Inferring differential leukocyte activity from antibody microarrays using a latent variable model

Joshua Ho[1,2]

[1] School of IT, The University of Sydney
[2] NICTA, Australian Technology Park, NSW, Australia
joshua@it.usyd.edu.au

Joint work with
R. Koundinya, T.S. Caetano,
C.G. dos Remedios, and M.A. Charleston

Human Immune System

- Many types of leukocytes (white blood cells), each have a different immune function:
  - B-cells
  - T-cells
  - NK-cells
  - Monocytes … and many more
- Distribution of leukocytes differs in different diseases
- Each type of leukocyte is characterized by a set of cell surface molecules called CD antigens

NICTA

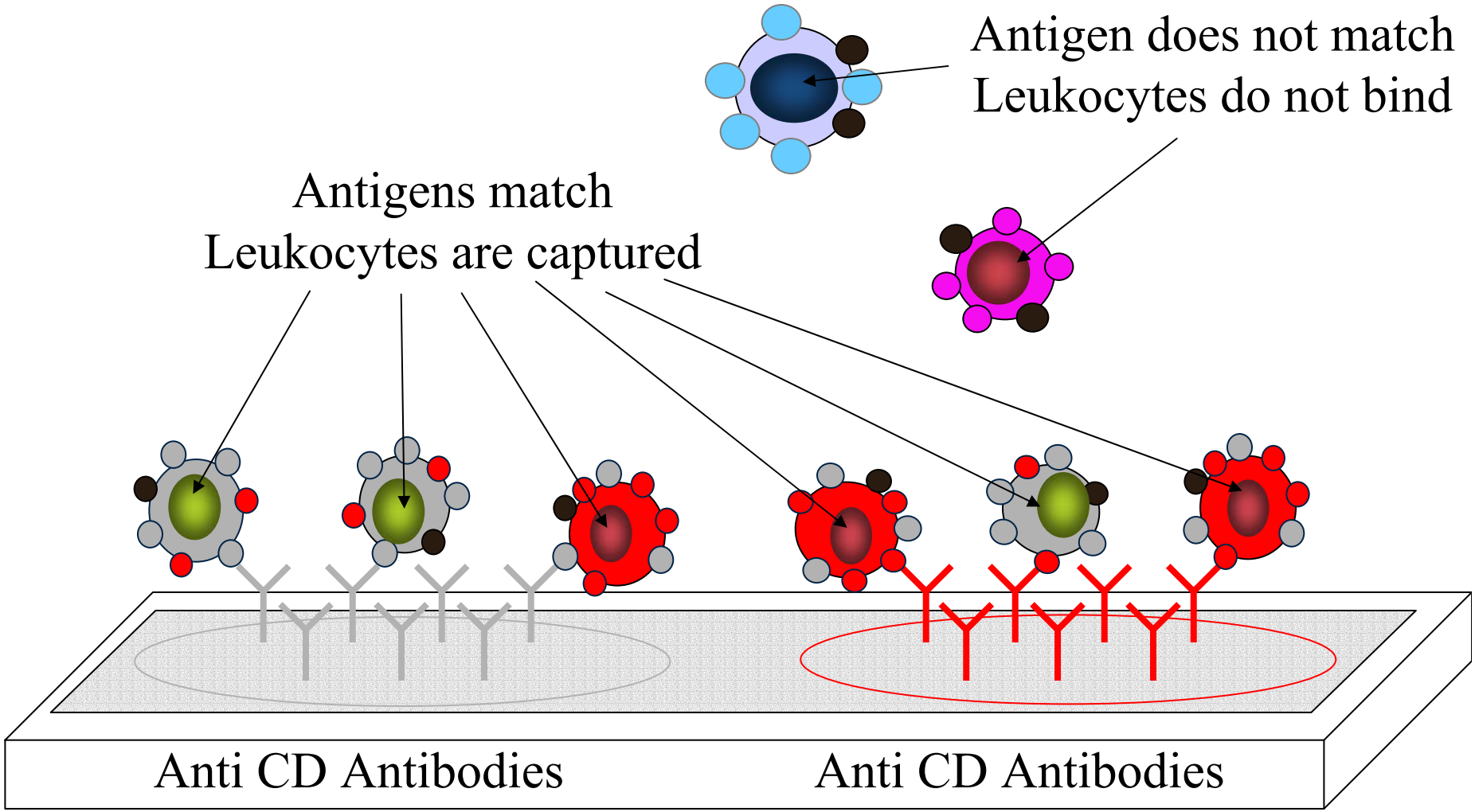
CD antigens

Leukocyte	CD antigens[a]
T cell (T)	TCR a/b TCR g/d CD1a CD2 CD3 CD4 CD5 CD7 CD8 CD9 CD11a CD11b CD11c CD16 CD25 CD28 CD29 CD31 CD37 CD38 CD43 CD44 CD45 CD45RA CD49d CD49e CD52 CD54 CD56 CD57 CD60 CD62L CD80 CD86 CD95 CD102 CD103 CD120a CD122 CD126 CD128 CD130 CD134 CD154
B cell (B)	CD1a CD2 CD5 CD9 CD11a CD11b CD11c CD19 CD20 CD21 CD22 CD23 CD24 CD25 CD29 CD31 CD32 CD37 CD38 CD40 CD44 CD45 CD45RA CD45RO CD49d CD52 CD54 CD62L CD77 CD79a CD79b CD80 CD86 CD95 CD102 CD120a CD122 CD126 CD130 CD138 HLA-DR l FMC7 k
Monocyte (M)	CD1a CD4 CD9 CD11a CD11b CD11c CD13 CD14 CD15 CD16 CD29 CD31 CD32 CD33 CD36 CD37 CD38 CD40 CD43 CD44 CD45 CD45RA CD45RO CD49d CD49e CD52 CD54 CD60 CD61 CD62L CD64 CD65 CD80 CD86 CD88 CD95 CD102 CD120a CD122 CD126 CD128 CD130 HLA-DR
Natural Killer (NK)	CD2 CD7 CD8 CD11a CD11b CD11c CD16 CD25 CD29 CD31 CD38 CD43 CD44 CD45 CD45RA CD45RO CD49d CD49e CD52 CD56 CD57 CD62L CD95 CD102 CD120a CD122 CD128 CD130
Others	CD10 CD34 CD41 CD42a CD62E CD62P CD66c CD71 CD117 CD135 CD235a

Note: [a]These relationships were extracted from the official poster of the Eight International Workshop on Human Leukocyte Differentiation Antigens.

NICTA

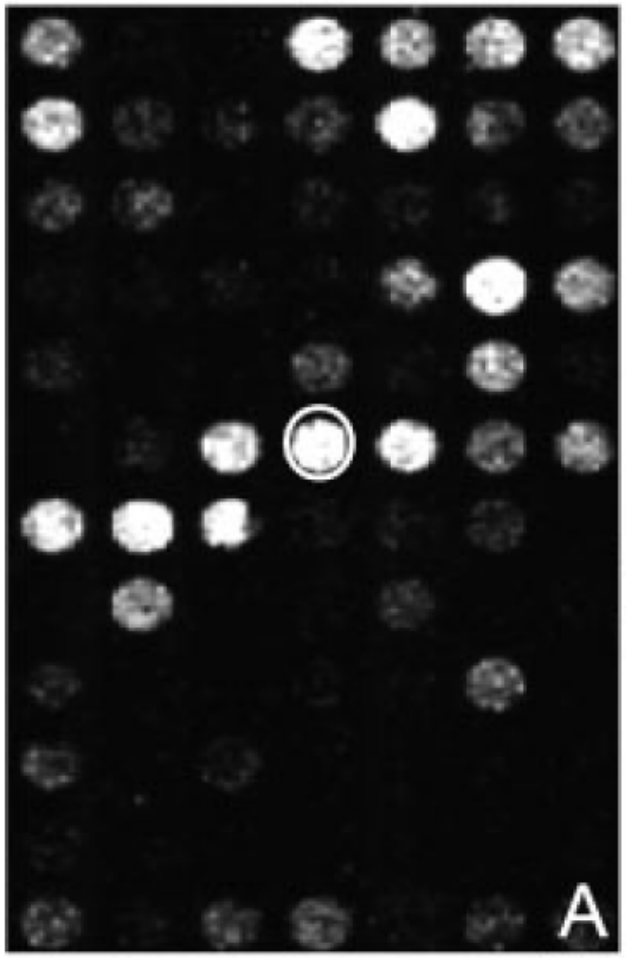
Antibody Microarray

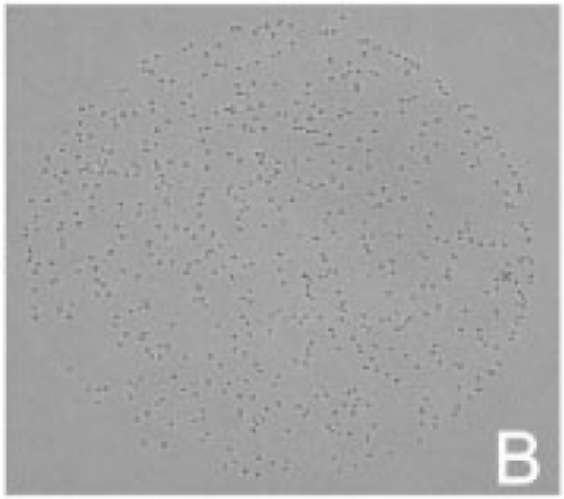


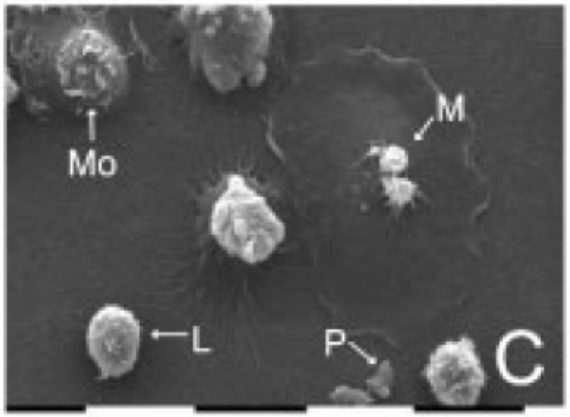
Courtesy of A. Brown

NICTA

Antibody Microarray







D

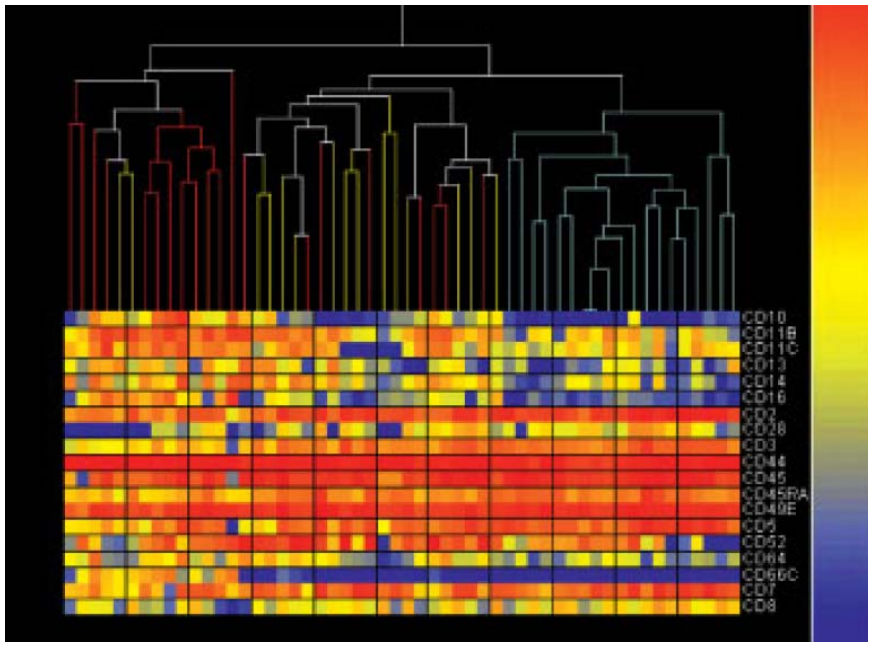
TCR α/β	TCR γ/δ	1a	2	3	4	5
7	8	9	10	11a	11b	11c
13	14	15	16	19	20	21
22	23	24	25	28	29	31
32	33	34	36	37	38	40
41	42a	43	44	45	45RA	45RO
49d	49e	52	54	56	57	60
61	62L	62E	62P	64	65	66c
71	77	79a	79b	80	86	88
95	102	103	117	120a	122	126
128	130	134	135	138	154	235a
HLA-DR	FMC7	κ	λ	sIg	TEST	TEST

Courtesy of A. Brown

NICTA

Current Analysis Approach

- Visualization
- Clustering
- Finding DE genes
- Train classifier

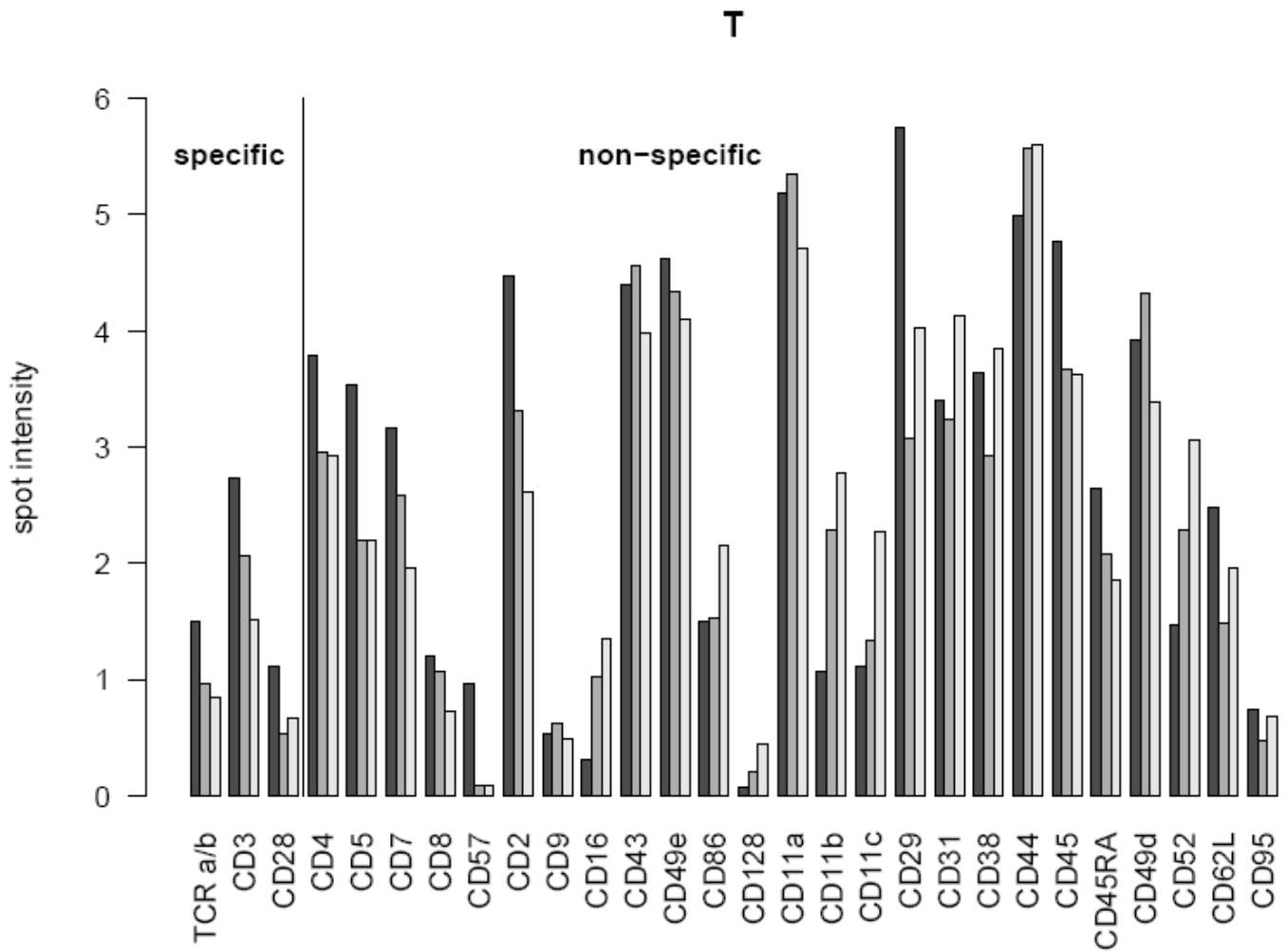


- But how can we infer changes at the leukocyte level?

Courtesy of A. Brown

NICTA

How difficult is antibody microarray analysis?



NICTA

How to infer leukocyte activity?

CD antigen	Description	Cellular expression	Healthy Mean ± SD	SAP Mean ± SD (p-value)	UAP Mean ± SD (p-value)	*p-Value (SAP-UAP)*	References
CD2	T cell adhesion and activation	T/NK	4.5 ± 0.9	3.4 ± 1.0 ($p = 0.04$)	2.6 ± 0.8 ($p<0.001$)*	0.01	Novel
CD3	Forms part of the T cell receptor complex	T	2.7 ± 0.5	NS	1.5 ± 0.6 ($p<0.001$)*	0.01	[36]
CD5	Binds CD166, costimulation	T/B	3.5 ± 0.9	2.2 ± 1.0 ($p = 0.03$)	2.2 ± ($p = 0.002$)	NS	Novel
CD7	T cell costimulation	T/NK	3.2 ± 0.9	NS	2.0 ± 1.0 ($p = 0.002$)	NS	Novel
CD8	Expressed on cytotoxic T cells	T/NK	1.3 ± 0.4	NS	0.7 ± 0.4 ($p = 0.002$)*	0.01	[37]
CD10	Zinc metalloproteinase	G	0.4 ± 0.2	NS	1.4 ± 1.1 ($p<0.001$)	0.01	Novel
CD11b	Binds CD54(ICAM-1) and iC3b	Mo/NK/G	1.1 ± 0.6	2.2 ± 1.0 ($p = 0.02$)	2.8 ± 0.9 ($p<0.001$)	NS	[31]
CD11c	Binds CD54(ICAM-1), fibrinogen and iC3b	Mo/NK/G	1.1 ± 0.4	NS	2.3 ± 0.7 ($p<0.001$)*	0.01	[32]
CD13	Zinc metalloproteinase	Mo/G	0.5 ± 0.4	1.1 ± 0.7 ($p = 0.05$)	1.1 ± 0.7 ($p = 0.01$)	NS	Novel
CD14	Lipopolysaccharide receptor	Mo	0.6 ± 0.4	1.2 ± 0.6 ($p = 0.03$)	1.9 ± 0.9 ($p<0.001$)*	0.02	[30, 40]
CD16	Fc receptor	Mo/NK/G	0.3 ± 0.2	1.0 ± 0.7 ($p = 0.01$)	1.4 ± 1.1 ($p<0.001$)	NS	[40]
CD28	T cell costimulation	T	1.2 ± 0.5	NS	0.7 ± 0.5 ($p = 0.02$)	NS	[44]
CD44	Leukocyte adhesion	T/B/NK/Mo/G	5.0 ± 0.7	5.6 ± 0.5 ($p = 0.05$)	5.6 ± 0.6 ($p = 0.005$)	NS	[34]
CD45	Leukocyte common antigen	T/B/NK/Mo/G	4.7 ± 0.7	NS	3.6 ± 1.5 ($p = 0.01$)	NS	Novel
CD45RA	Expressed on naïve T cells	T	2.7 ± 0.7	NS	1.9 ± 0.8 ($p = 0.002$)	NS	Novel
CD49e	Integrin, forms VLA-5 complex with CD29	T/B/NK/Mo	4.6 ± 0.4	NS	4.1 ± 0.8 ($p = 0.02$)	NS	Novel
CD52	Involved in ADCC	T/B/Mo	1.5 ± 1.1	2.7 ± 1.8 ($p = 0.05$)	3.1 ± 1.6 ($p = 0.002$)	NS	Novel
CD64	Phagocytosis and ADCC	Mo	0.6 ± 0.2	NS	1.3 ± 0.6 ($p<0.001$)	NS	Novel
CD66c	Adhesion	G	0.1 ± 0.1	NS	0.9 ± 0.8 ($p<0.001$)*	0.01	Novel

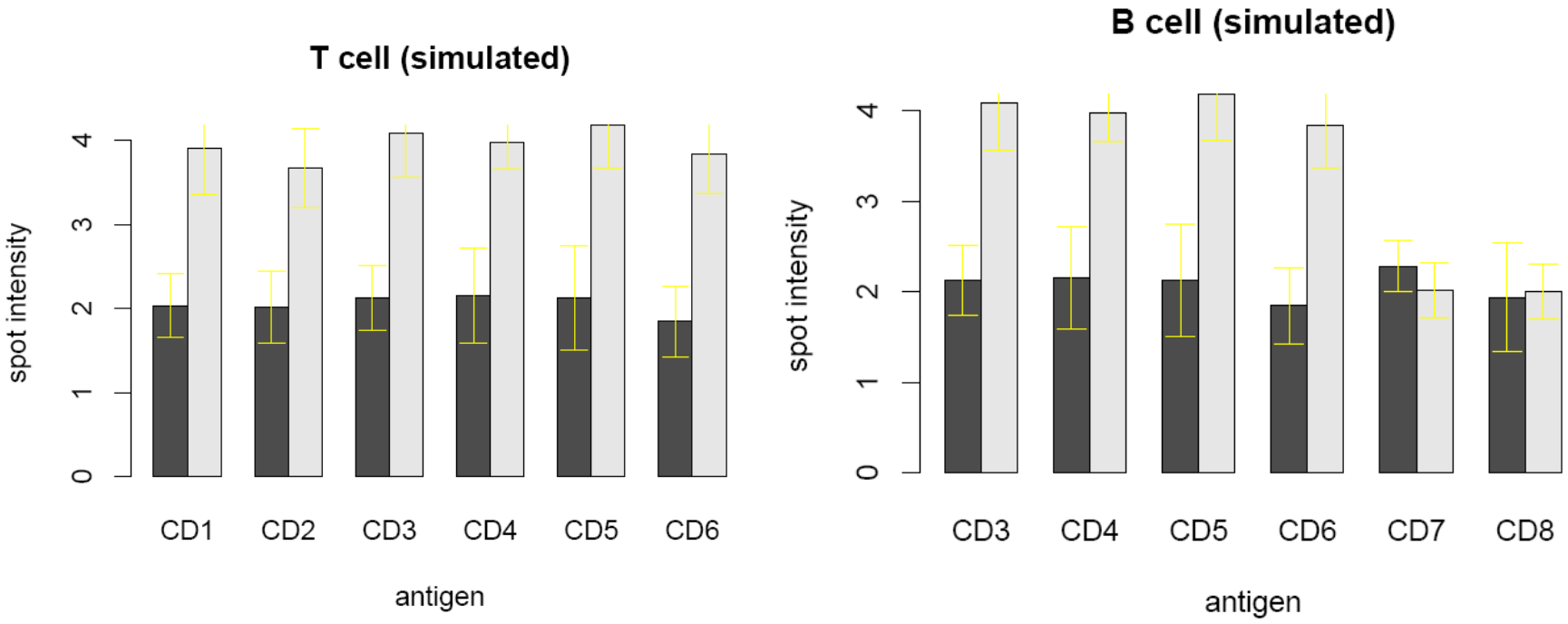
- Conclusion: T decreased, Mo increased

Courtesy of A. Brown

NICTA

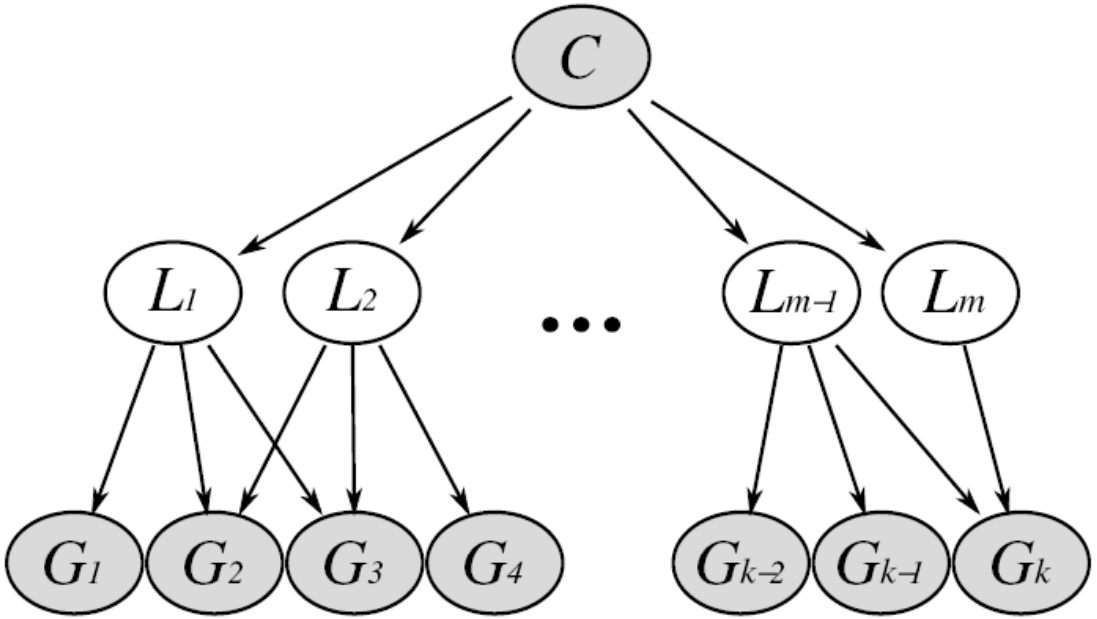
Is GSEA an answer?

- No. Simulation shows GSEA is not satisfactory
- GSEA q-val for enrichment for simulated T and B cells are 0.7 and 0.69 (not sig.)



NICTA

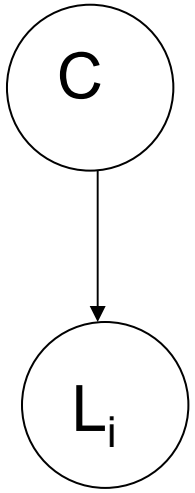
Our approach: latent variable model



- C = class (observed) = {normal, disease1,disease2,…}
- L = leukocyte activity (latent) = {activated, deactivated}
- G = antigen expression (observed) = Gaussian distribution

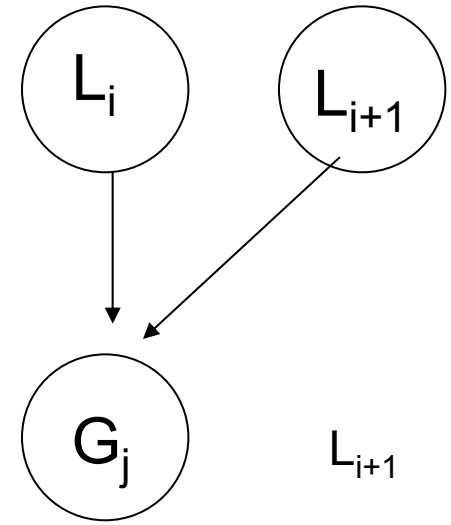
NICTA

Conditional Probability Models

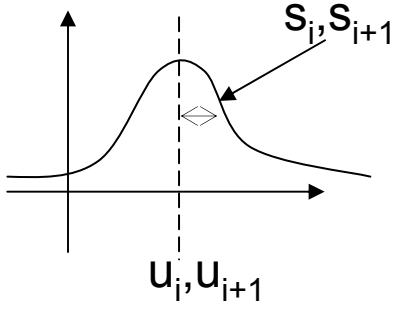


- $P(L_i|C)$ =

	class1	class2
active	0.3	0.8
inactive	0.7	0.2



- $P(G_j|L_i,L_{i+1})$ =

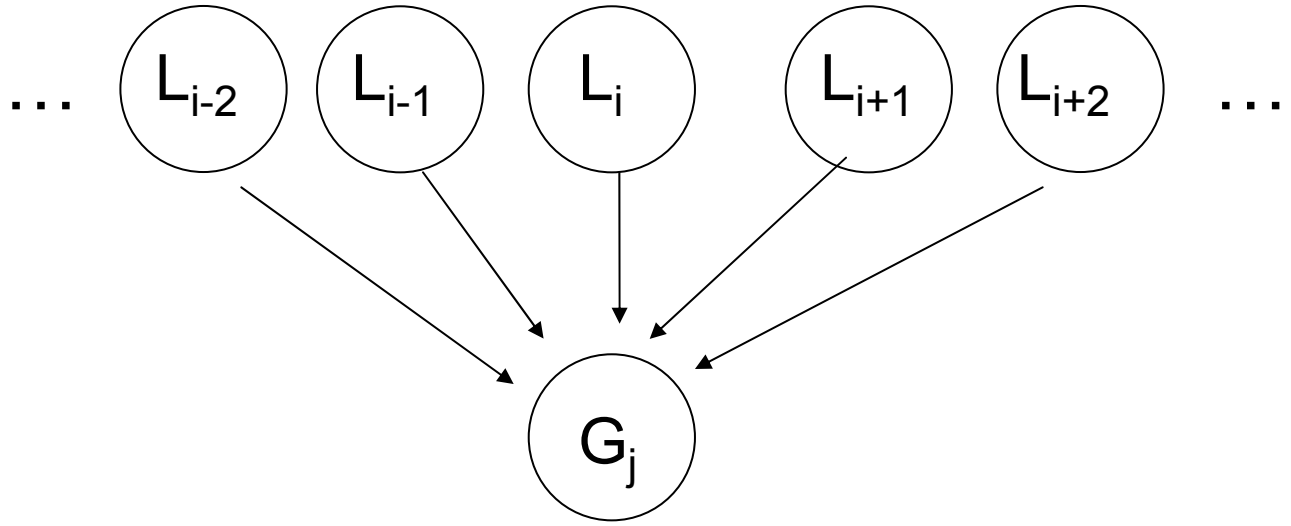


		Li: active	Li: inactive
L_{i+1}	active	$u_{1,1}$, $s_{1,1}$	$u_{0,1}$, $s_{0,1}$
	inactive	$u_{1,0}$, $s_{1,0}$	$u_{0,0}$, $s_{0,0}$

NICTA

Model simplification

- n parents = $2x2^n$ parameters in the model
- Expensive and inaccurate for small sample dataset

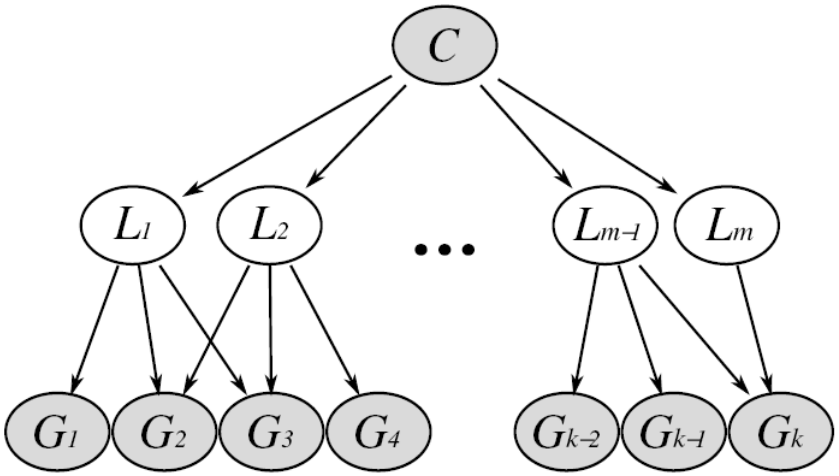


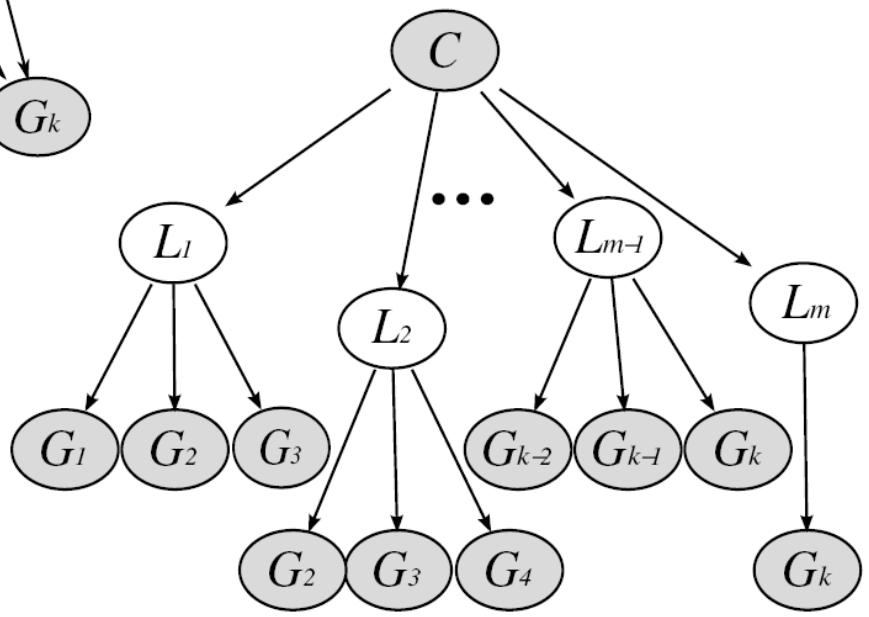
- Solution: decomposition
- $P(G_j|L_i,L_{i+1},\ldots L_{i+m}) = P(G_j|L_i)\ P(G_j|L_{i+1})\ldots P(G_j|L_{i+m})$
- Only 2n parameters

NICTA

Model Simplification

- Problem: data duplication!



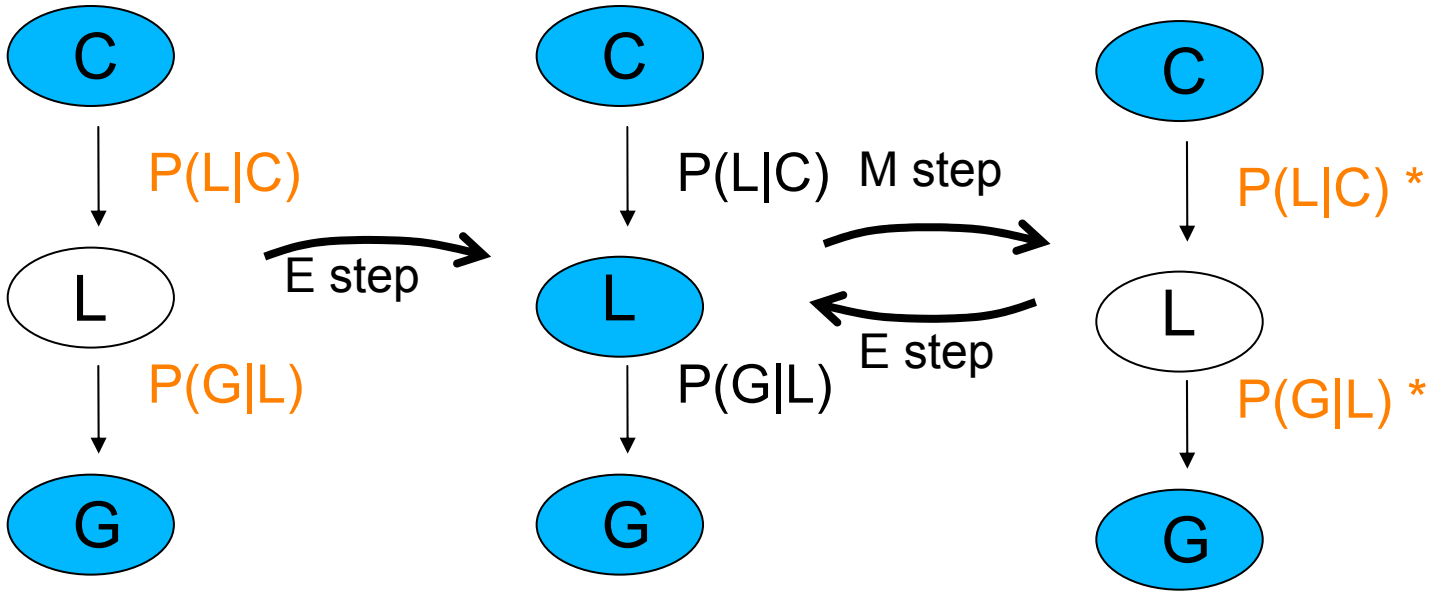


- Solution: variance of each G_i be proportional to the original number of parents

NICTA

LVM Parameter Learning

- Expectation Maximization (EM) algorithm
- E-step: infer expected distribution of the latent variables
- M-Step: find the maximum likelihood estimates of the model parameters
- Repeat E- & M- steps until convergence



NICTA

Model Analysis

$P(L_i|C)$ =

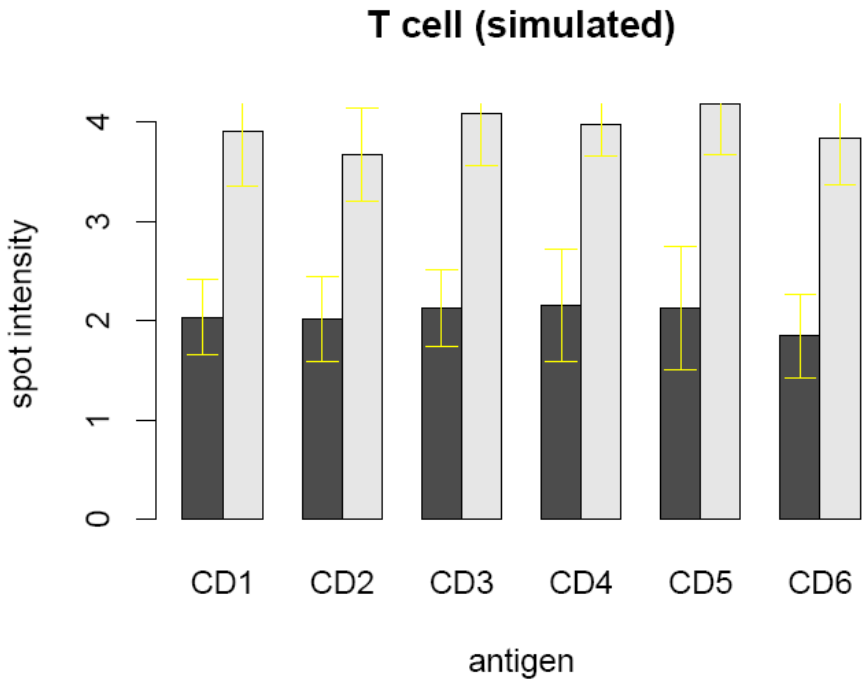
	class1	class2
active	0.3	0.8
inactive	0.7	0.2

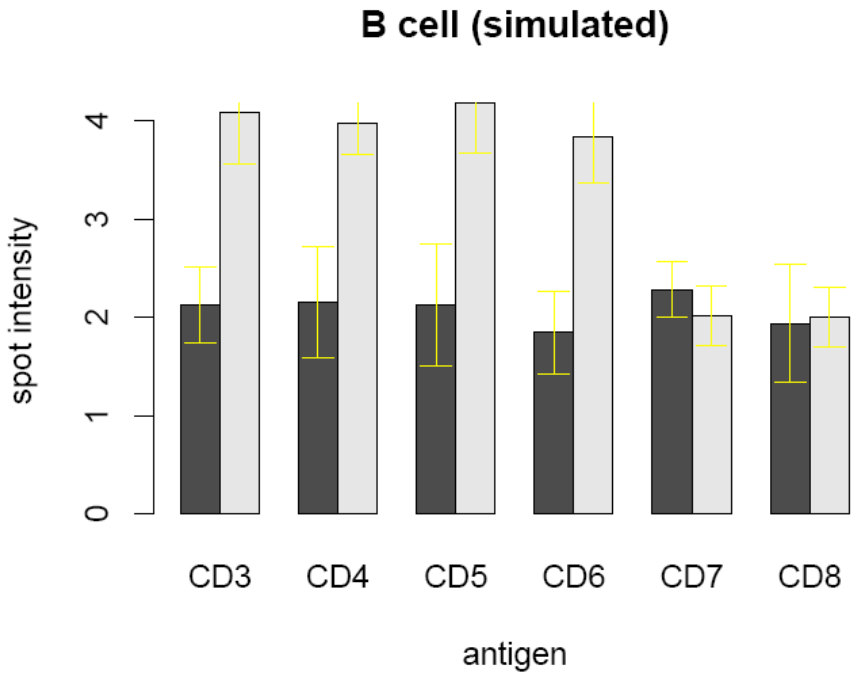
- Question: how to determine differential leukocyte activity from $P(L_i|C)$?
- Solution: Total correlation

$$C_{\text{tot}}(L_i, C) = \sum_{l \in S(L_i)} \sum_{c \in S(C)} p(l,c) \log \left[\frac{p(l,c)}{p(l)p(c)} \right]$$

NICTA

Simulated data

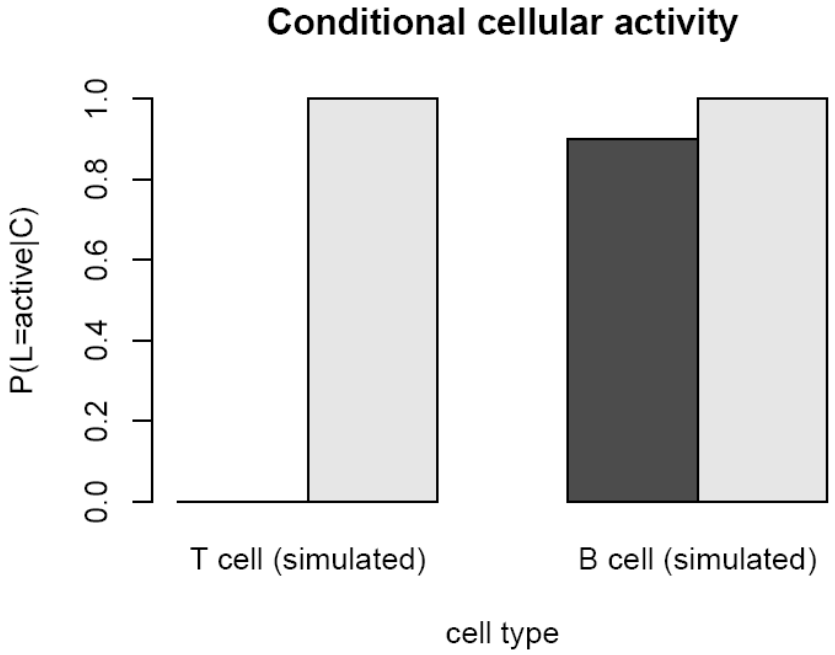


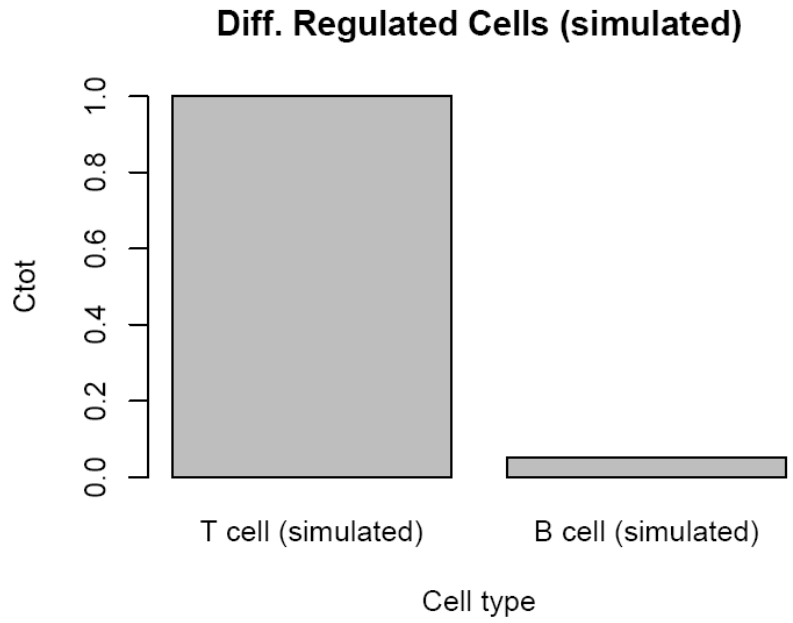


- T cells are differentially activated
- B cells' activity remains unchanged

NICTA

Result: simulated data





- Very effective in inferring differential leukocyte activity in simulated data

NICTA

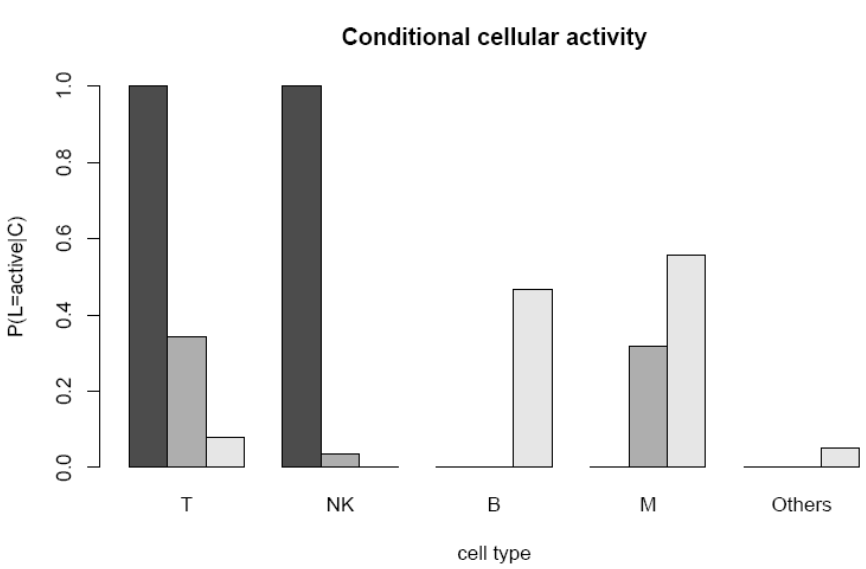
Real data – cardiovascular diseases

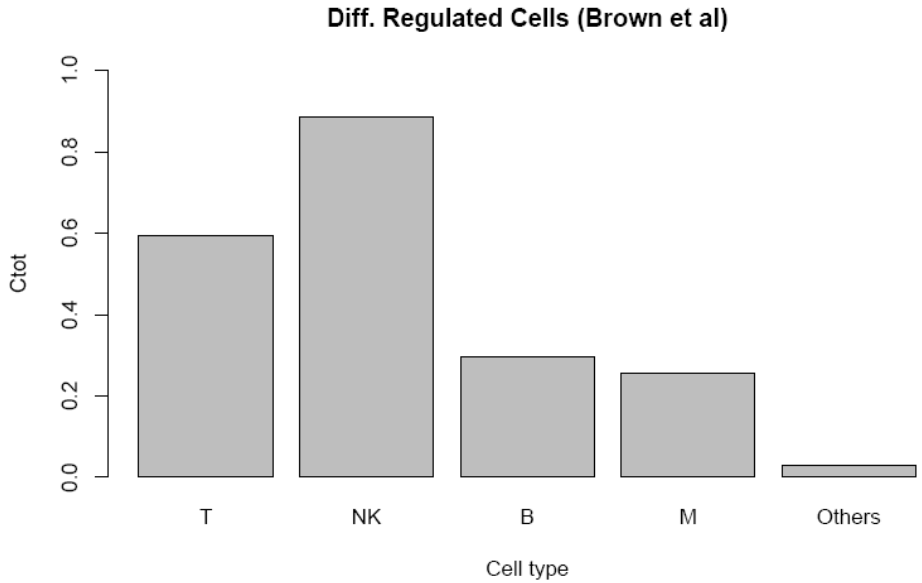
- Brown et al. – coronary artery diseases
  - Control: 19 healthy donor
  - Stable angina pectoris (SAP): 15 patients
  - Unstable angina pectoris (UAP): 19 patients
- Lui et al. – heart failure
  - Control: 19 healthy donor
  - Ischemic heart disease (IHD) – 22 patients
  - Idiopathic dilated cardiomyopathy (IDCM) – 15 patients

NICTA

Result: Brown et al.

Our Results





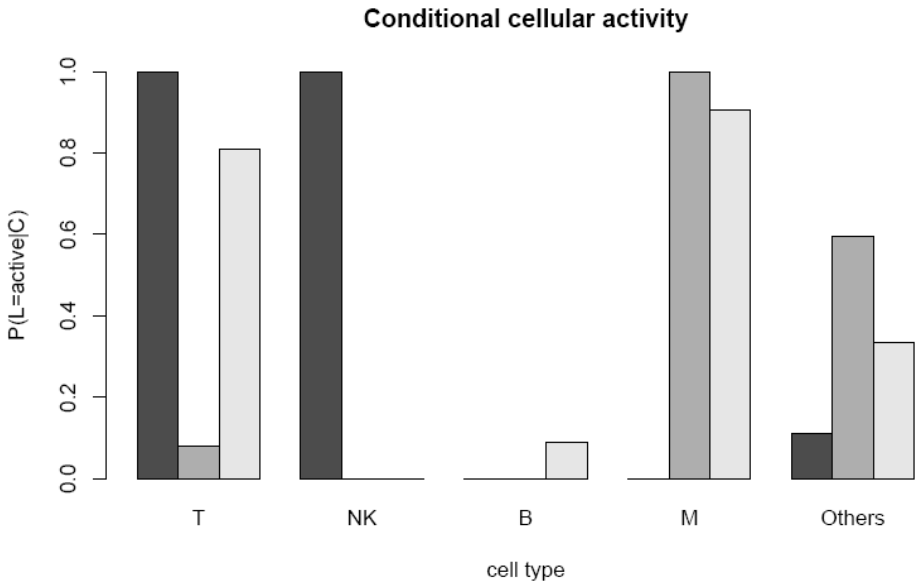
GSEA Result

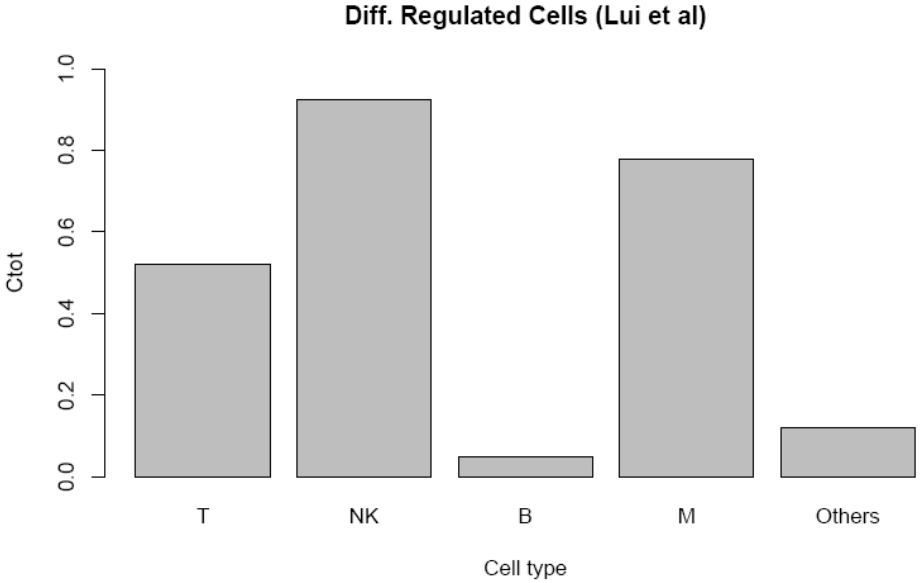
Analysis	Up-regulated in control (FDR)	Up-regulated in disease (FDR)
control vs. SAP	**T (0.11)** , B (0.66), NK (0.49)	M (0.33), Others (0.87)
control vs. UAP	T (0.27), NK (0.54)	M (0.62), Others (0.95), B (0.9)

NICTA

Result: Lui et al.

Our Results





GSEA Results

Analysis	Up-regulated in control (FDR)	Up-regulated in disease (FDR)
control vs. IHD	**T (0.051) , B (0.17), NK (0.17)**	M (0.64), Others (0.63)
control vs. IDCM	**T (0.25), B (0.15), NK (0.15)**	M (0.34), Others (0.67)

NICTA

Interpretation

- We found that cardiovascular diseases has:
  - Decreased T cell activity
  - Decreased NK cell activity
  - Increased monocyte activity
- Consistent with previous findings
- GSEA results
  - Significant drop in T and NK cells in only some cases
  - Failed to identify increases in monocyte activity
  - Significant B cell enrichment in HF is not consistent with visual inspection and known biology

NICTA

Summary

- We developed a latent variable model to infer differential leukocyte activity from antibody microarrays
- We show its usefulness with simulated dataset and two cardiovascular datasets
- Our result is better than GSEA-based analysis given the known biology
- We demonstrates how incorporating underlying biological processes using a probabilistic model is a very powerful approach in analyzing microarray data

NICTA

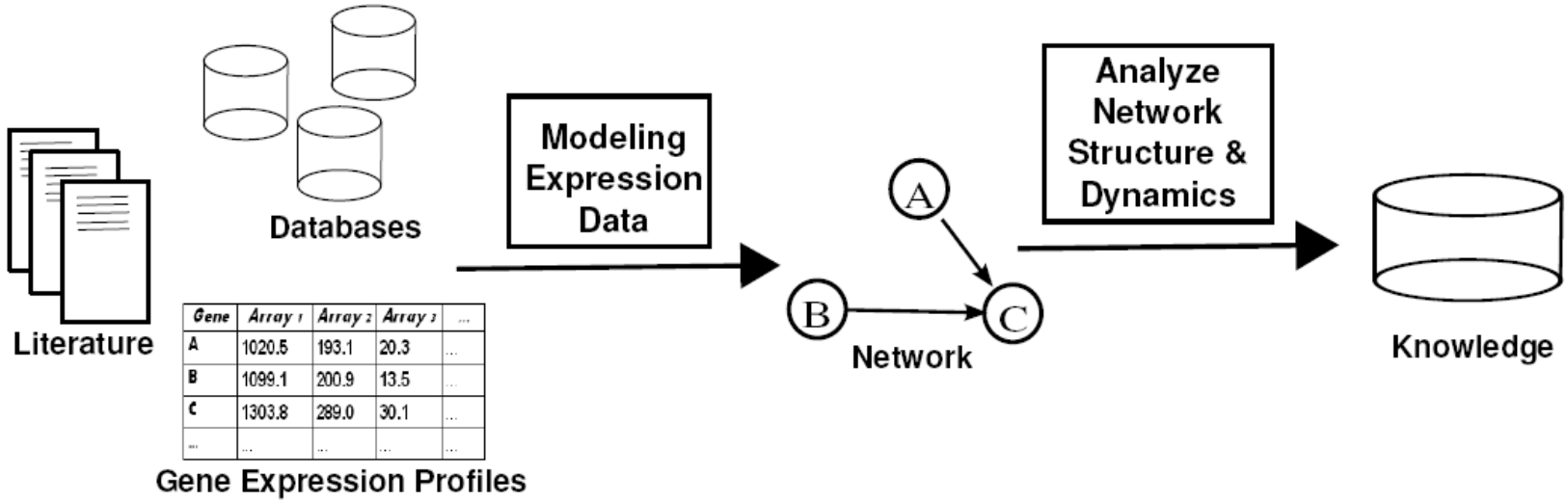
Acknowledgement

- Joint work with
  - Dr. Michael Charleston
  - Prof. Cristobal dos Remedios
  - Dr. Tiberio Ceatano (NICTA)
  - Rajeev Koundinya

- Funding:
  - The University of Sydney
  - NICTA
  - GIW student bursary
  - Sydney Bioinformatics

NICTA

Systems Biology



- Analysis of high-throughput data in the context of large interacting systems
- Need to model the biological data generation process, not just association mining
- Recover systems level properties that are not apparent when considering individual components independently

NICTA