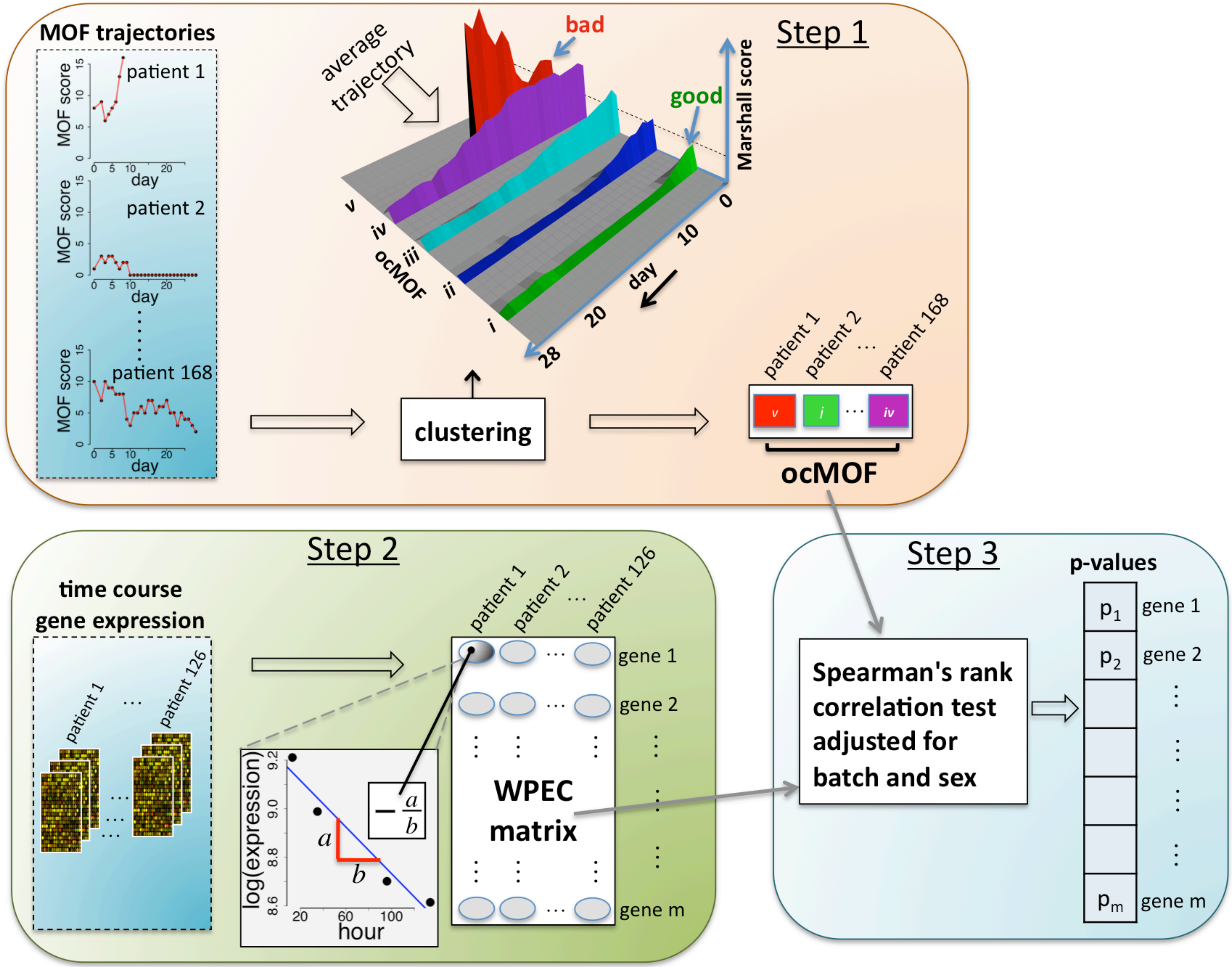

Expression Dynamics Over *Early* Time Interval
Patients
0
100
200
300
400
500
600
700
Hours Since Injury

Within-Patient Expression Change
WPEC

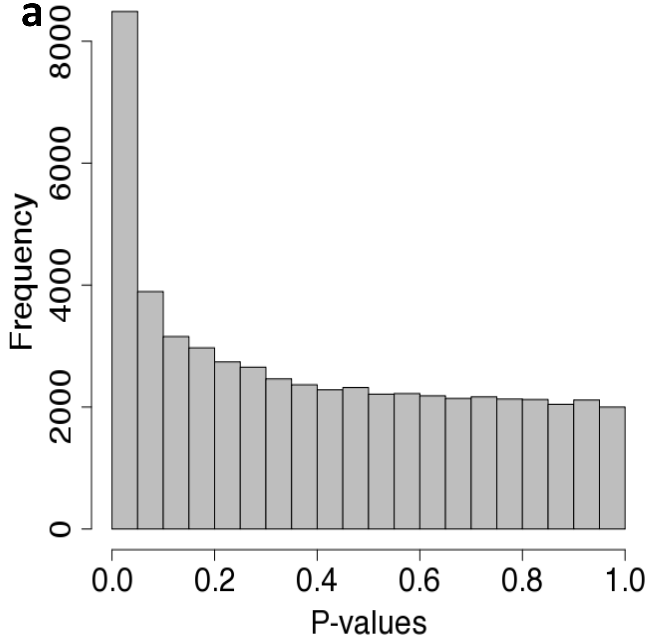



WPEC
matrix
=
U
D
V^T
Correlate
Statistically significant clinical variables
TRDB
database


WPEC
matrix
=
U
D
V^T
Among ~400 clinical variables, ocMOF is the most significant…
Correlate
Statistically significant clinical variables
TRDB
database


Step 1
MOF trajectories
patient 1
patient 2
patient 168
MOF score
day
average trajectory
bad
good
Marshall score
ocMOF
day
clustering
patient 1
patient 2
...
patient 168
v
i
iv
ocMOF
Step 2
time course gene expression
patient 1
patient 126
patient 1
patient 2
patient 126
gene 1
gene 2
gene m
log(expression)
hour
$-\frac{a}{b}$
WPEC matrix
Step 3
Spearman's rank correlation test adjusted for batch and sex
p-values
p_1 gene 1
p_2 gene 2
p_m gene m

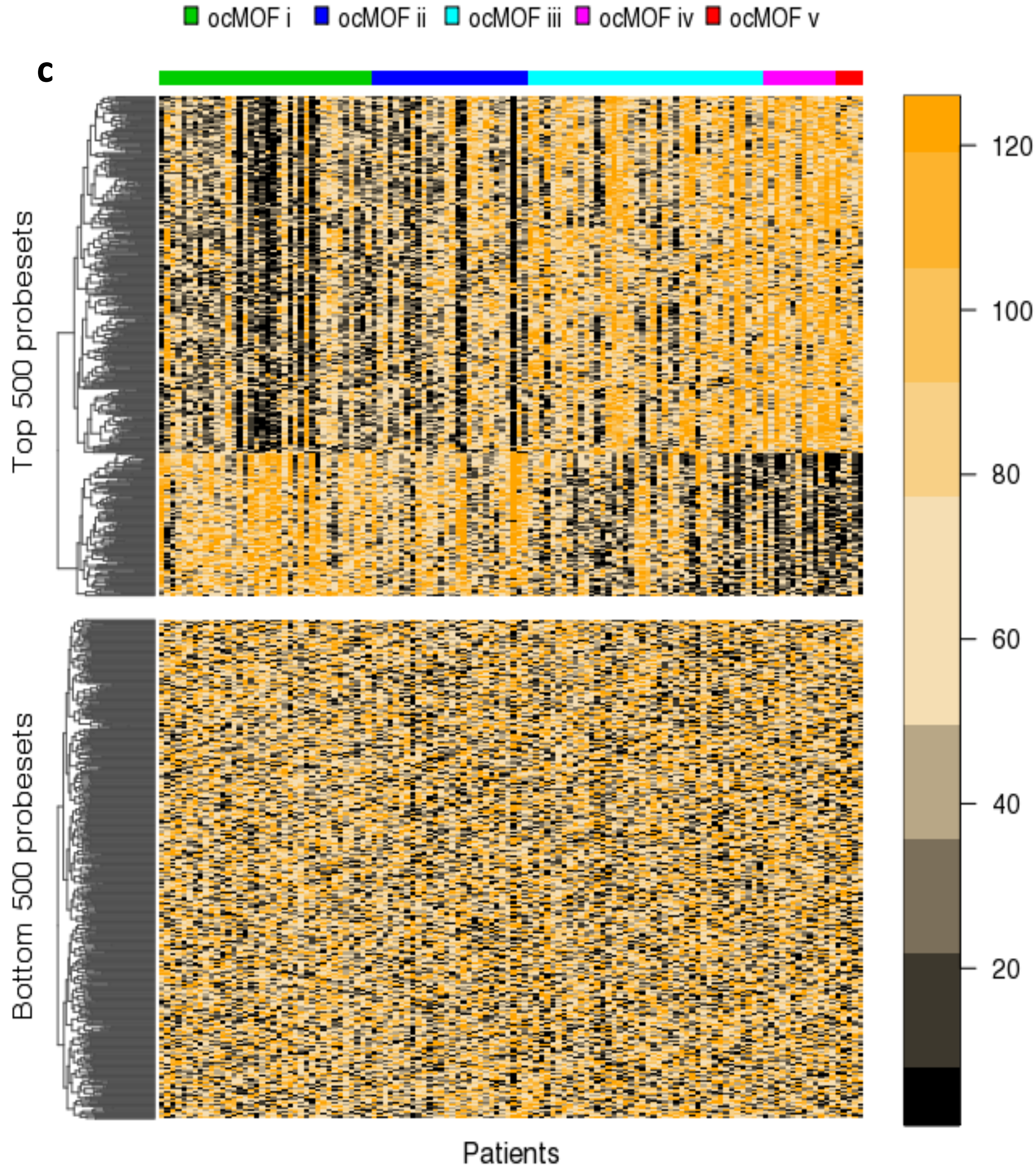
a



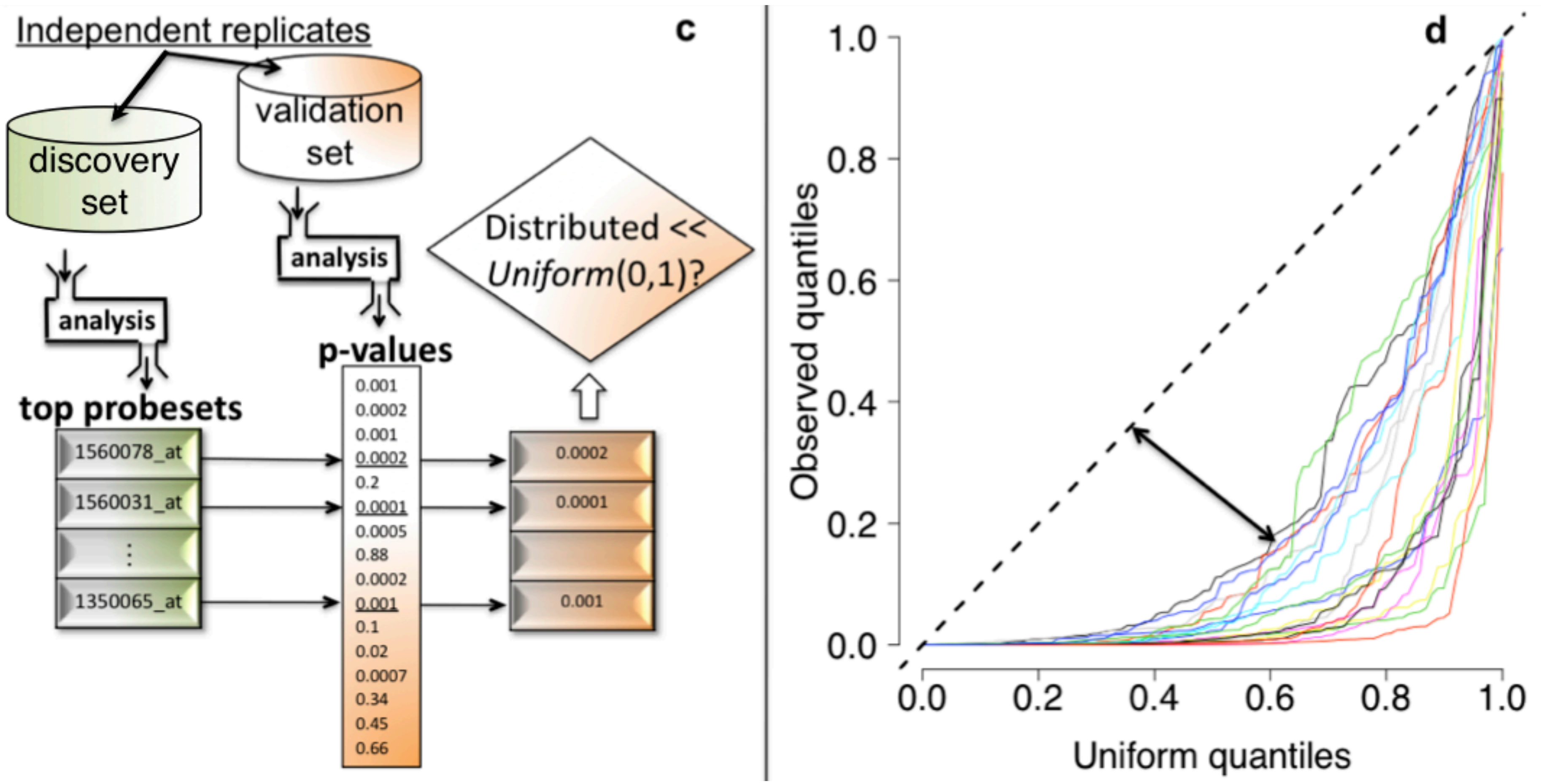
b

FDR	Number of significant probesets
1%	671
5%	2061
10%	3661

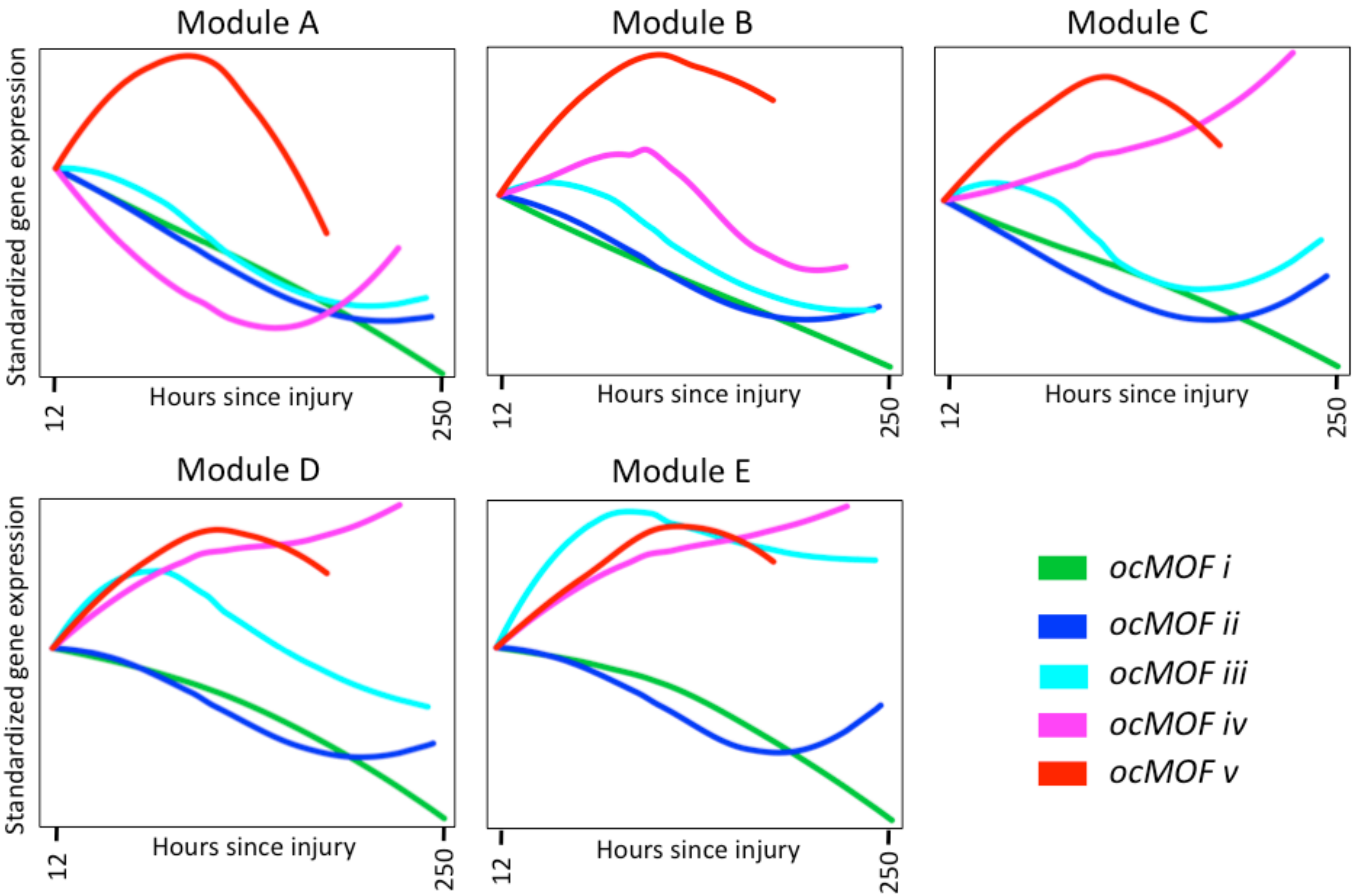
c



Verifying Reproducibility of Associations



Co-expression Modules



Significant Pathways

- p38 MAPK signaling
- Antigen presentation pathway
- Toll-like receptor signaling
- Interferon signaling
- Interleukin-6 signaling

Drug Targets

Monocyte HLA-DR and Interferon-Gamma Treatment in Severely Injured Patients—A Critical Reappraisal More Than a Decade Later

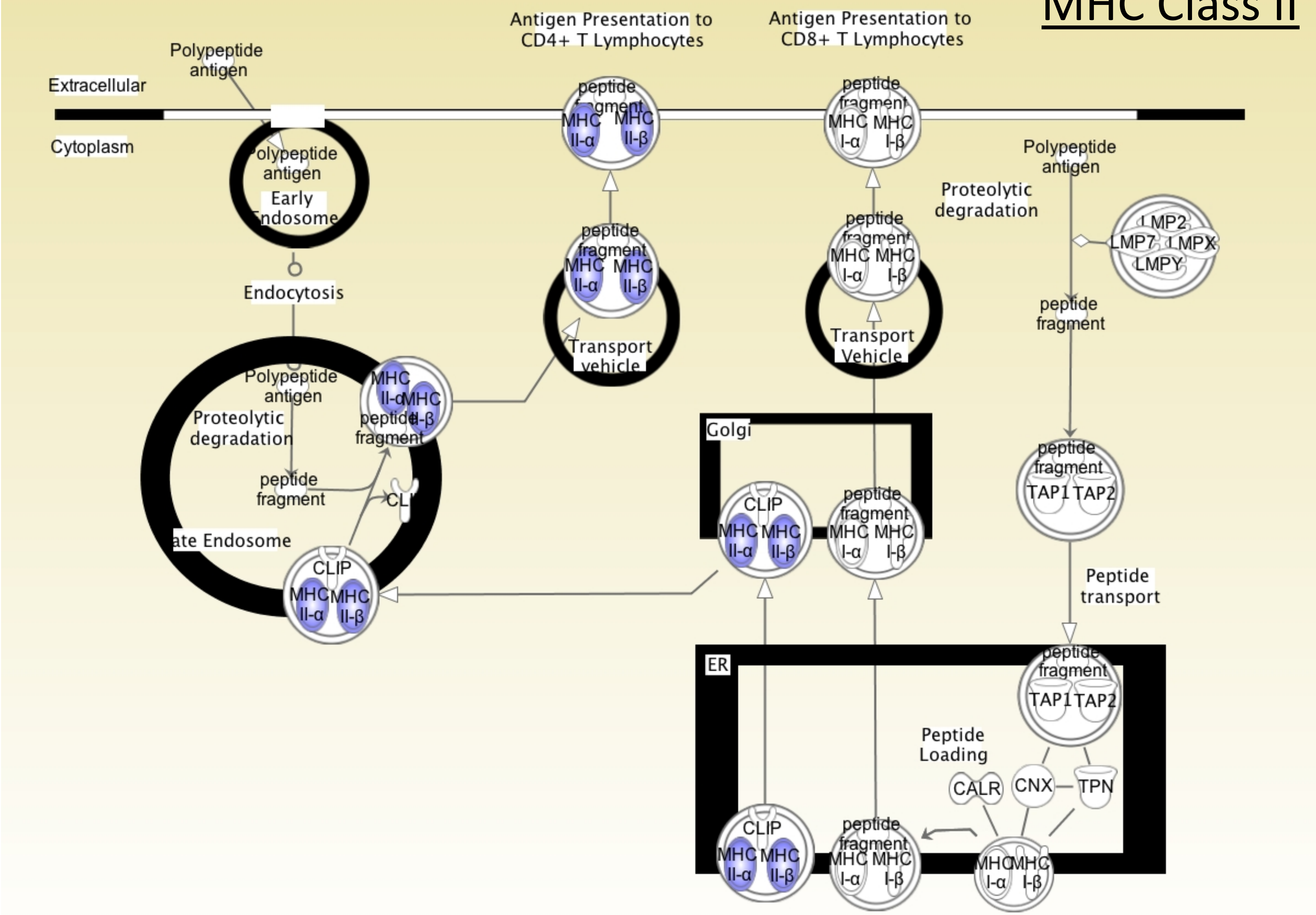
Matthias Turina, MD, Ashley Dickinson, Sarah Gardner, BS, Hiram C Polk Jr, MD, FACS

Nature Reviews Drug Discovery **2**, 717-726 (September 2003) | doi:10.1038/nrd1177

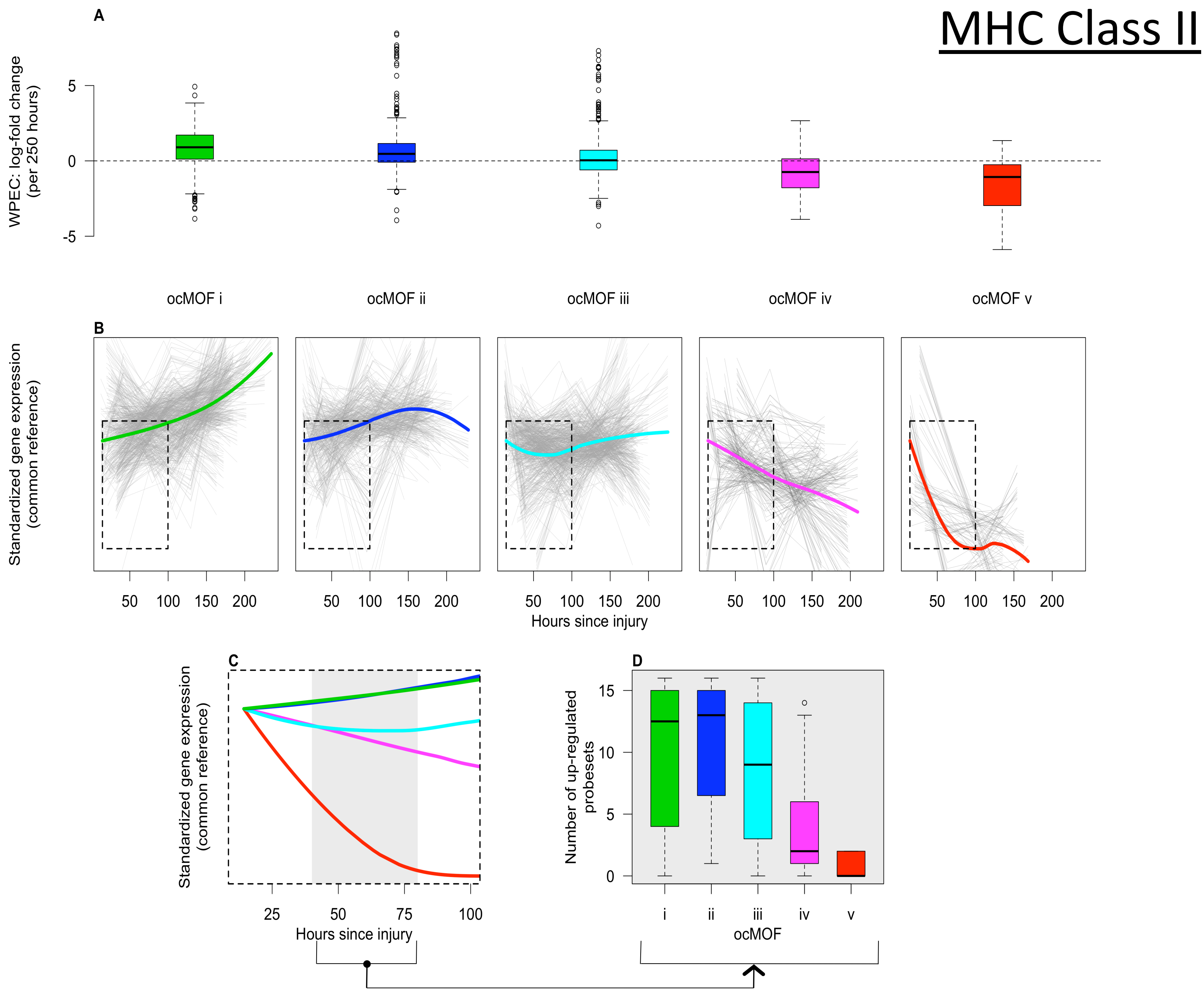
p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases

Sanjay Kumar[1], Jeffrey Boehm[1] & John C. Lee[1] About the authors

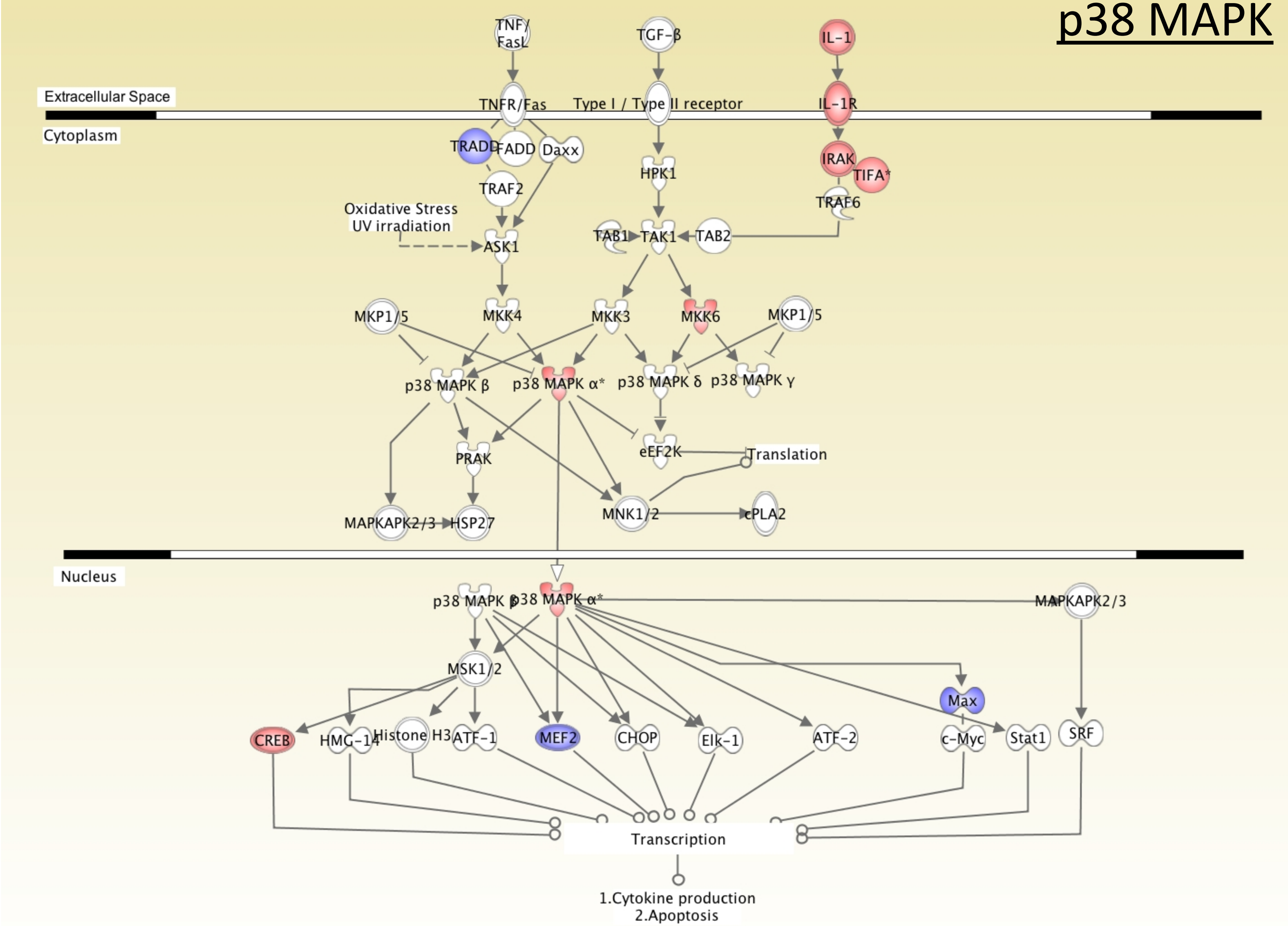
MHC Class II



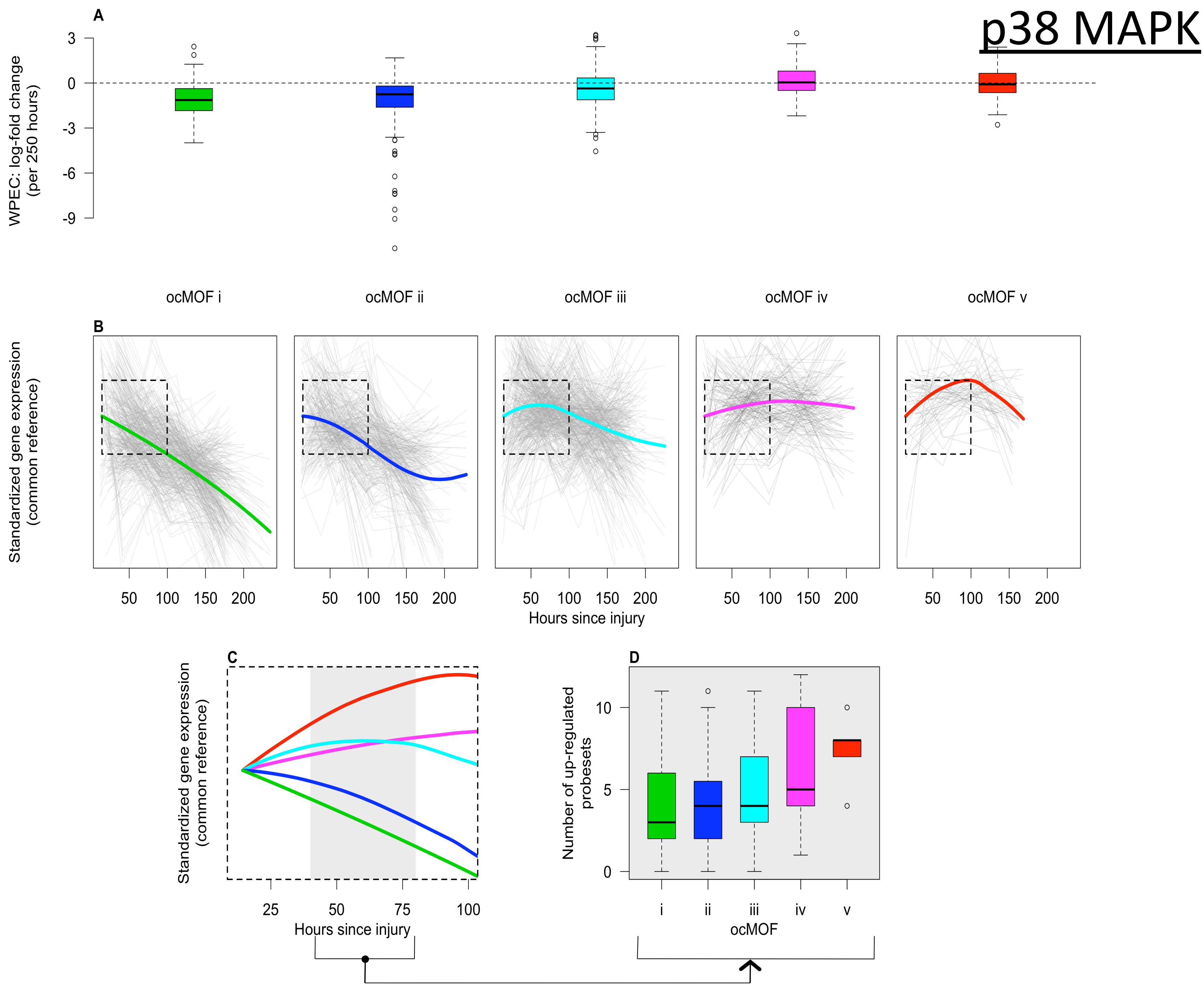
Antigen Presentation to CD4+ T Lymphocytes
Antigen Presentation to CD8+ T Lymphocytes
Polypeptide antigen
Extracellular
Cytoplasm
Early Endosome
Endocytosis
Transport vehicle
Transport Vehicle
Proteolytic degradation
peptide fragment
CLIP
MHC II-α
MHC II-β
MHC I-α
MHC I-β
LMP2
LMP7
LMPX
LMPY
TAP1
TAP2
Peptide transport
Golgi
ER
Peptide Loading
CALR
CNX
TPN


MHC Class II
A
WPEC: log-fold change (per 250 hours)
ocMOF i
ocMOF ii
ocMOF iii
ocMOF iv
ocMOF v
B
Standardized gene expression (common reference)
Hours since injury
C
Standardized gene expression (common reference)
Hours since injury
D
Number of up-regulated probesets
i
ii
iii
iv
v
ocMOF

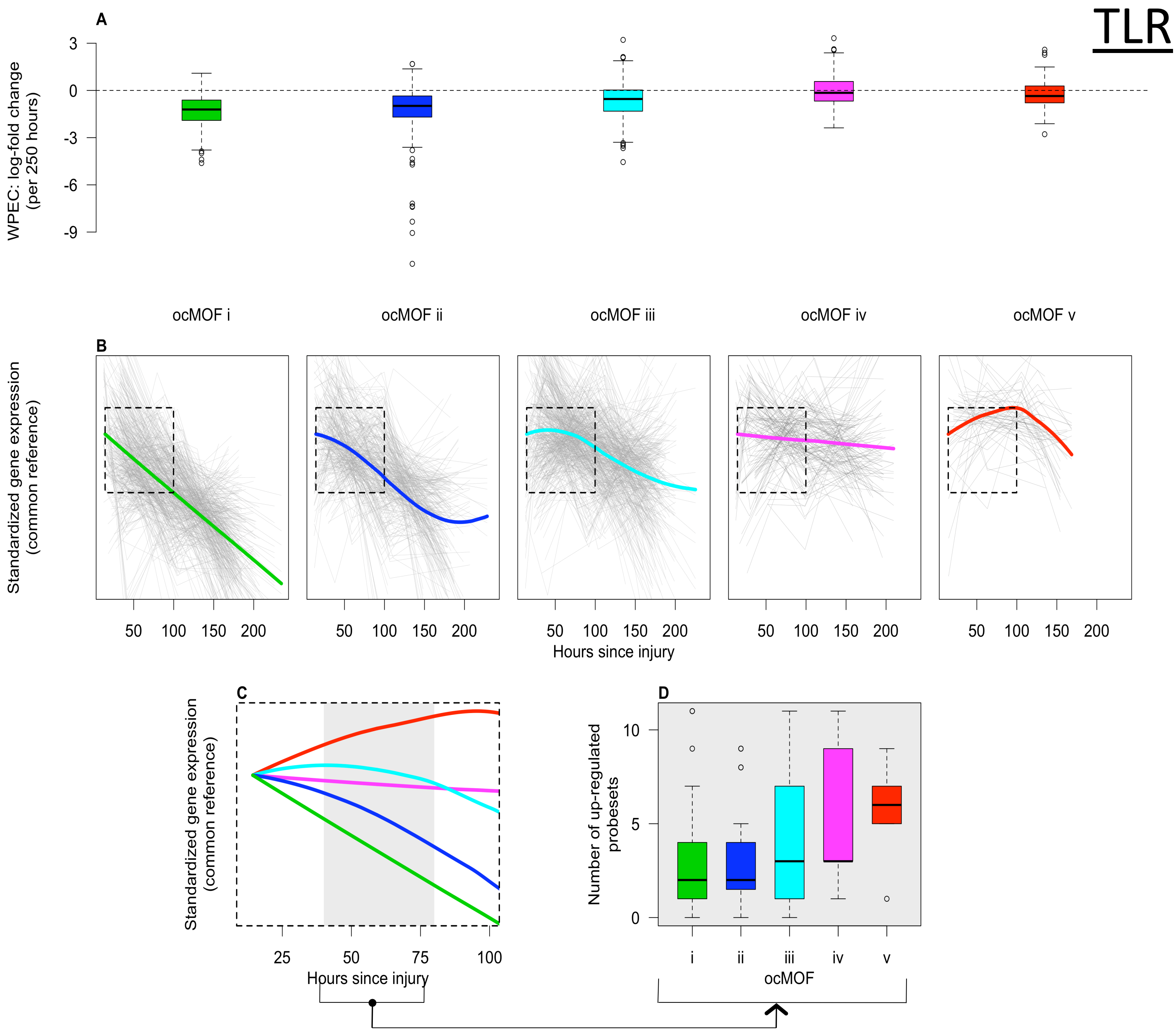
p38 MAPK

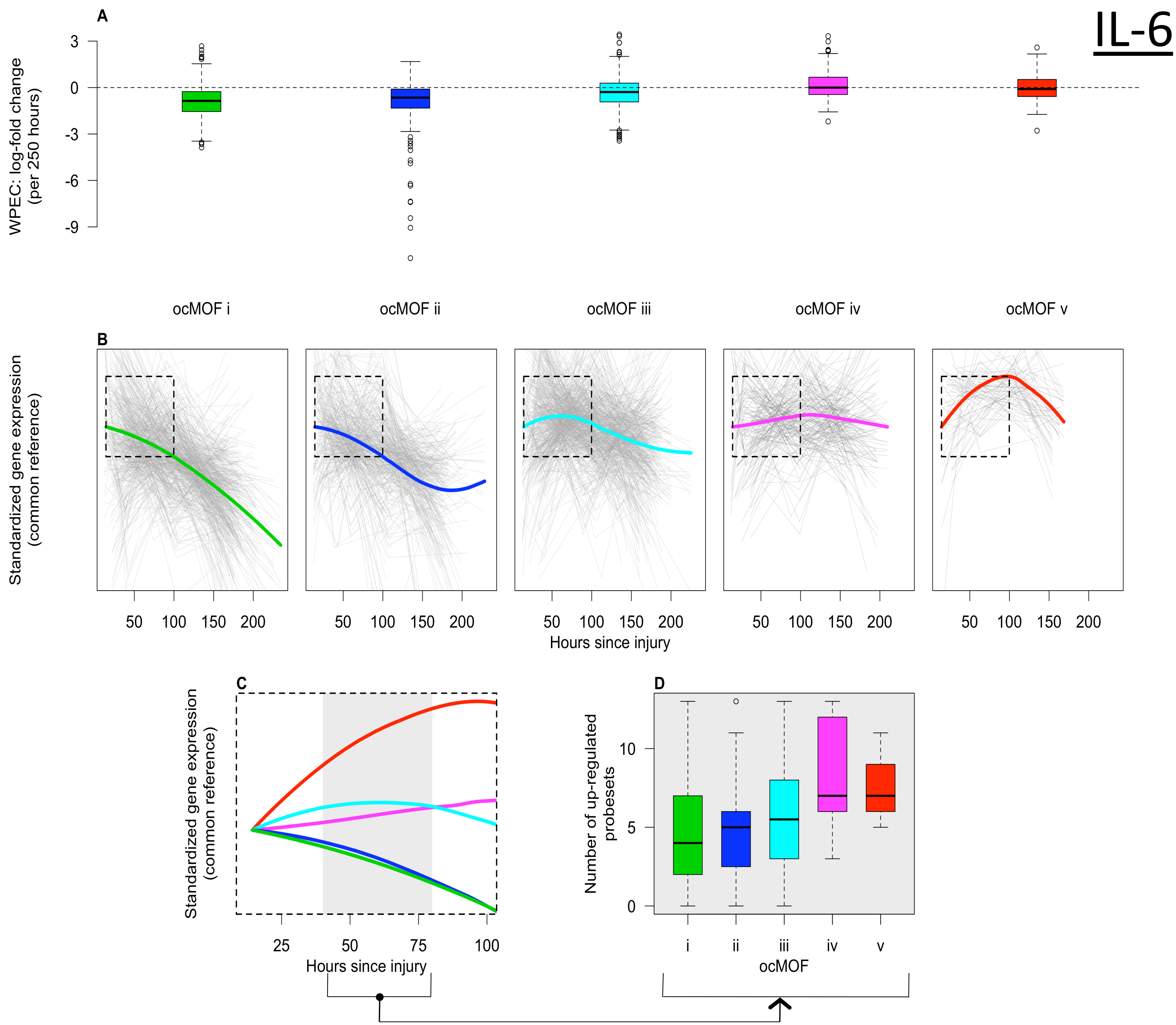


TNF/FasL
TGF-β
IL-1
Extracellular Space
TNFR/Fas
Type I / Type II receptor
IL-1R
Cytoplasm
TRADD
FADD
Daxx
HPK1
IRAK
TIFA*
TRAF2
TRAF6
Oxidative Stress
UV irradiation
ASK1
TAB1
TAK1
TAB2
MKP1/5
MKK4
MKK3
MKK6
MKP1/5
p38 MAPK β
p38 MAPK α*
p38 MAPK δ
p38 MAPK γ
PRAK
eEF2K
Translation
MAPKAPK2/3
HSP27
MNK1/2
cPLA2
Nucleus
p38 MAPK β
p38 MAPK α*
MAPKAPK2/3
MSK1/2
Max
CREB
HMG-14
Histone H3
ATF-1
MEF2
CHOP
Elk-1
ATF-2
c-Myc
Stat1
SRF
Transcription
1.Cytokine production
2.Apoptosis



p38 MAPK
A
WPEC: log-fold change
(per 250 hours)
3
0
-3
-6
-9
ocMOF i
ocMOF ii
ocMOF iii
ocMOF iv
ocMOF v
B
Standardized gene expression
(common reference)
50
100
150
200
Hours since injury
C
Standardized gene expression
(common reference)
25
50
75
100
Hours since injury
D
Number of up-regulated
probesets
10
5
0
i
ii
iii
iv
v
ocMOF


TLR
A
WPEC: log-fold change
(per 250 hours)
3
0
-3
-6
-9
ocMOF i
ocMOF ii
ocMOF iii
ocMOF iv
ocMOF v
B
Standardized gene expression
(common reference)
50
100
150
200
Hours since injury
C
Standardized gene expression
(common reference)
25
50
75
100
Hours since injury
D
Number of up-regulated
probesets
10
5
0
i
ii
iii
iv
v
ocMOF



IL-6
A
WPEC: log-fold change
(per 250 hours)
3
0
-3
-6
-9
ocMOF i
ocMOF ii
ocMOF iii
ocMOF iv
ocMOF v
B
Standardized gene expression
(common reference)
50
100
150
200
Hours since injury
C
Standardized gene expression
(common reference)
25
50
75
100
Hours since injury
D
Number of up-regulated
probesets
10
5
0
i
ii
iii
iv
v
ocMOF

Summary

- Simplification of complex longitudinal phenotype
- Capture expression dynamics within patients – robust and translational
- Reproducible clinical associations
- Biologically meaningful within-patient changes in gene expression explain outcome

To appear, *PLoS Medicine*

Acknowledgements

Joint work with:

Keyur Desai

Chuen-Seng Tan

Jeffrey T. Leek

Ronald Maier

Ronald Tompkins

Glue grant members

Funding for this project:

U54 GM62119 (PI: Ron Tompkins)

R01 HG002913

gluegrant.org