

# **Deciphering the Transcription Patterns of a Malaria Parasite Using Multivariate Curve Resolution**

**Robert M. Flight & Peter D. Wentzell**

**Department of Chemistry, Dalhousie University  
Halifax, NS Canada**

**[rflight@dal.ca](mailto:rflight@dal.ca)   [peter.wentzell@dal.ca](mailto:peter.wentzell@dal.ca)**

# Malaria

- Blood borne disease caused by *Plasmodium* parasites, transmitted by *Anapholes* mosquito
- Infects hundreds of millions and kills millions annually
- No vaccine, current drug treatments becoming ineffective

# High Throughput Analysis

DNA



Transcription



RNA Polymerase

mRNA



Translation



Ribosomes

Protein

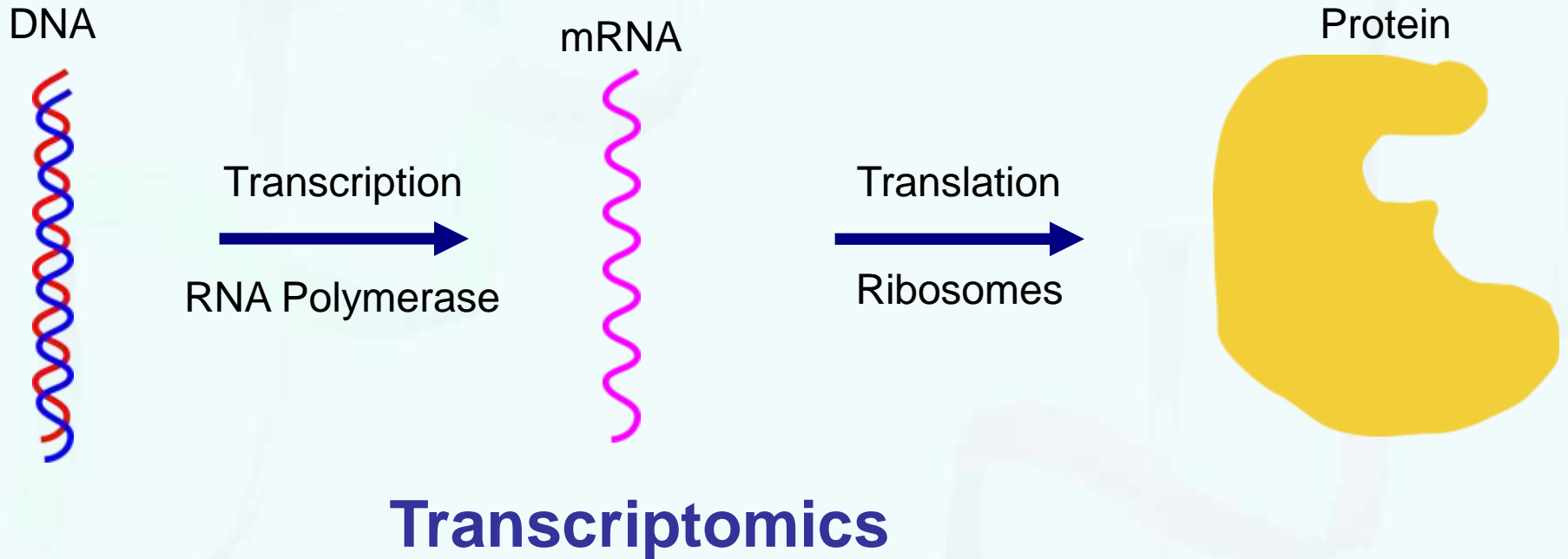


**Genomics**

**Transcriptomics**

**Proteomics**

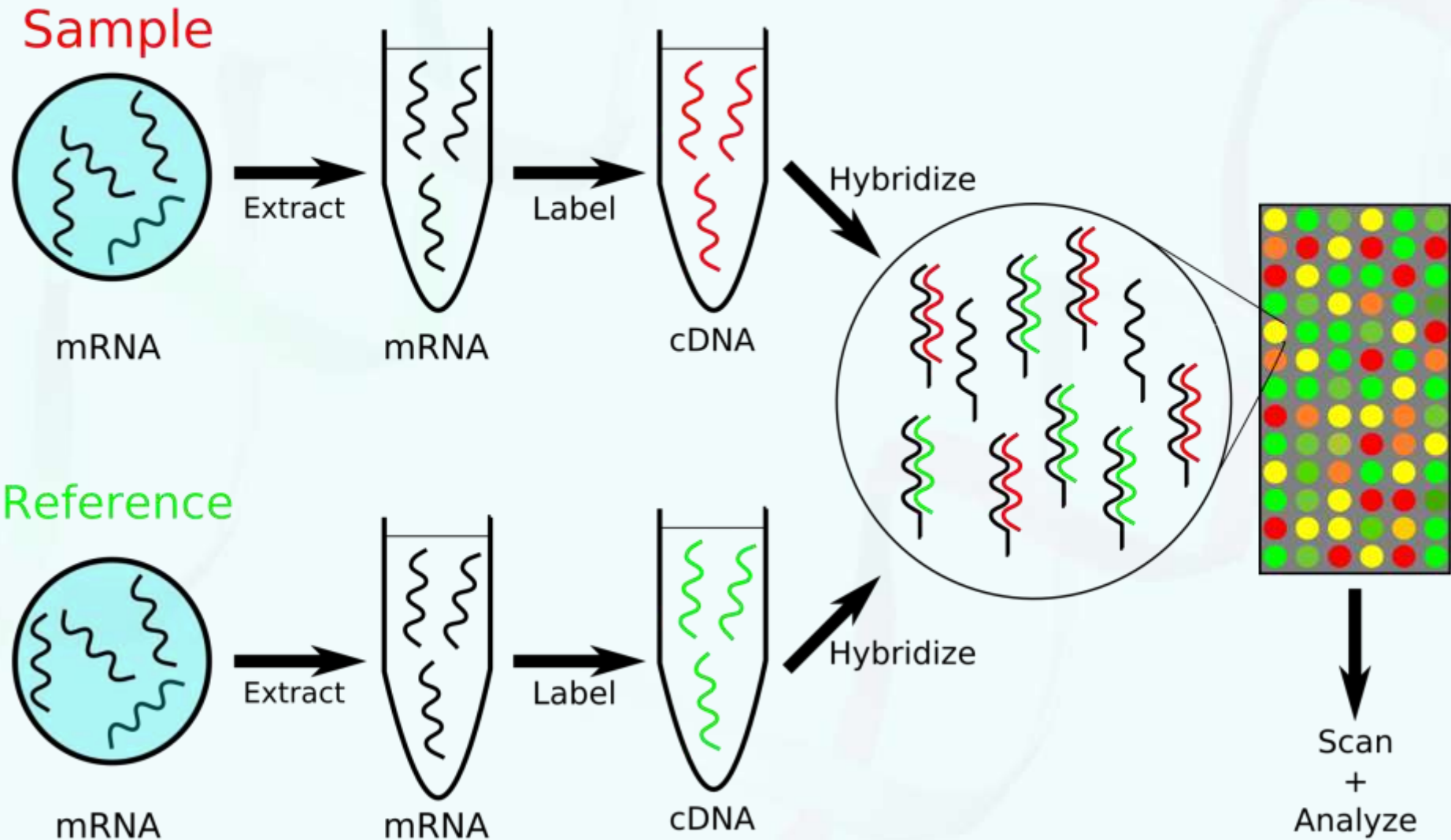
# High Throughput Analysis

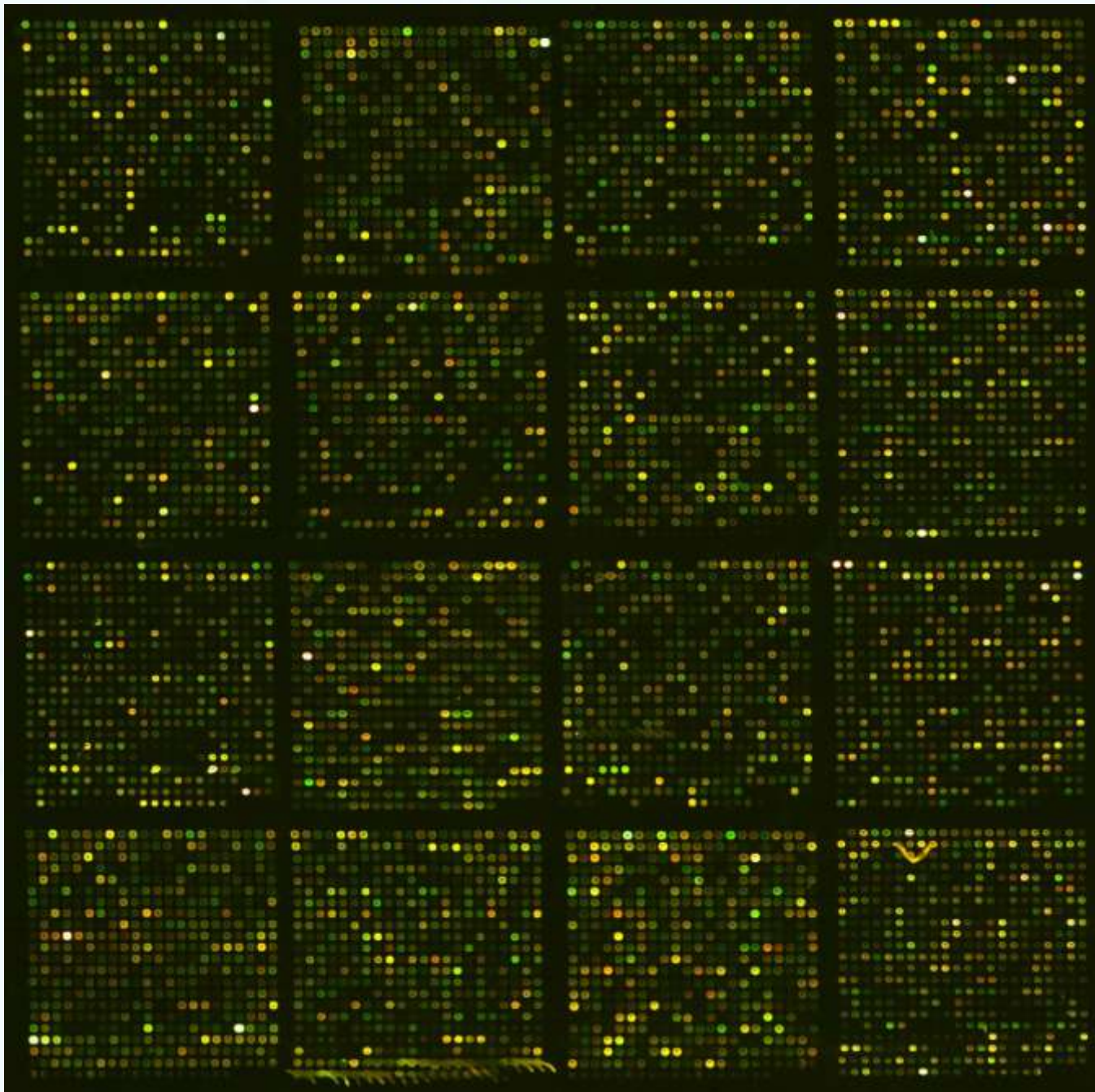


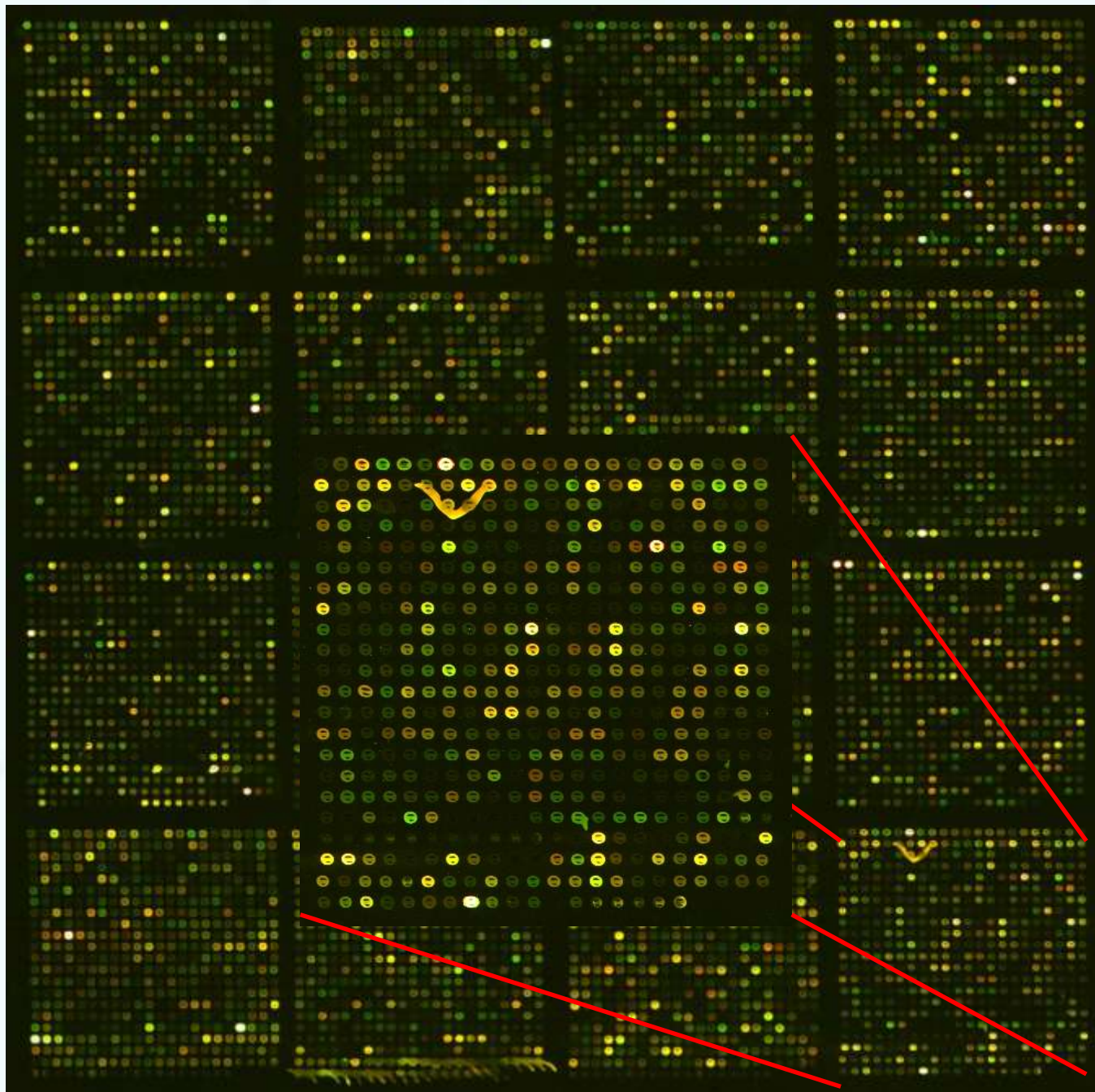
# DNA Microarray

- High throughput measure for mRNA transcripts
- DNA oligomers fixed to solid substrate
- Analyzed using:
  - Clustering
  - Support Vector Machines

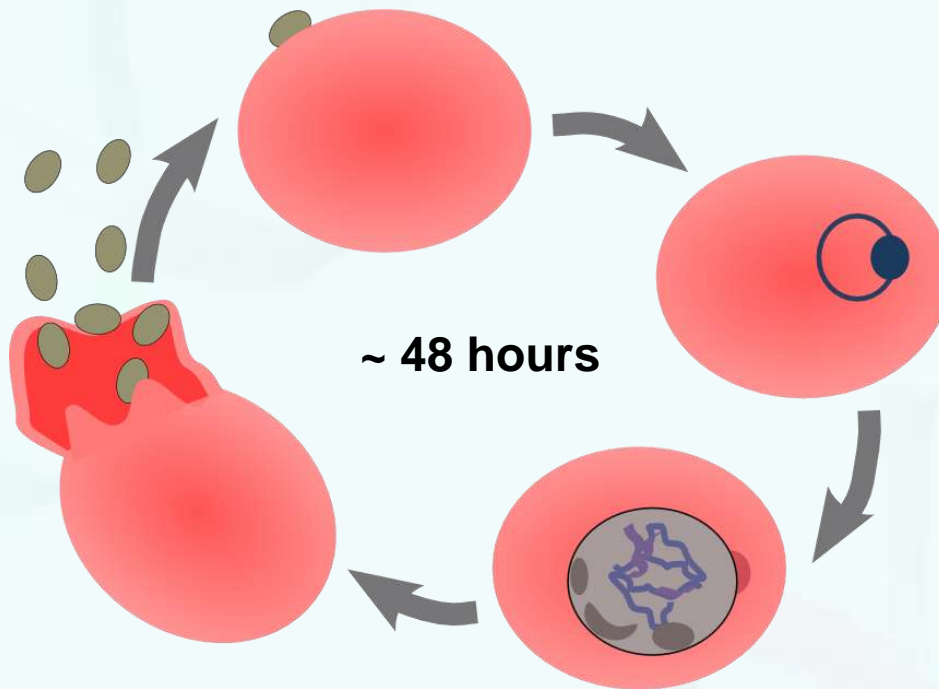
# Microarray Experiment





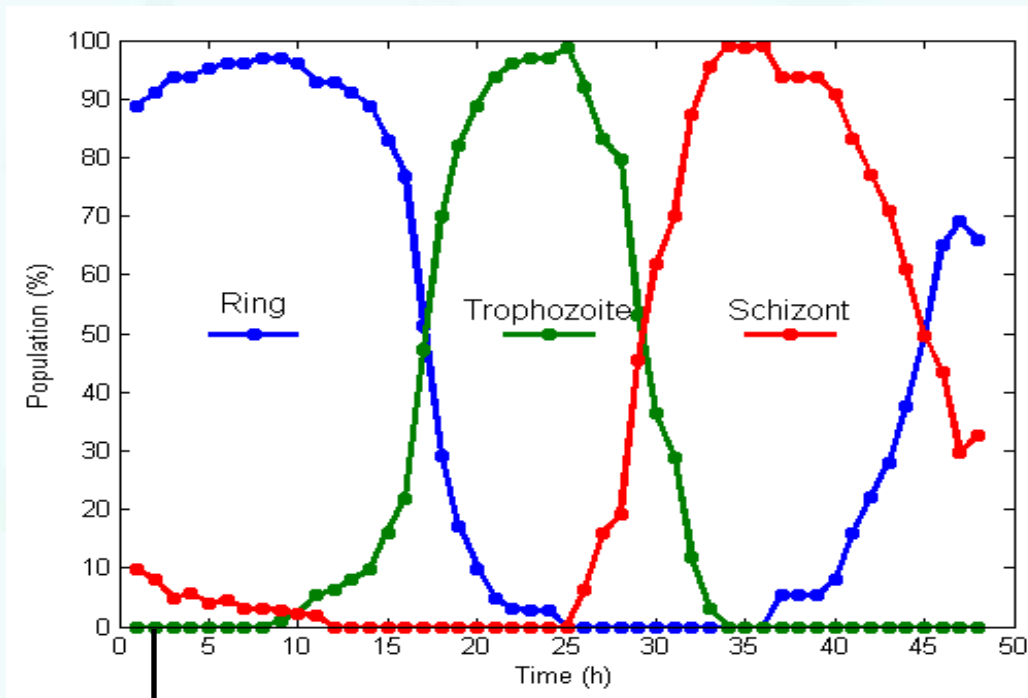


# Experiment

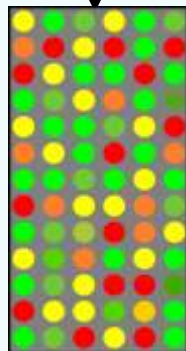


- Measurements taken throughout intraerythrocytic developmental cycle
- 55 timepoints with 9 replicates over 48 hours

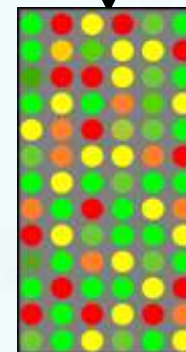
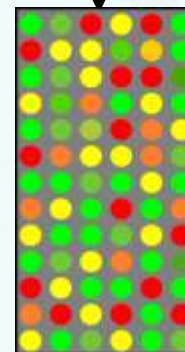
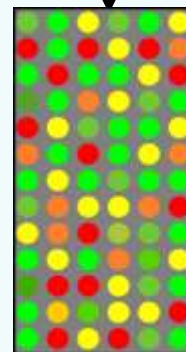
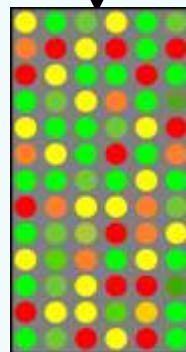
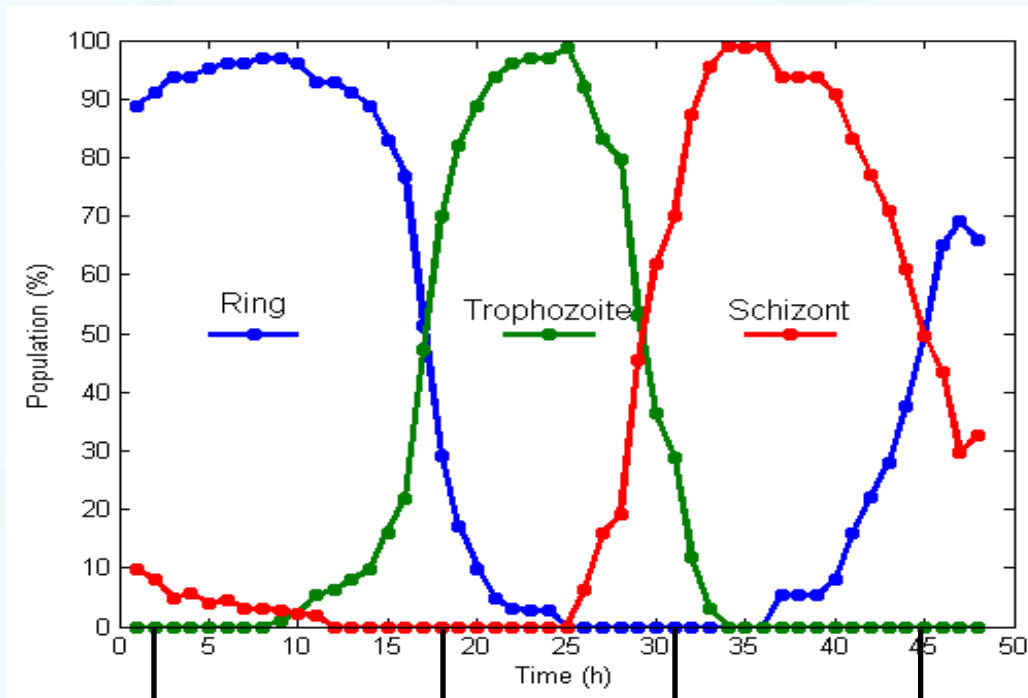
# Experiment



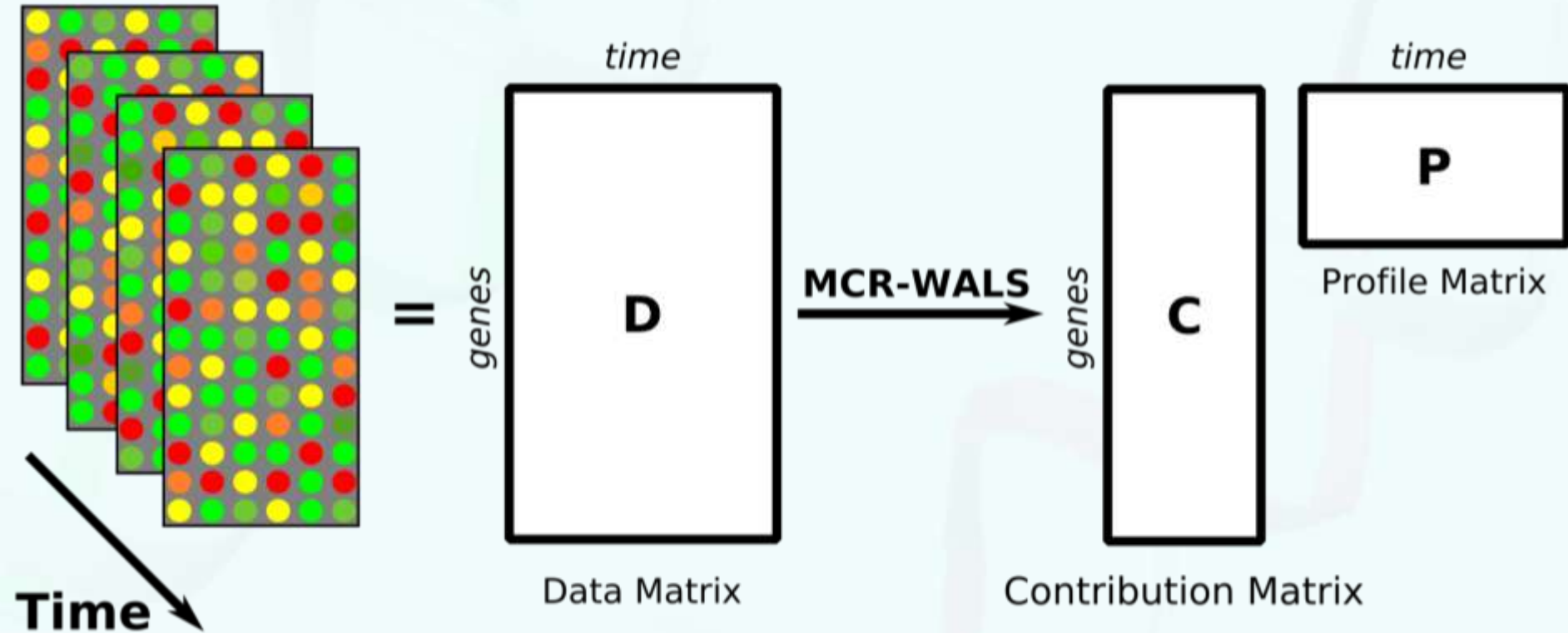
Extract mRNA



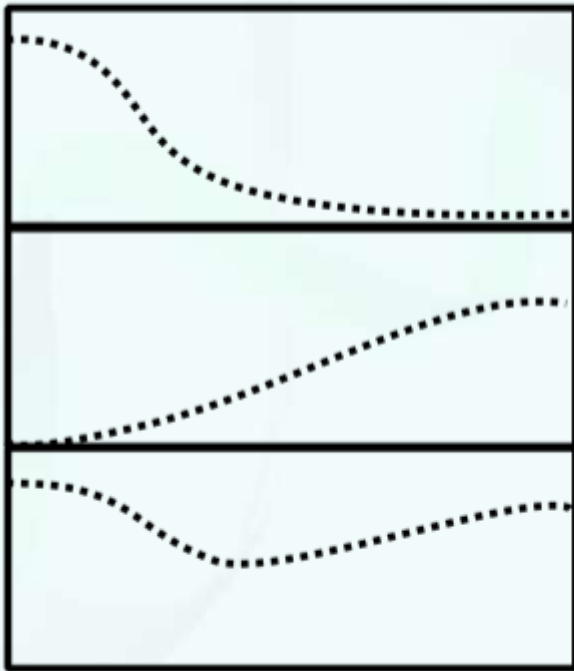
# Experiment



# Time-course Data



# Simple Example

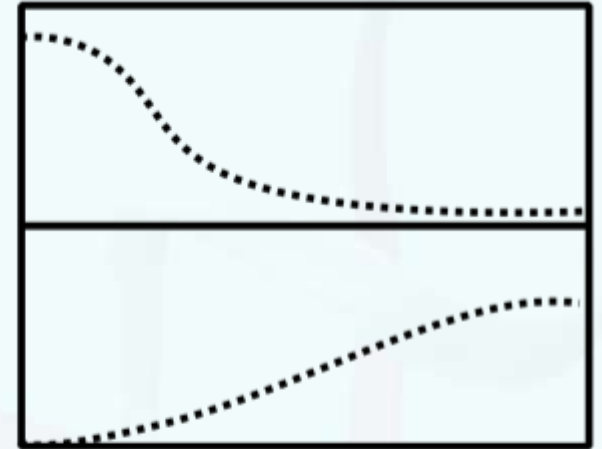


**D**

=

<b>1</b>	<b>0</b>
<b>0</b>	<b>1</b>
<b>1</b>	<b>1</b>

**C**



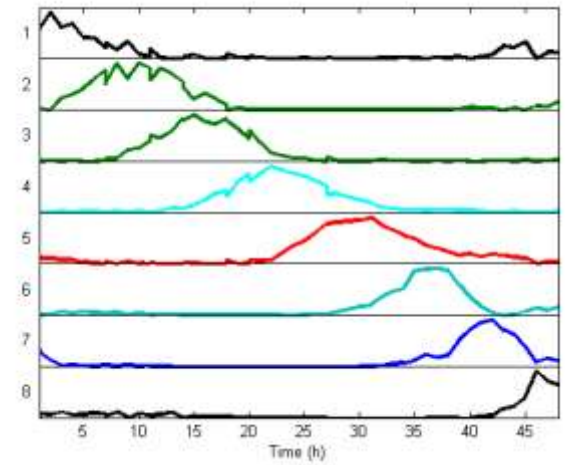
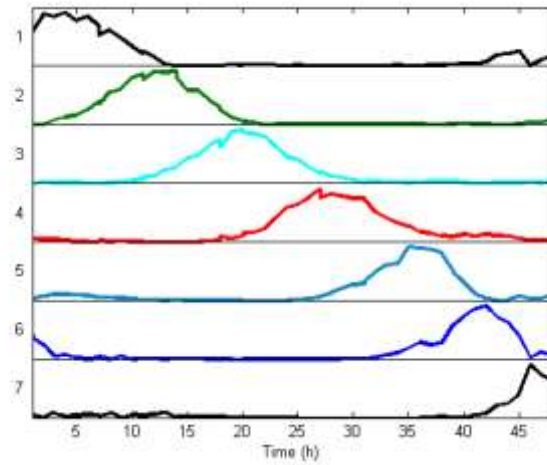
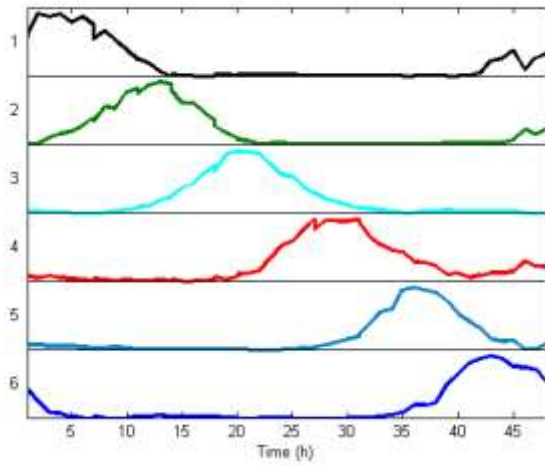
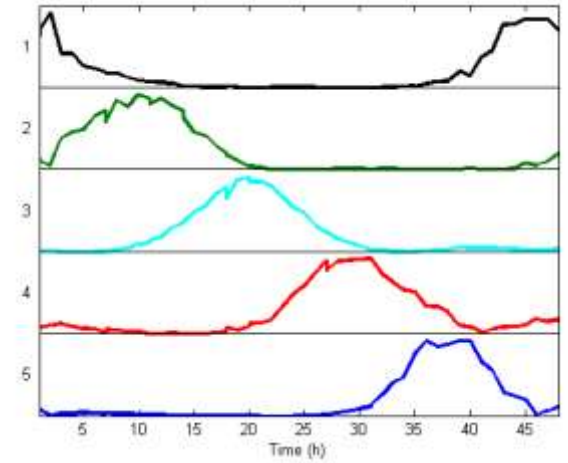
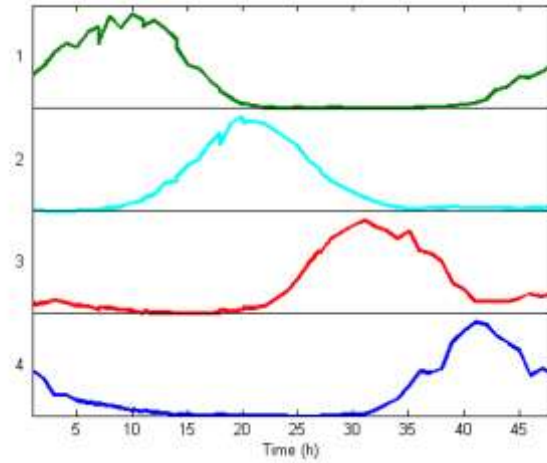
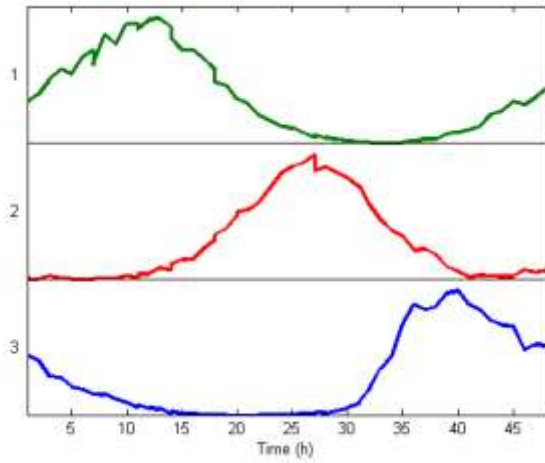
**P**

# MCR-WALS

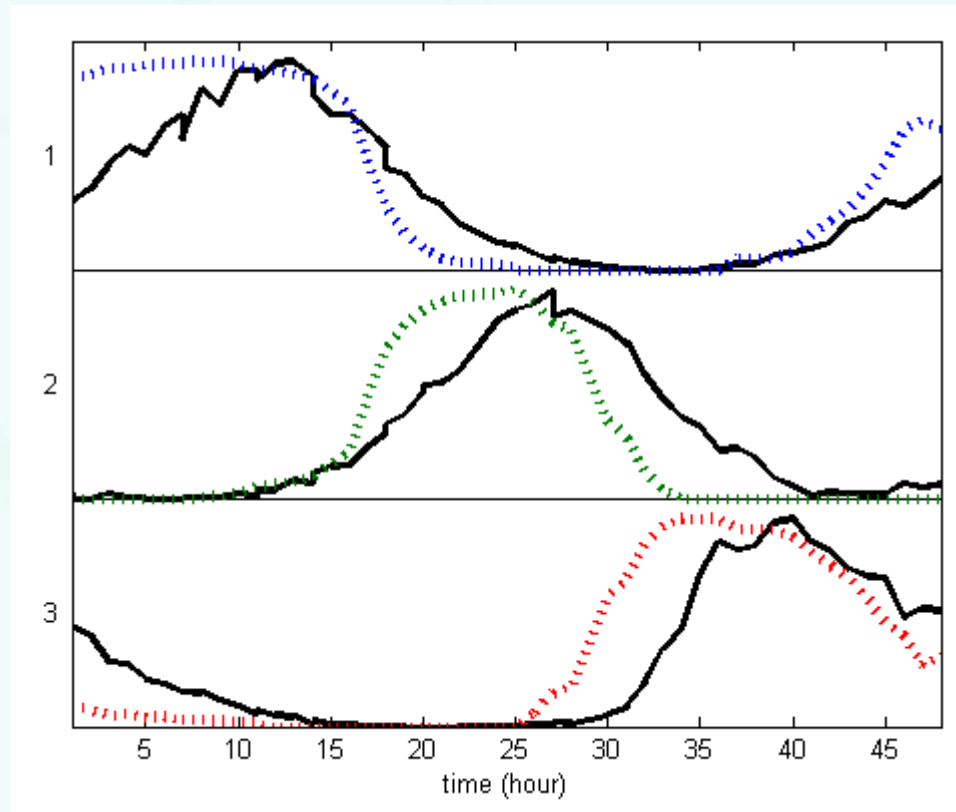
- **Weighted Alternating Least Squares**
- Incorporates measurement error
- At each stage of ALS, data points are **weighted** according to variance
- Impute missing points by assignment of large variance
- Uses raw ratios
  - Bilinear relationship
  - ALS non-negativity constraint

# Procedure

- Extract pixel intensities from both channels
- Calculate ratios and ratio variances
  - Orthogonal regression
  - Composite model of proportional errors and bootstrap error models
- **MCR-WALS**
  - 3 - 8 components



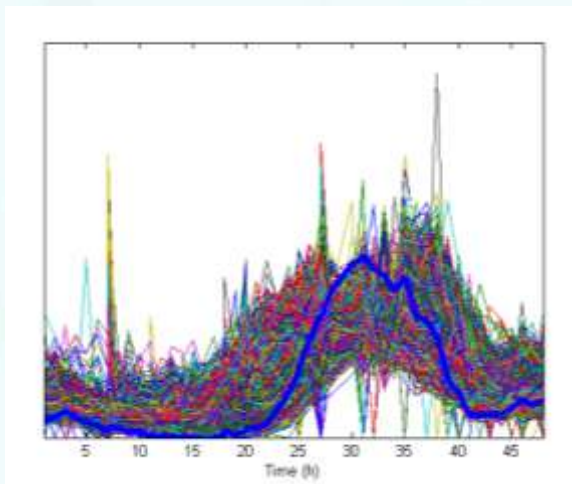
# Correlation to Physiology?



- Correlate MCR profiles and physiological profiles
- Physiological lead MCR-profiles

# Gene Ontology

- Look at terms associated with genes whose expression match the MCR profiles



PF14\_0124  
 PF11\_0114  
 PF14\_0218  
 PF13\_0326  
 PFI0520w  
 PFE0255w  
 PFL0415w  
 PFB0385w  
 PF14\_0749  
 PF10\_0016  
 PF10\_0015  
 PFL2570w

**GOSSIP**

Geop by MicroDiscovery and ITB

File Contig Analysis Help

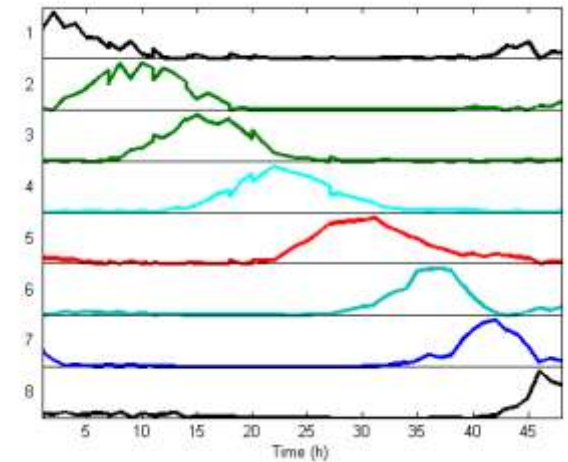
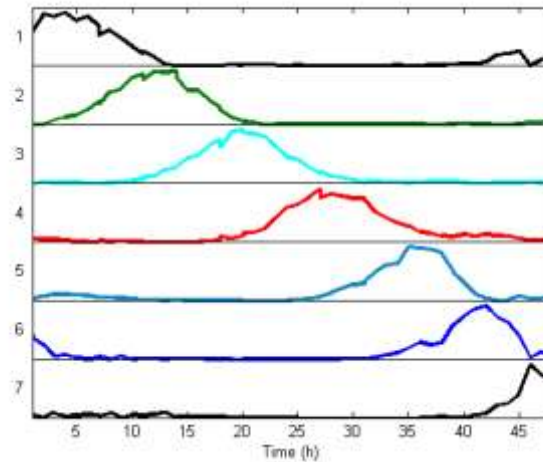
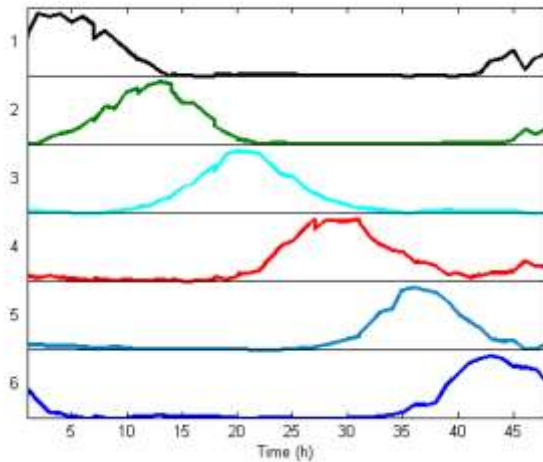
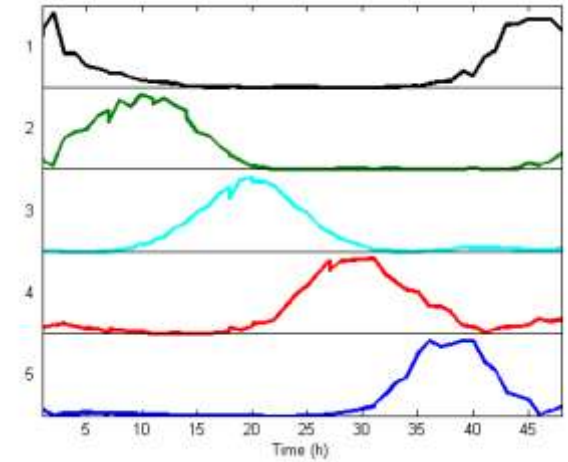
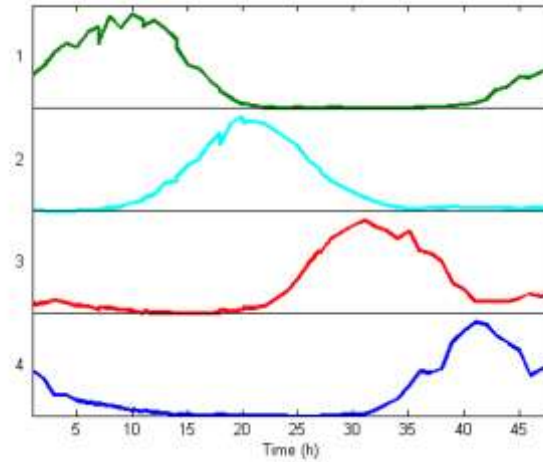
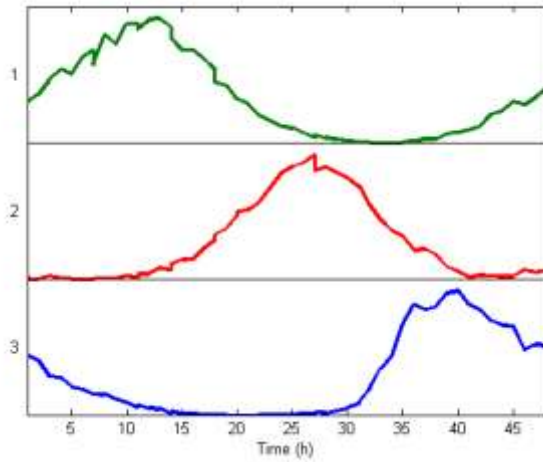
ID	GO Term	Name	FDR (Fisher's exact)	FWER (FDR-wise)	single test p-value	# in Test Set	# in Reference Set
1	43229	ribosome organon	4.1437e-05	3.00677e-03	2.94386e-11	132	348
2	43228	lipid droplet	4.1437e-05	3.00677e-03	2.94386e-11	132	348
3	43227	membrane bound organelle	4.1437e-05	3.00677e-03	8.52236e-11	107	251
4	43229	ribosome organon	4.1437e-05	3.00677e-03	8.52236e-11	107	251
5	5127	cytoplasm	2.82406e-05	2.57123e-03	8.675e-10	118	239
6	5128	basal body	2.82406e-05	2.57123e-03	1.1749e-09	66	108
7	4260	DNA replication	4.00546e-05	7.14056e-03	2.62001e-09	21	11
8	5127	cytoplasm	2.80796e-07	6.18208e-07	1.33725e-08	138	244
9	30811	apoptosis	2.51857e-07	7.01703e-07	2.95125e-08	50	88
10	4262	DNA metabolism	5.68846e-07	1.48271e-06	4.45446e-08	22	28
11	5094	chromosome	3.26173e-07	2.25493e-06	6.94525e-08	12	22
12	30804	apoptosis	2.2428e-08	7.07133e-08	1.3628e-07	6	0
13	5657	replicative life cycle	2.45273e-06	1.13188e-05	3.28416e-07	30	1
14	5622	cell	3.66634e-05	8.0032081e-05	3.08334e-06	146	483
15	4739	cell of nucleus	0.00015202	0.000542502	1.28396e-05	33	52
16	43883	transcription initiation	0.00028863	0.000302103	2.3895e-05	61	119
17	50676	cellular physiological process	0.000591155	0.000750505	4.32084e-05	176	319
18	3967	cellular process	0.00046688	0.00058480	8.98354e-05	173	690
19	5152	metabolism	0.000468	0.00071457	8.00700e-05	144	504
20	44210	cellular metabolism	0.00066668	0.00081337	8.00807153	144	507
21	5024	cellular activity	0.00062579	0.00075048	8.007954027	138	367
22	3967	DNA metabolic process	0.00044805	0.0004804	8.00794019	6	1
23	1663	DNA replication fork	0.00044303	0.000291	8.00790334	4	0
24	42679	DNA transcription	0.00044303	0.000291	8.00790334	4	0
25	42679	DNA-dependent DNA replication	0.00044303	0.000291	8.00790334	4	0
26	3810	transcription	0.0002622	0.0002622	8.00790334	11	18
27	313	lipid droplet	0.0002622	0.0002622	8.00790334	6	0
28	44238	protein metabolic process	0.0002622	0.0002622	8.00790334	138	470
29	5120	nucleosome, nucleolus	0.0002622	0.0002622	8.00790334	52	154
30	5607	ribosome	0.0002622	0.0002622	8.00790334	6	1
31	2547	plastid, thylakoid	0.0002622	0.0002622	8.00790334	4	1
32	transcription initiation	0.0002622	0.0002622	8.00790334	6	1	

# Enriched Terms

<b>GOID</b>	<b>GO Term</b>	<b>P-value</b>
GO:0006260	DNA replication	5.0e-009
GO:0006259	DNA metabolism	5.0e-009
GO:0006261	DNA-dependent DNA replication	1.6e-007
GO:0005694	chromosome	9.5e-007
GO:0043283	biopolymer metabolism	6.4e-006
GO:0009987	cellular process	7.5e-005
GO:0050875	cellular physiological process	7.5e-005
GO:0005657	replication fork	1.9e-004
GO:0030894	replisome	3.0e-004
GO:0006270	DNA replication initiation	6.7e-004
GO:0007067	mitosis	6.7e-004
GO:0000087	M phase of mitotic cell cycle	6.7e-004
GO:0043226	organelle	9.4e-004
GO:0043229	intracellular organelle	9.4e-004
GO:0043231	intracellular membrane-bound organelle	1.7e-003
GO:0043227	membrane-bound organelle	1.7e-003
GO:0000279	M phase	2.2e-003
GO:0008152	metabolism	2.2e-003
GO:0005663	DNA replication factor C complex	4.3e-003
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	4.5e-003
GO:0000278	mitotic cell cycle	6.6e-003
GO:0008094	DNA-dependent ATPase activity	6.8e-003

# Enriched Terms

<b>GOID</b>	<b>GO Term</b>	<b>P-value</b>
GO:0006260	DNA replication	5.0e-009
GO:0006259	DNA metabolism	5.0e-009
GO:0006261	DNA-dependent DNA replication	1.6e-007
GO:0005694	chromosome	9.5e-007
GO:0043283	biopolymer metabolism	6.4e-006
GO:0009987	cellular process	7.5e-005
GO:0050875	cellular physiological process	7.5e-005
GO:0005657	replication fork	1.9e-004
GO:0030894	replisome	3.0e-004
GO:0006270	DNA replication initiation	6.7e-004
GO:0007067	mitosis	6.7e-004
GO:0000087	M phase of mitotic cell cycle	6.7e-004
GO:0043226	organelle	9.4e-004
GO:0043229	intracellular organelle	9.4e-004
GO:0043231	intracellular membrane-bound organelle	1.7e-003
GO:0043227	membrane-bound organelle	1.7e-003
GO:0000279	M phase	2.2e-003
GO:0008152	metabolism	2.2e-003
GO:0005663	DNA replication factor C complex	4.3e-003
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	4.5e-003
GO:0000278	mitotic cell cycle	6.6e-003
GO:0008094	DNA-dependent ATPase activity	6.8e-003



■ Transcription & Translation

■ Metabolism

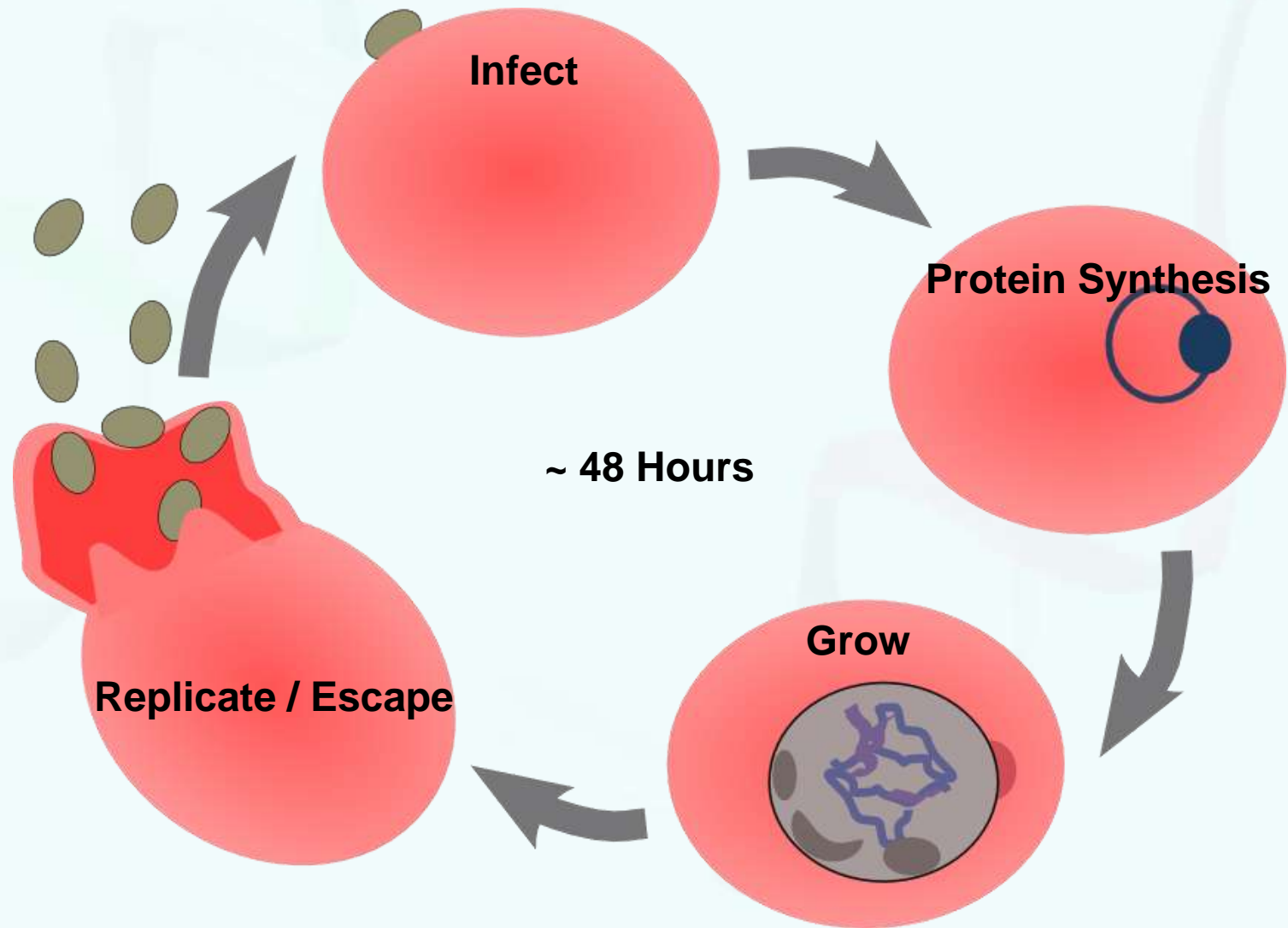
■ DNA Replication

■ Protein Modification

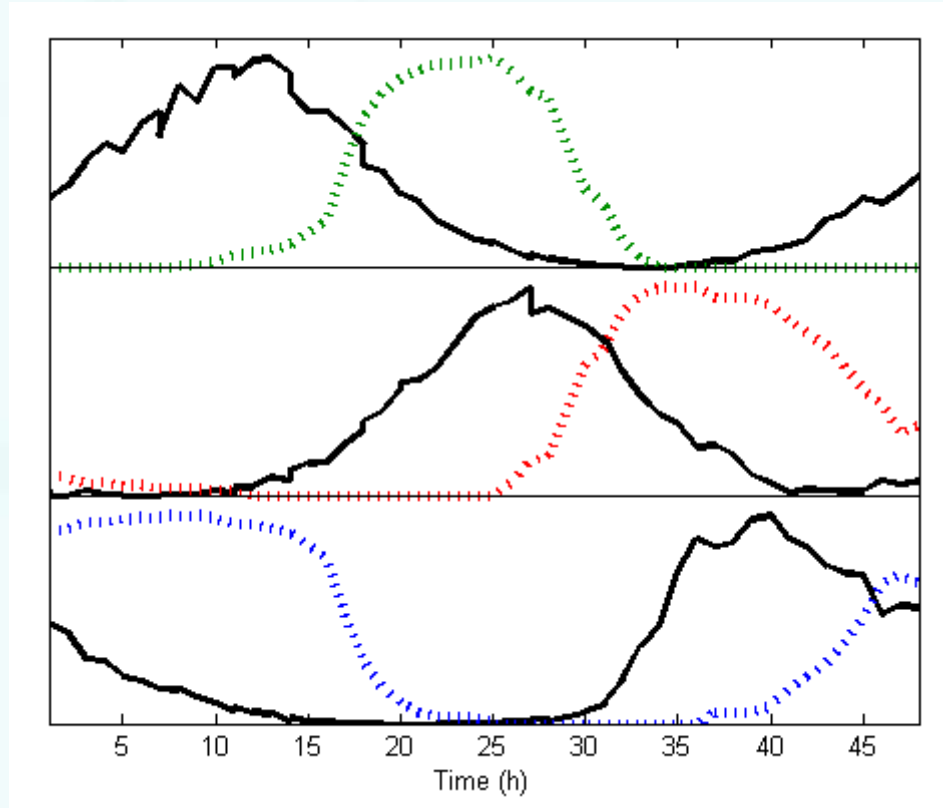
■ Cell Invasion

■ No significant terms

# Intraerythrocytic Developmental Cycle



# MCR Profiles vs Biology



- MCR profiles lead physiological profiles

# Conclusions

- MCR-WALS can extract biologically relevant information from time-course microarray data
- GO term analysis allows the assignment of biologically relevant functions, and aids interpretation of profiles
- Future work
  - Improve algorithm
  - Use contribution matrix
  - Other types of biological data

# Acknowledgements

## Wentzell Group

Tobias Karakach

Hannes Hochreiner

Yazeed Sobaih



**CGSS**

**CAC 2006**

1981



“What a day.... I must have spread malaria across half the country.”

