Analyzing the transcriptional response of *P. falciparum* to drugs and stress

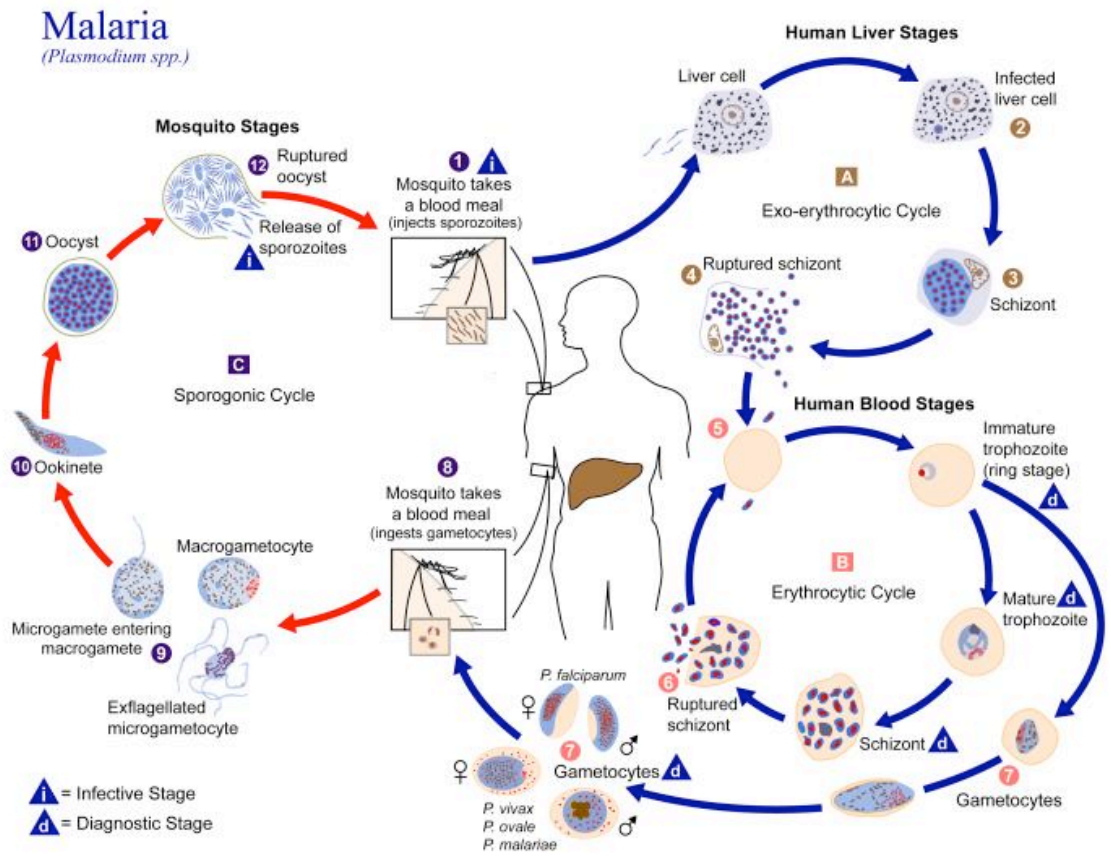
Jean-Philippe Vert

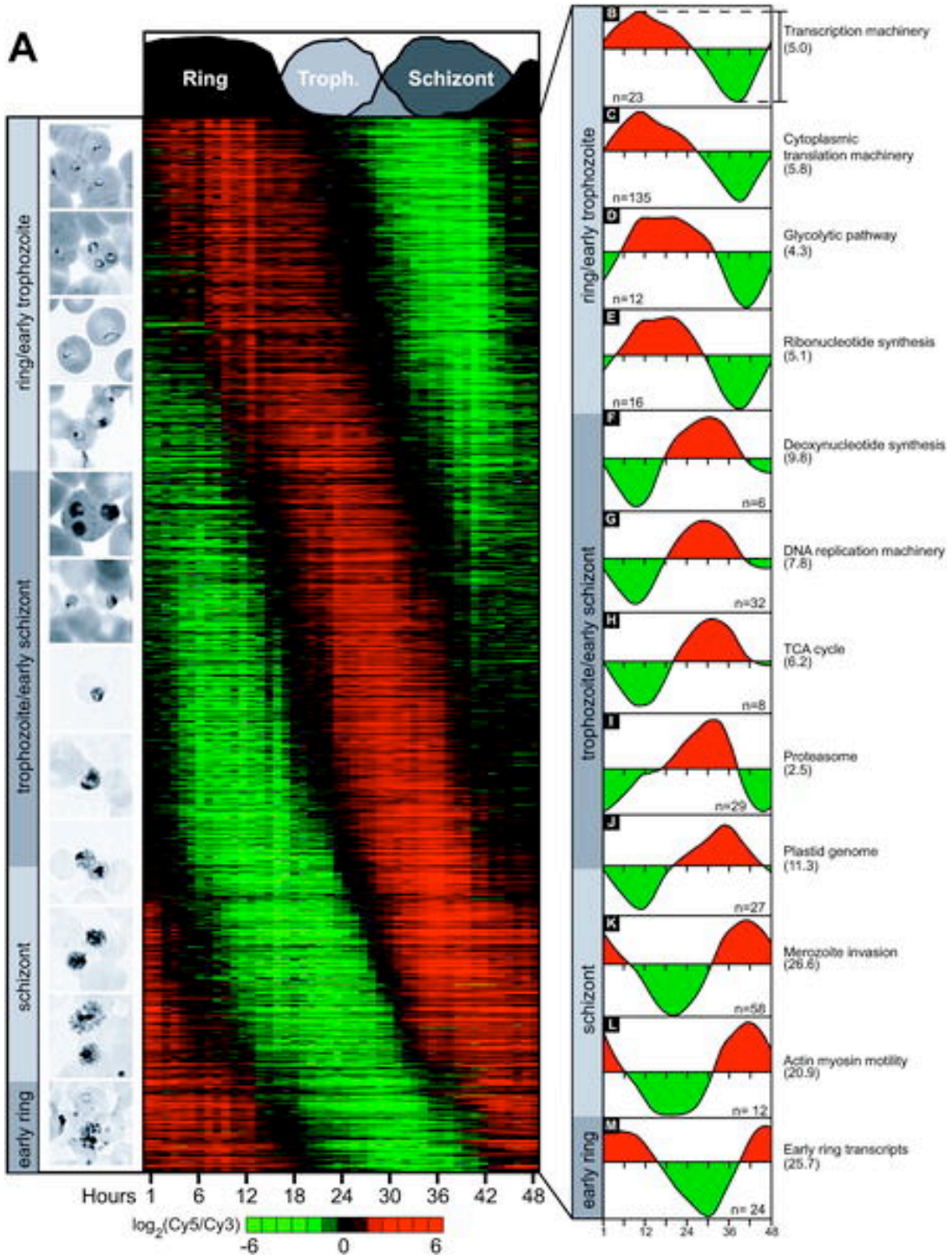
Mines ParisTech / Institut Curie / Inserm

*« Bioinformatique de P. falciparum » w*orkshop

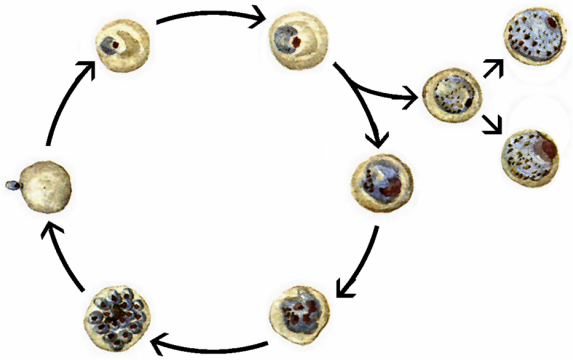
Paris, France, Jan 21, 2009

Plasmodium life cycle





Intraerythrocytic Developmental Cycle (IDC)



75% of the genes are activated only once during the IDC

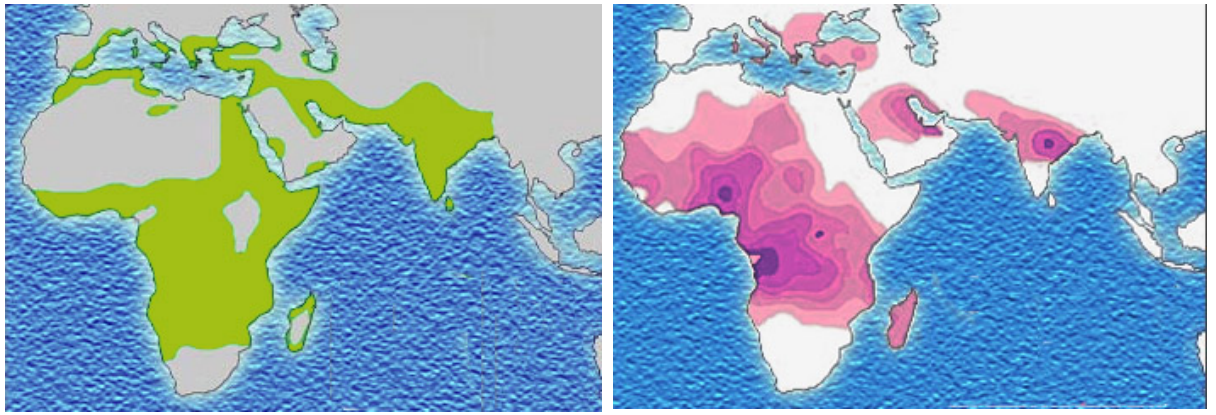
(Bozdech et al., 2003)

Motivations

- Study the response of the parasite to a drug/ stress at the transcriptome level
- Work with synchronized colonies
- Questions:
  - Is there a response?
  - If yes, which genes / pathways are involved?

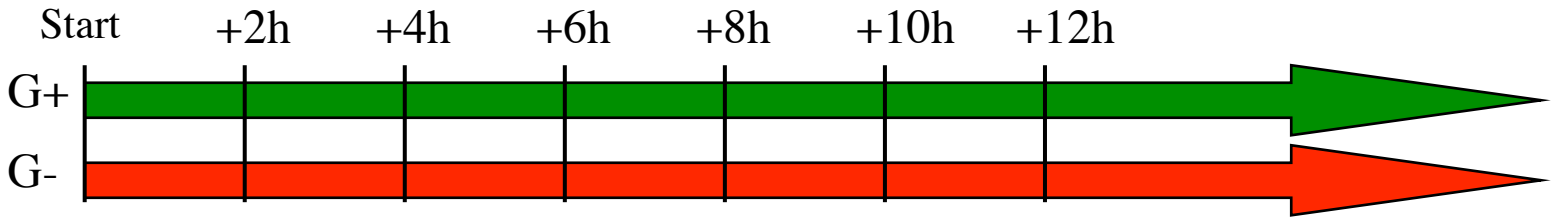
Application 1: HbAS

Malaria

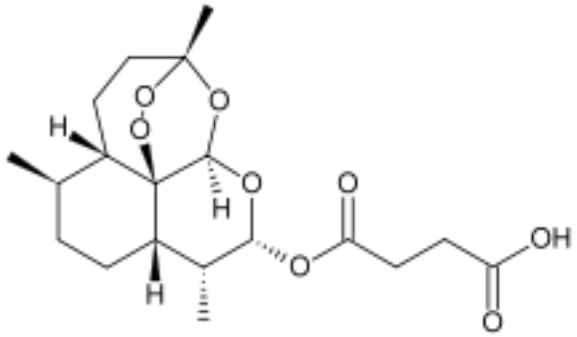


Drepanocytosis

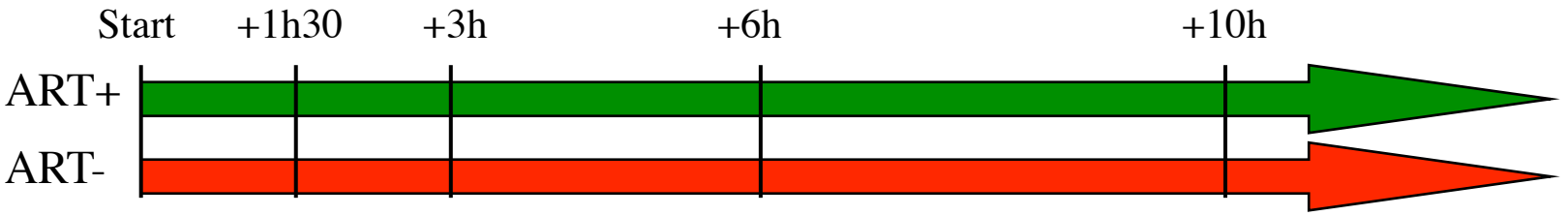
- Blood of HbAS heterozygous patients with drepanocytosis
- Compare low oxygen pressure G+ vs normal conditions G- in HbAS blood.



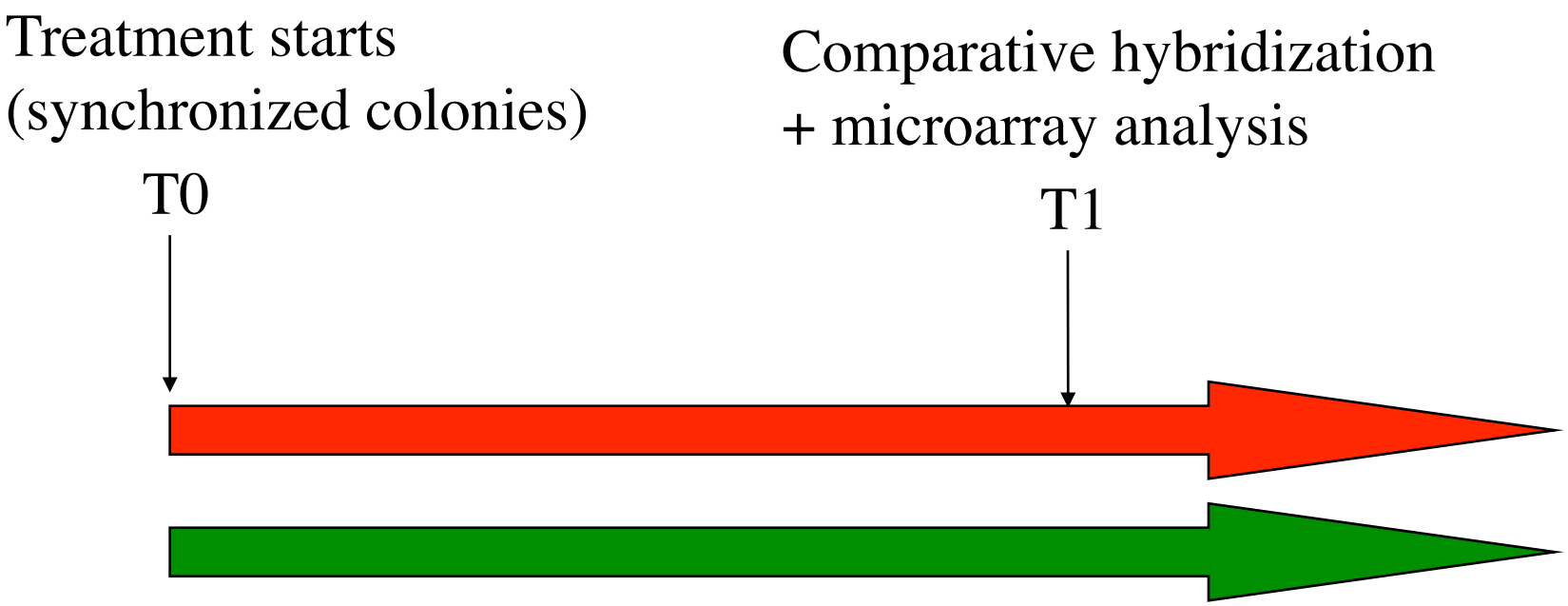
Application 2: ART



- Artesunate is part of the artemisin group of anti-malarial drugs
- Follow the effect of artesunate injection



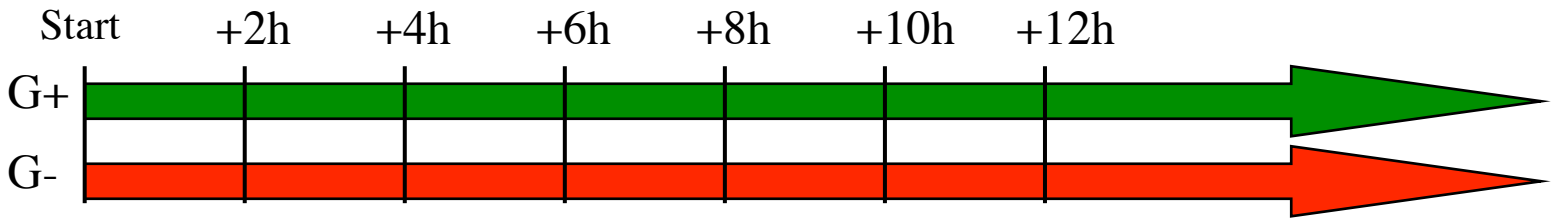
Experimental setup



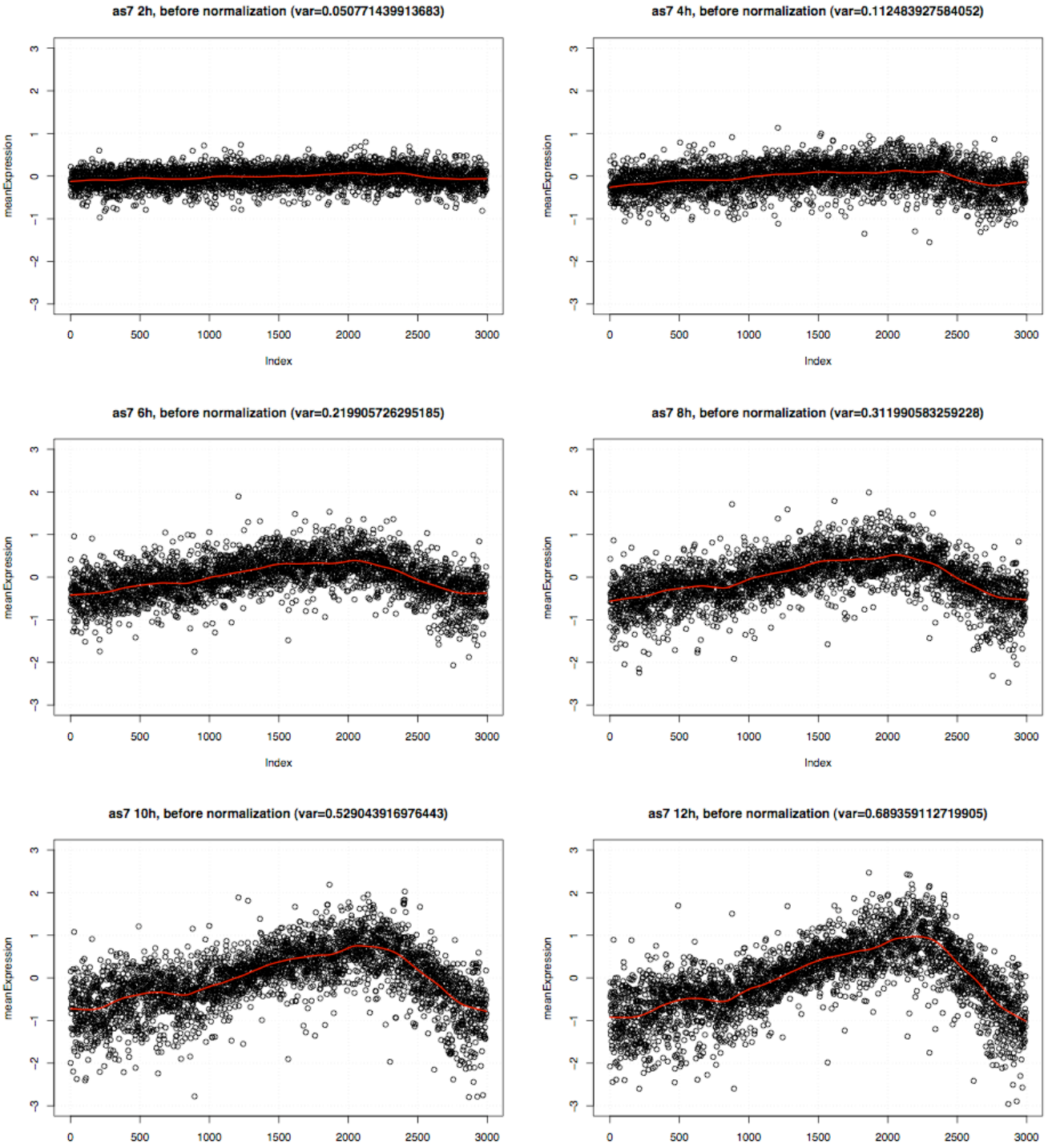
Results for HbAS

Experiment	Overexpressed	Underexpressed	Not differentially expressed
2h	67	39	5481
4h	114	202	5317
6h	845	927	3813
8h	1018	1156	3494
10h	1203	1373	3087
12h	1188	1332	3108

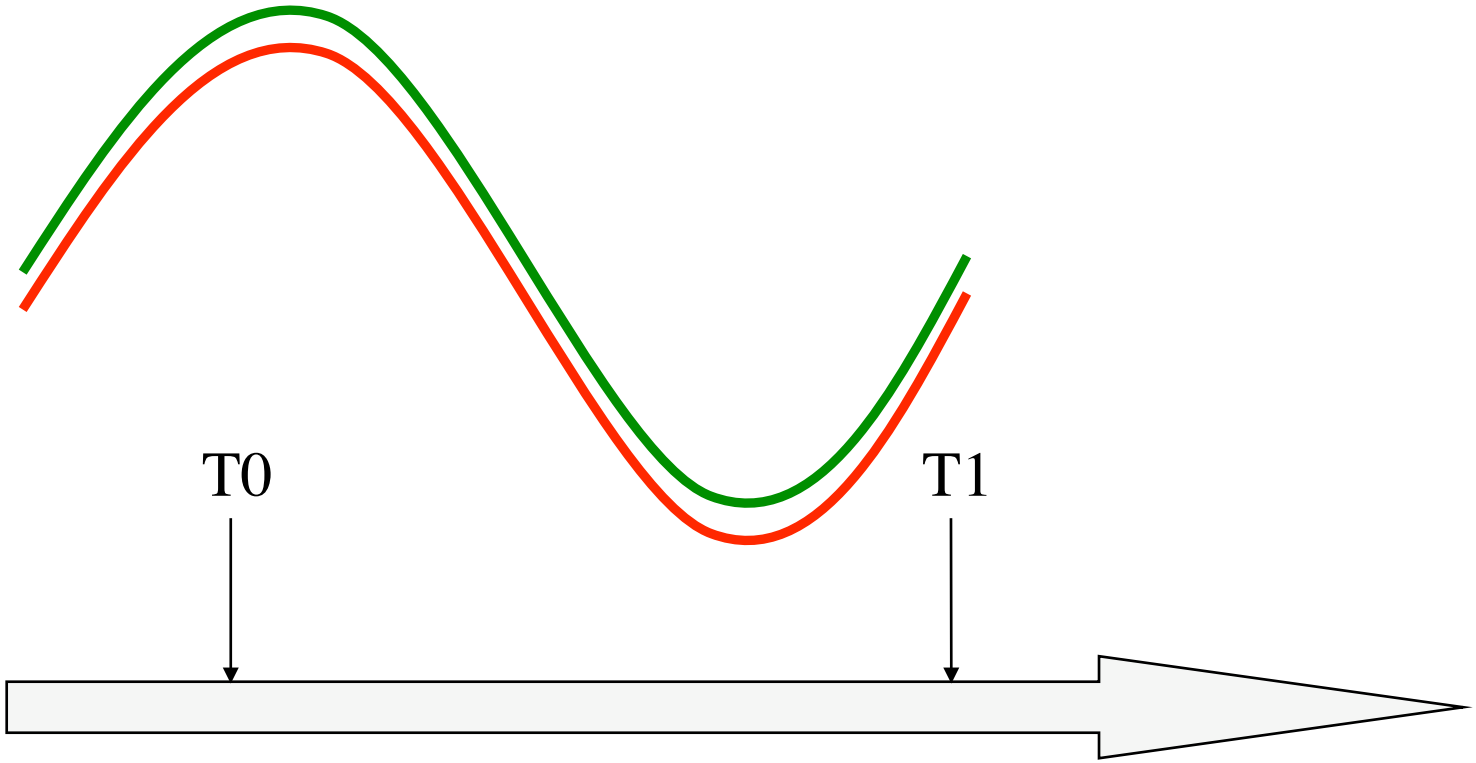
Many genes!



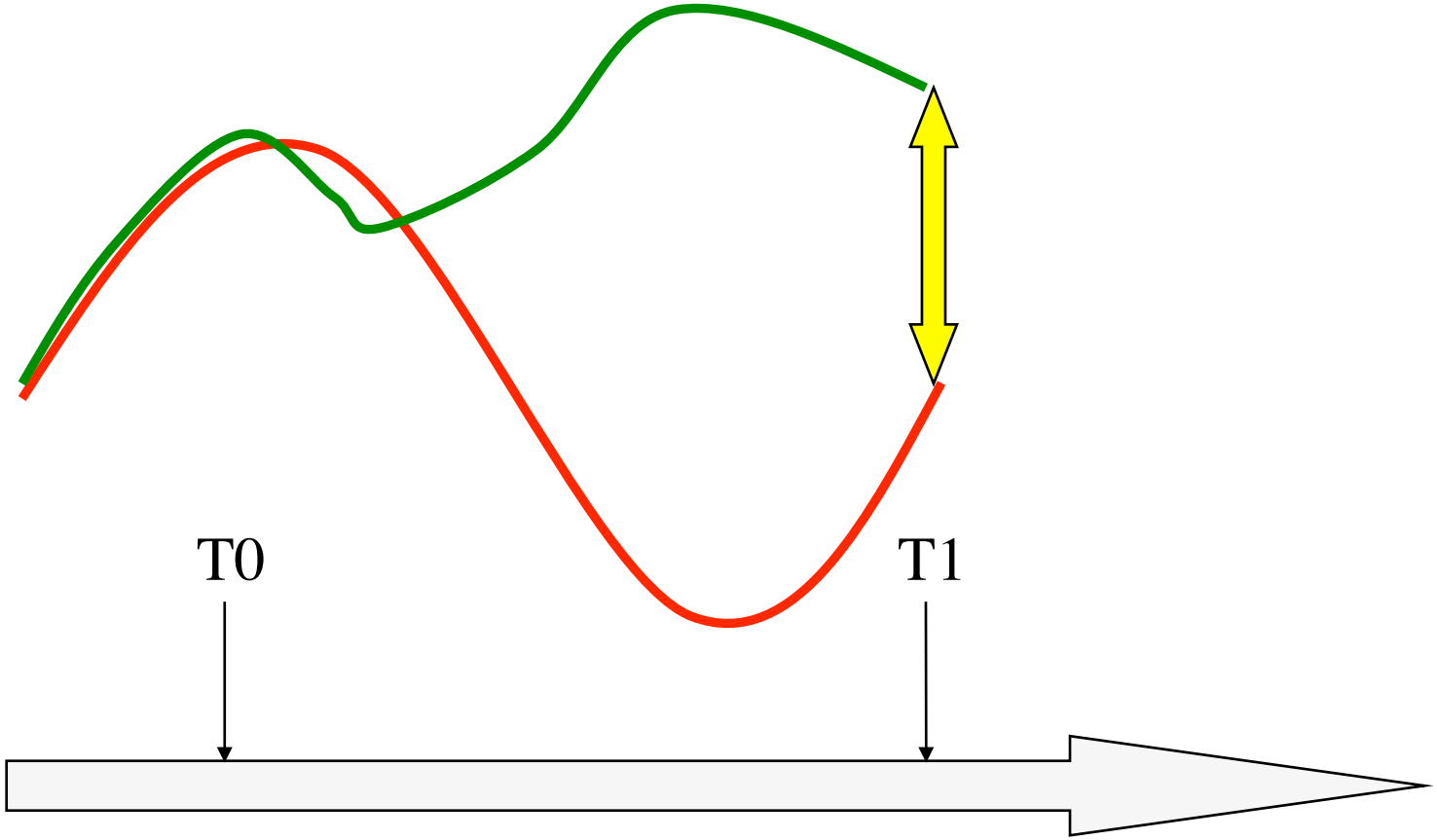
But…


as7 2h, before normalization (var=0.050771439913683)
as7 4h, before normalization (var=0.112483927584052)
as7 6h, before normalization (var=0.219905726295185)
as7 8h, before normalization (var=0.311990583259228)
as7 10h, before normalization (var=0.529043916976443)
as7 12h, before normalization (var=0.689359112719905)
meanExpression
Index

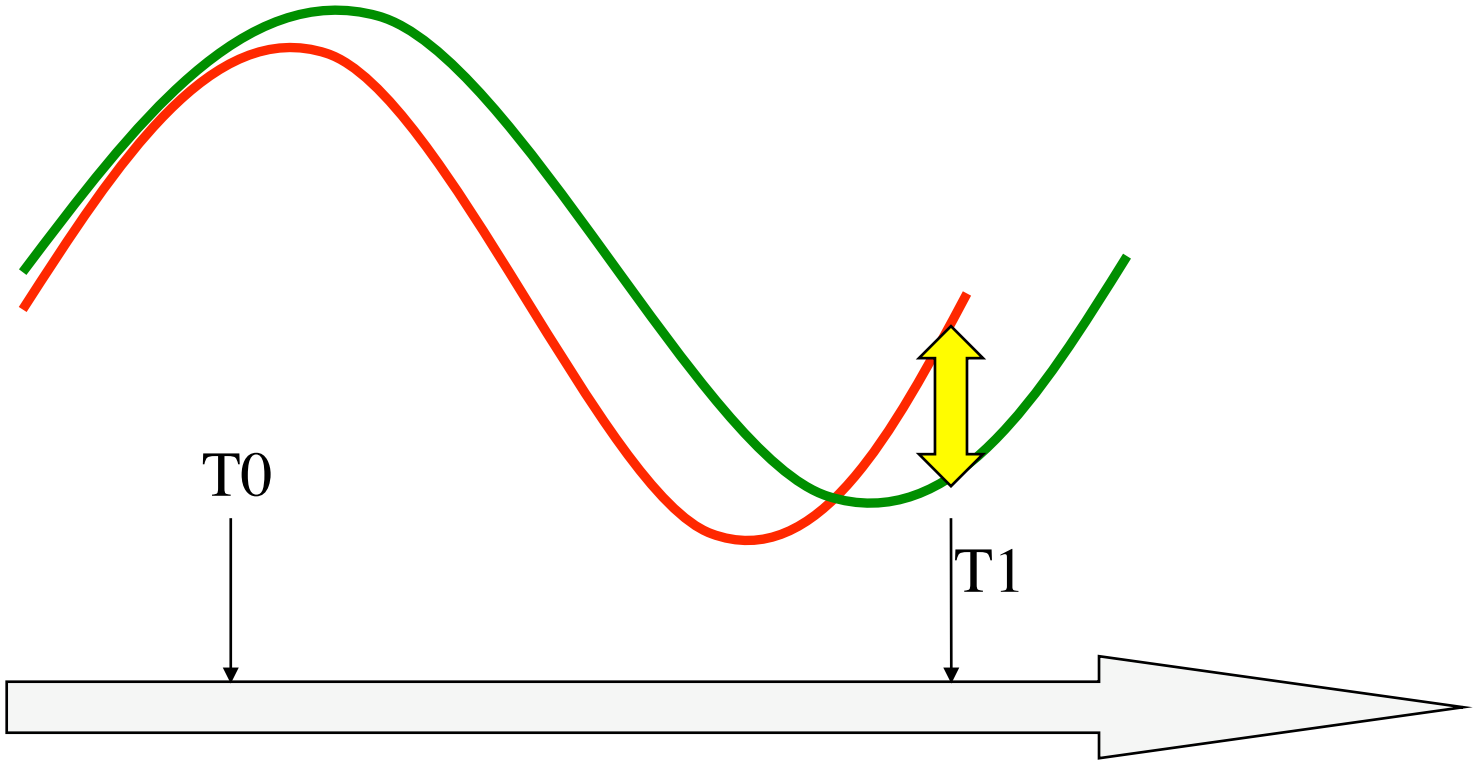
No treatment



Treatment with effect on a gene



Treatment with effect of the cell cycle speed



The Valid experiment

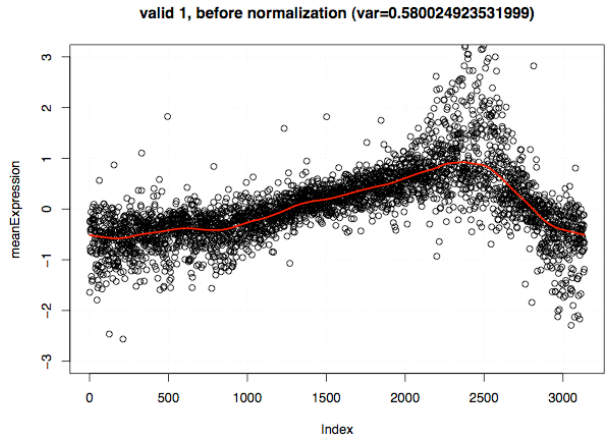
- 4 experiments (x 4 arrays) to assess the effect of pure age shift
- No treatment

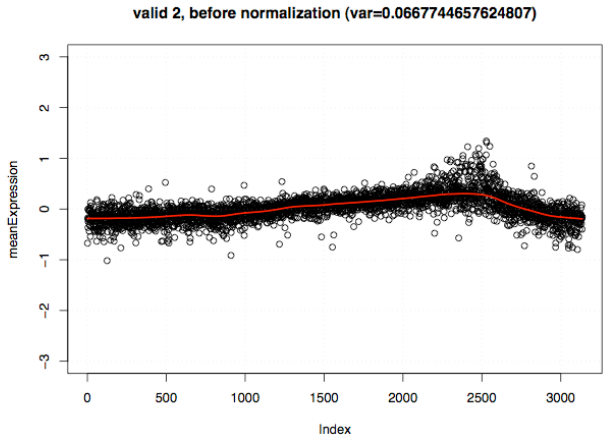
Experiment	Condition 1	Condition 2
Valid 1	4h	10h
Valid 2	6h	8h
Valid 3	8h	12h
Valid 4	10h	12h

Results for the Valid experiment

Experiment	Overexpressed	Underexpressed	Not differentially expressed
Valid 1	897	711	4436
Valid 2	157	116	5786
Valid 3	529	173	5385
Valid 4	290	77	5712

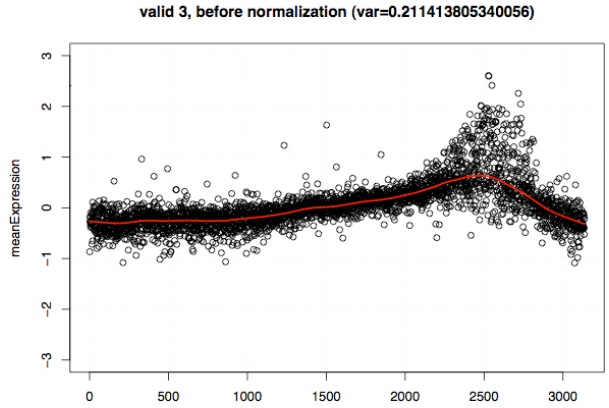
Δ=6h

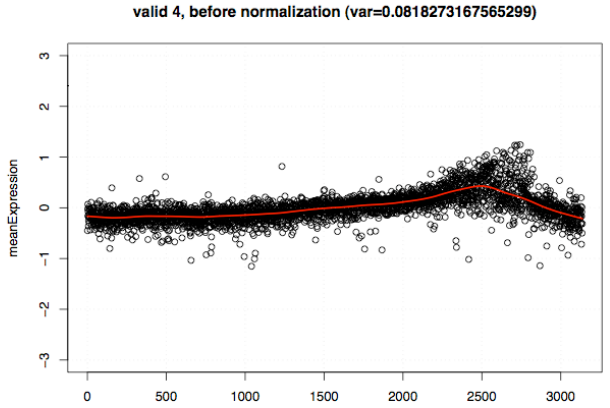




Δ=2h

Δ=4h





Δ=2h

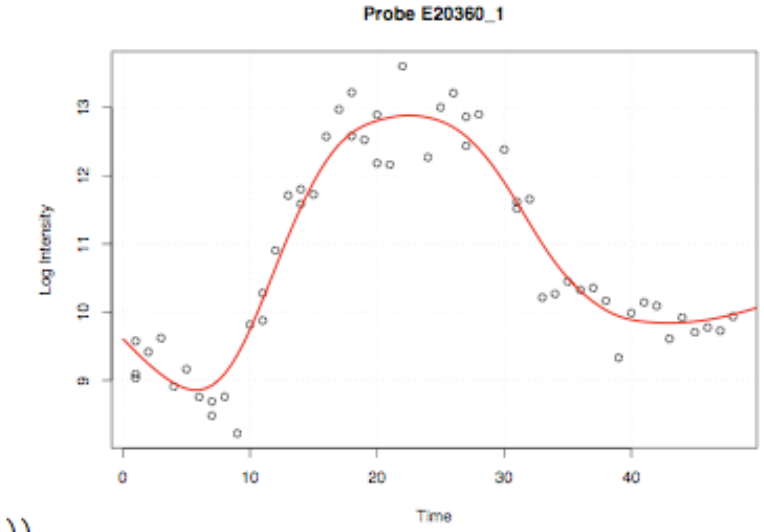
Problem

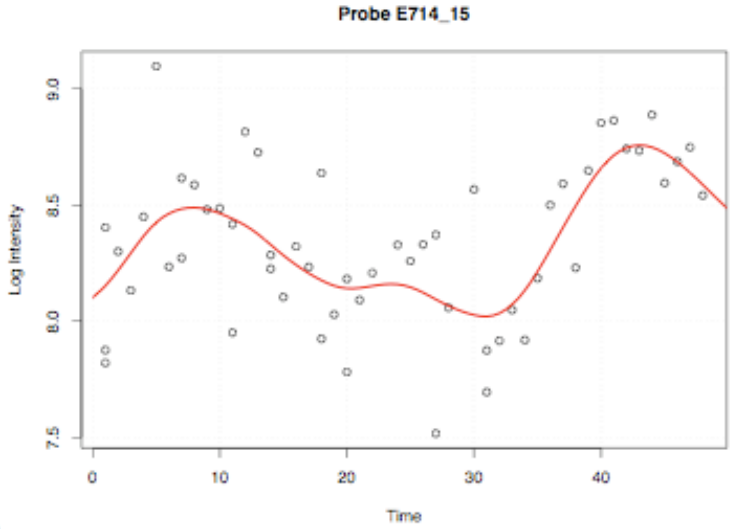
- Most treatments of interest have an effect both on the **gene expression** and on the **cell cycle speed**
- Can we isolate these individual effects?
- Our strategy
  - Use a reference IDC (e.g., de Risi's)
  - Estimate the age of each population
  - Remove the time shift effect based on the estimated ages

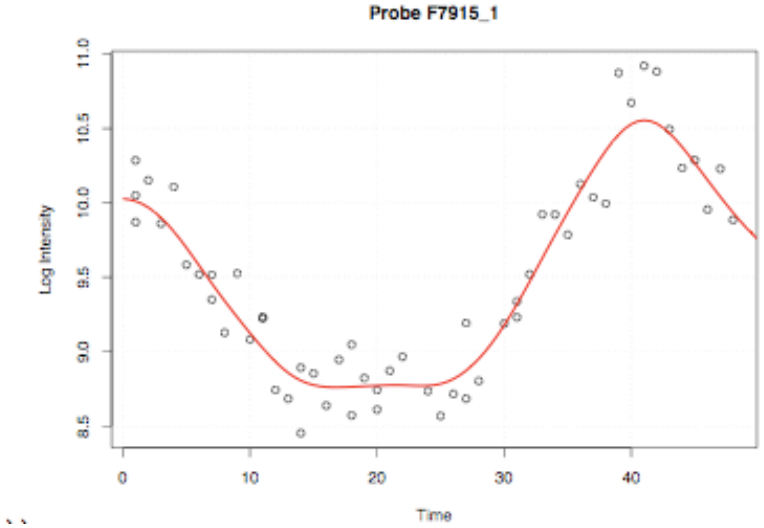
Age estimation

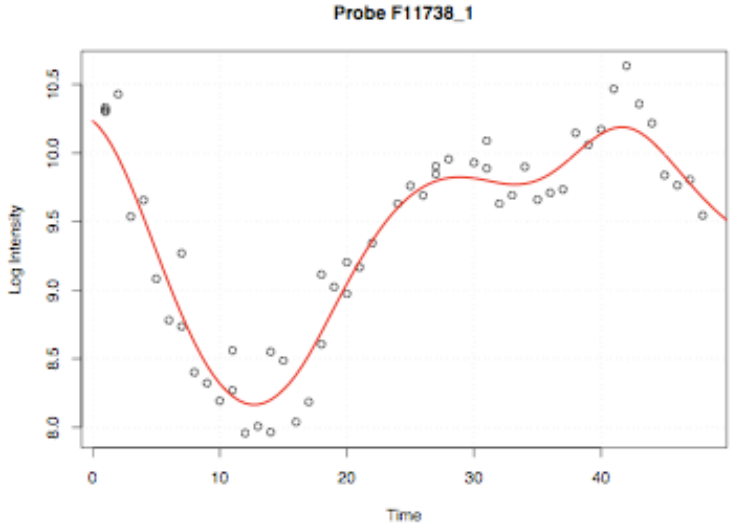
- Problem: given a comparative hybridization data, estimate the age T(red) and T(green) of both channels
- **Method 1 (single-channel)**: estimate each channel separately
- **Method 2 (double-channel)**; estimate both channels simultaneously
- Both methods require a reference cell cycle

Reference cell cycle estimation









Method 1: single channel

$$\begin{aligned}
\text{simple:} \quad & \min_{t} \| f - \psi(t) \| , \\
\text{offset:} \quad & \min_{t,\beta} \| f - \psi(t) - \beta \| , \\
\text{scale:} \quad & \min_{t,\alpha} \| f - \alpha\psi(t) \| , \\
\text{offset/scale:} \quad & \min_{t,\alpha,\beta} \| f - \alpha\psi(t) - \beta \| .
\end{aligned}$$

Pros: do not require both channels to be synchronized (e.g., one pool is ok); no constraint on the amount of phase shift.

Cons: the log-intensity per channel varies a lot!

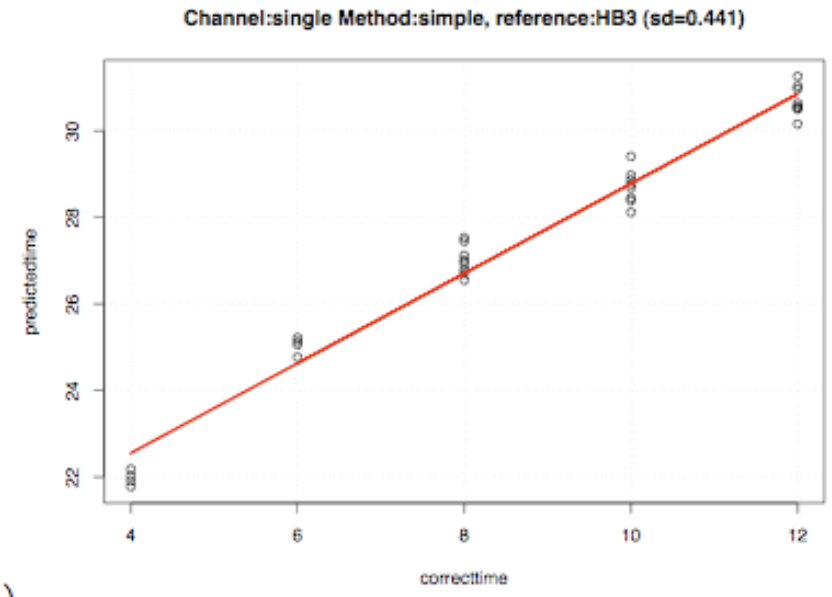
Method 2: double-channel

$$\begin{aligned}
\text{simple:} \quad & \min_{t_1,t_2} \| M - \psi(t_1) + \psi(t_2) \| , \\
\text{offset:} \quad & \min_{t_1,t_2,\beta} \| M - \psi(t_1) + \psi(t_2) + \beta \| , \\
\text{scale:} \quad & \min_{t_1,t_2,\alpha} \| M - \alpha \left(\psi(t_1) + \psi(t_2) \right) \| , \\
\text{offset/scale:} \quad & \min_{t_1,t_2,\alpha,\beta} \| M - \alpha \left(\psi(t_1) + \psi(t_2) \right) + \beta \| .
\end{aligned}$$

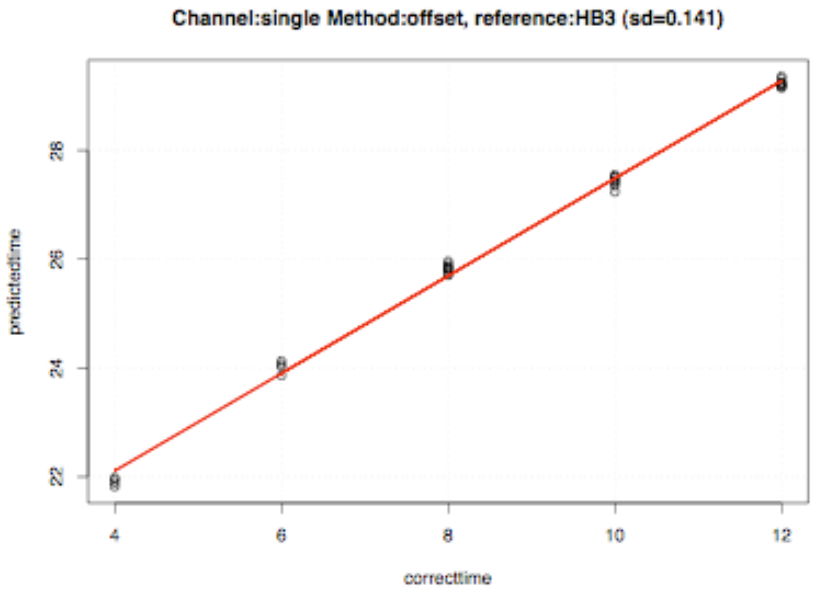
Pros: works with M (more reproducible that log(intensity)).

Cons: Both channels must be synchronized; t1 and t2 must be different enough; allowing a scaling factor is necessary but dangerous.

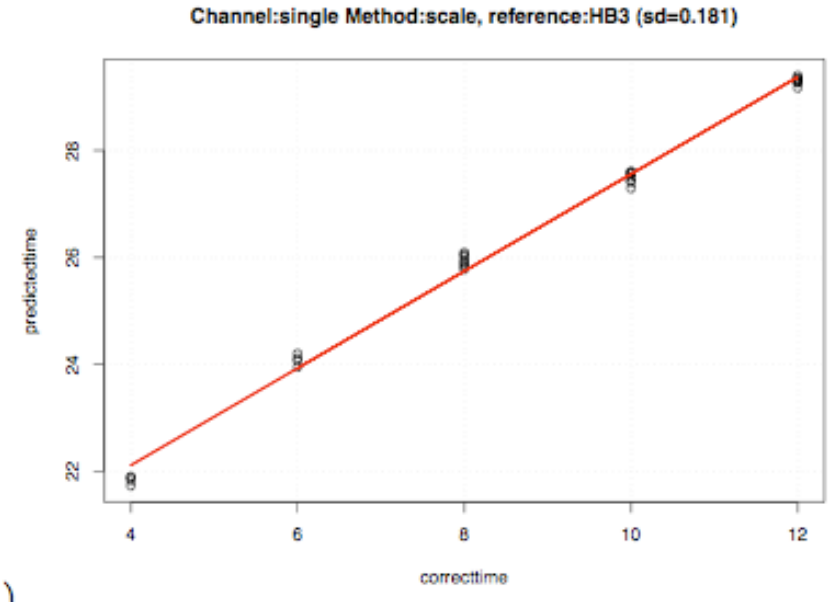
Results: single-channel



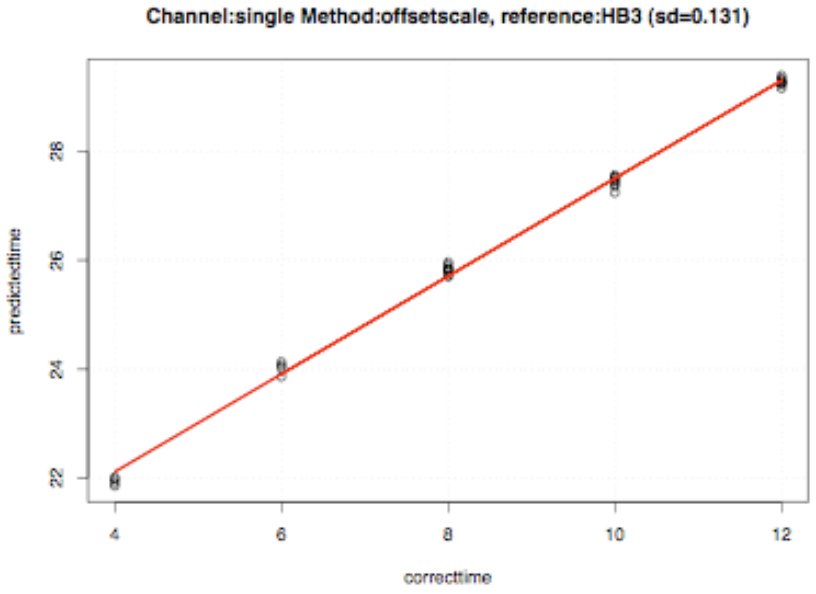
)



))

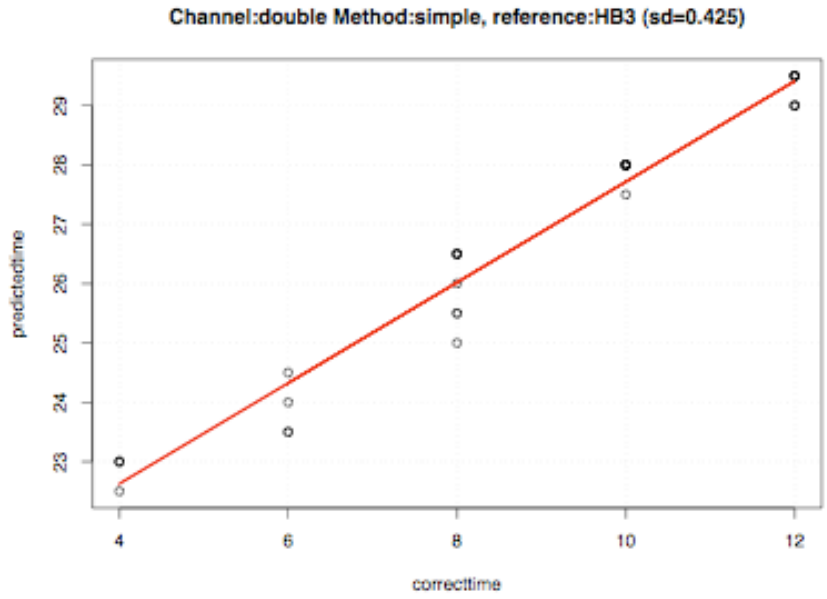


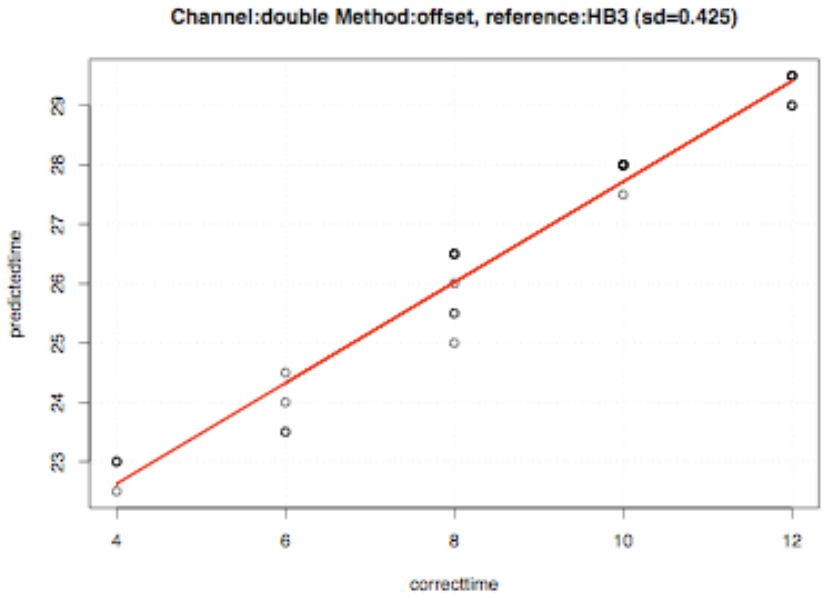
)



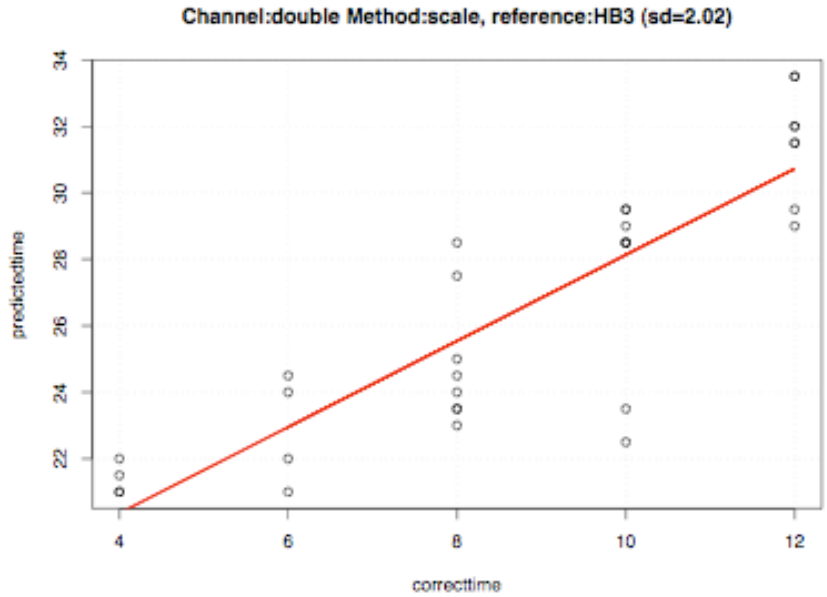
))

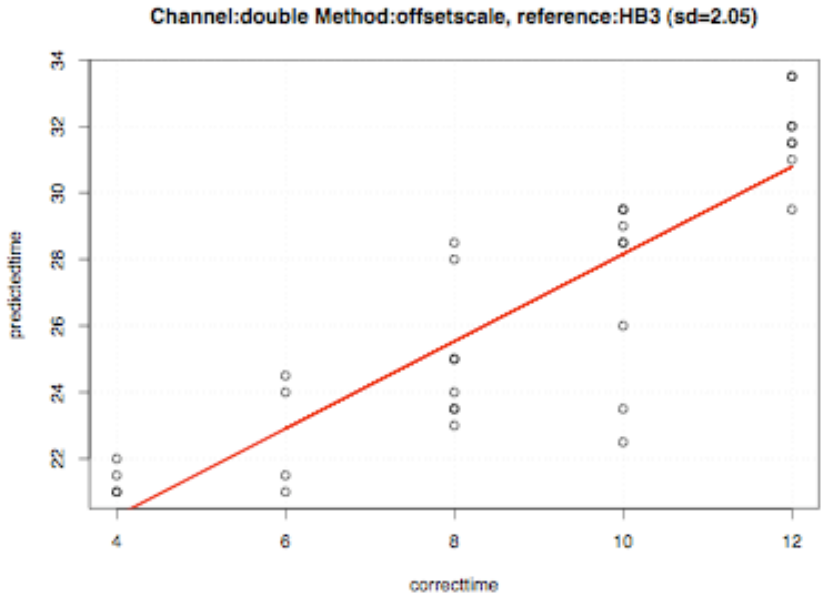
Results: double-channel





))



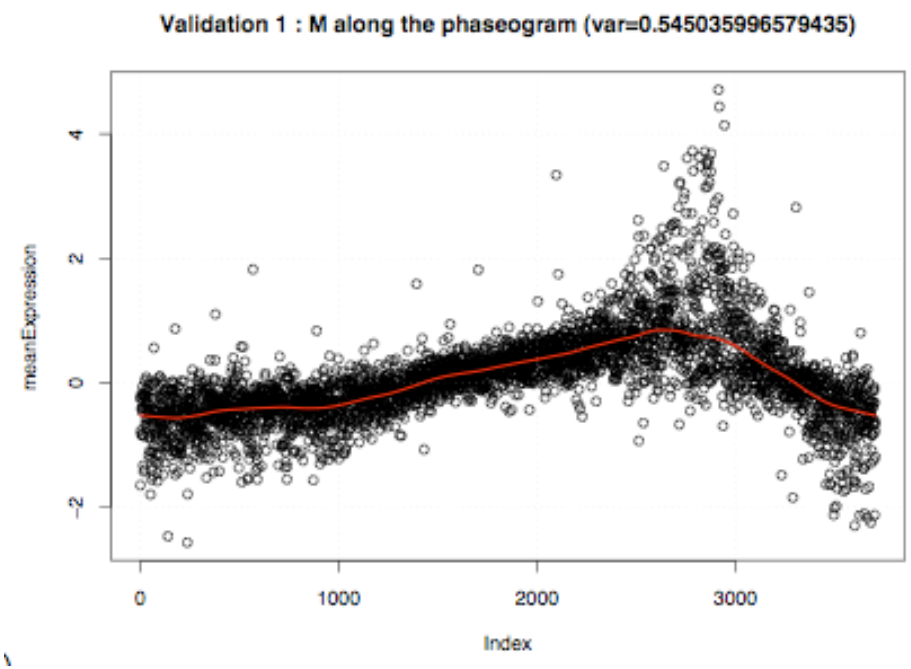


9

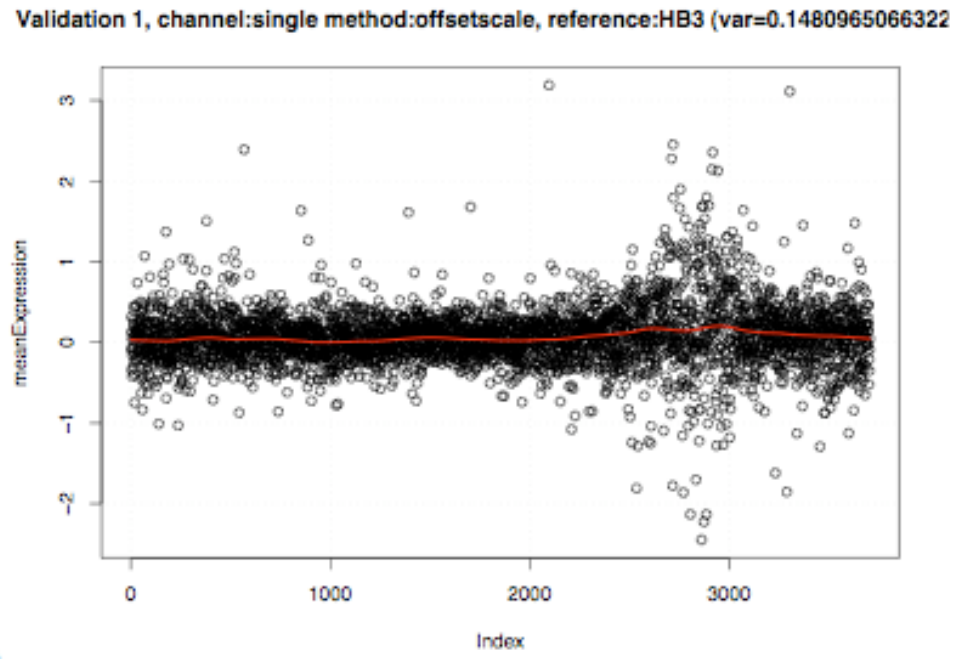
Age estimation: summary

- Single-channel outperforms double-channel
- **Single channel + offsetscale leads to an unbiased age estimation with standard deviation of 10 minutes (on a cycle of 48 hours)**
- Excellent linearity between De Risi's age and time measured at Pasteur.
- Time is slightly faster in Pasteur than in De Risi's lab.

Data normalization



Before

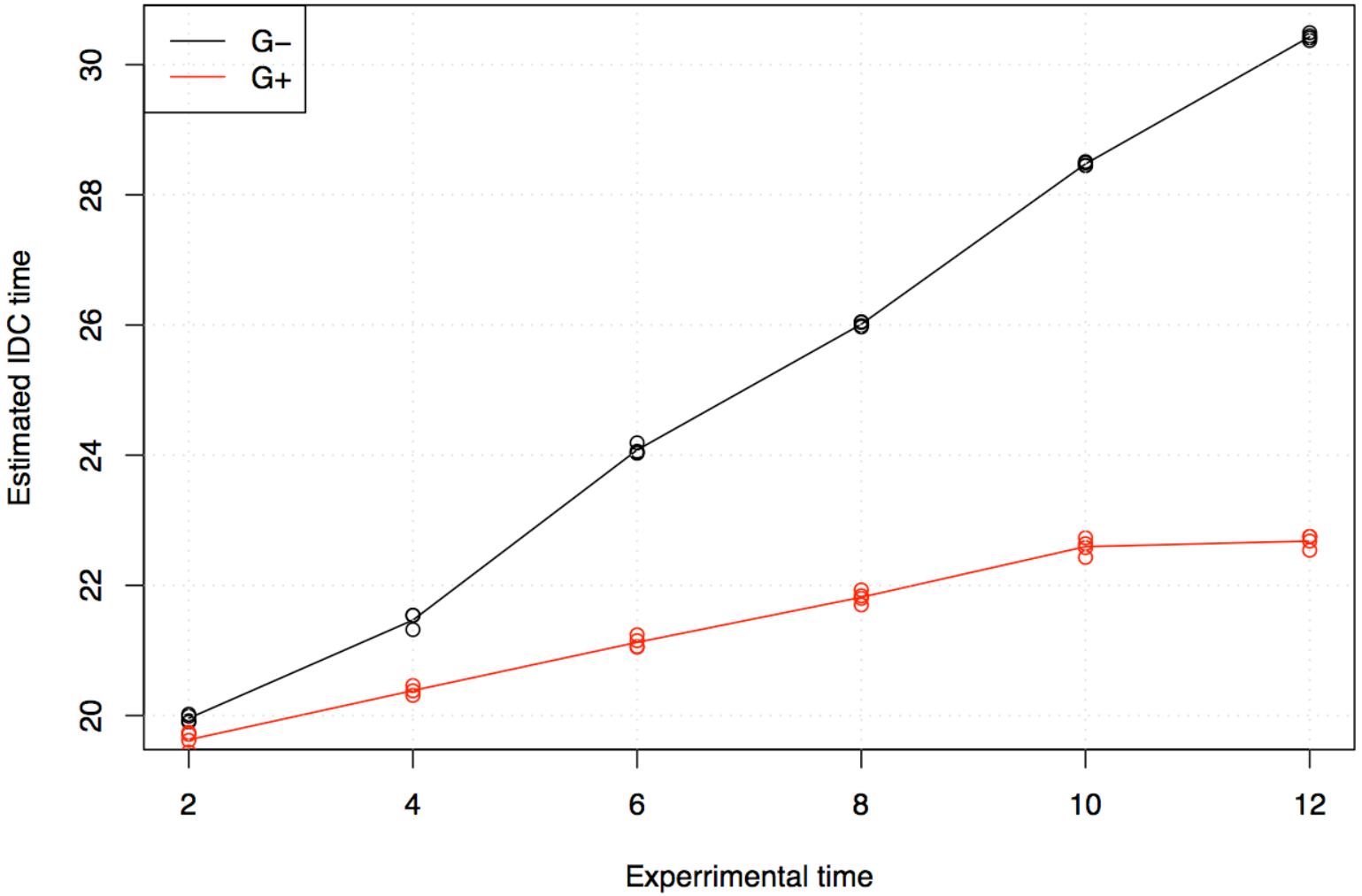


After

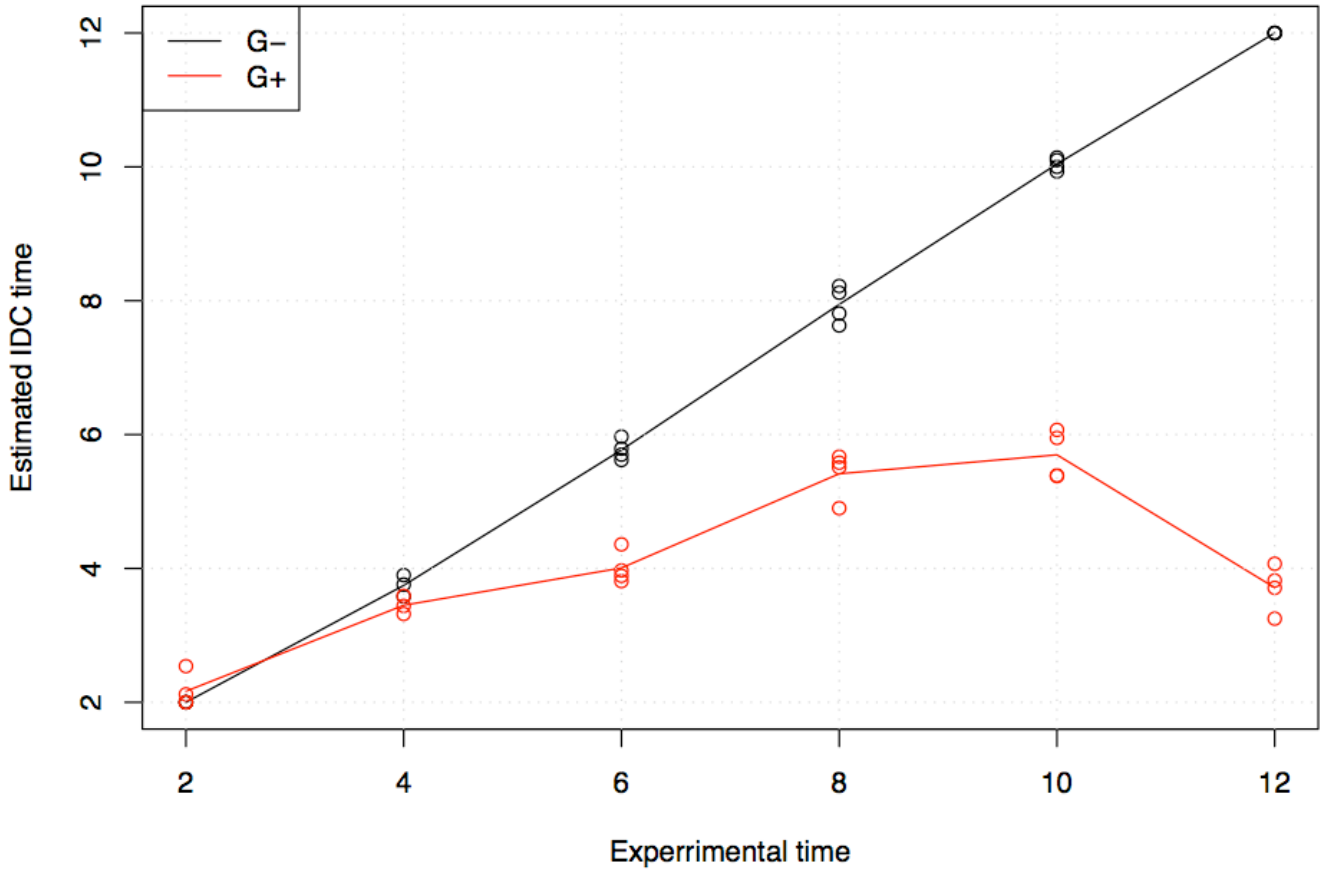
Data normalization: summary

- The main « wave effect » due to phase shift has disappeared
- However many gene remain differentially expressed, especially in the zones of large correction
- Why?
  - Difference between De Risi's and Pasteur's data
  - Imperfection of phase estimation
  - Nonlinearity of the normalization

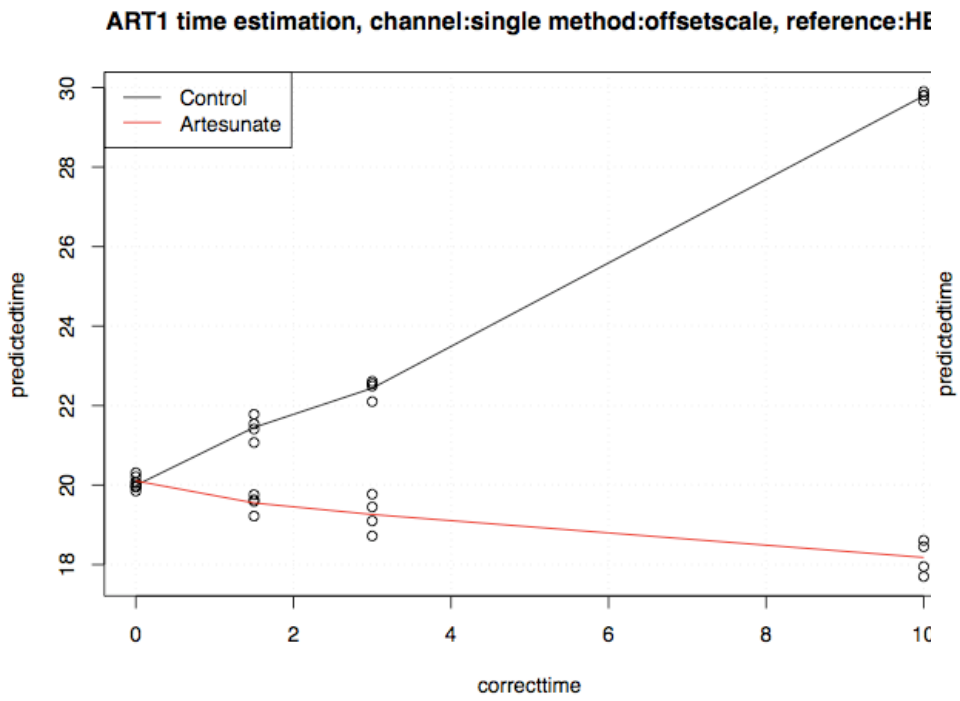
HbAS results: age estimation



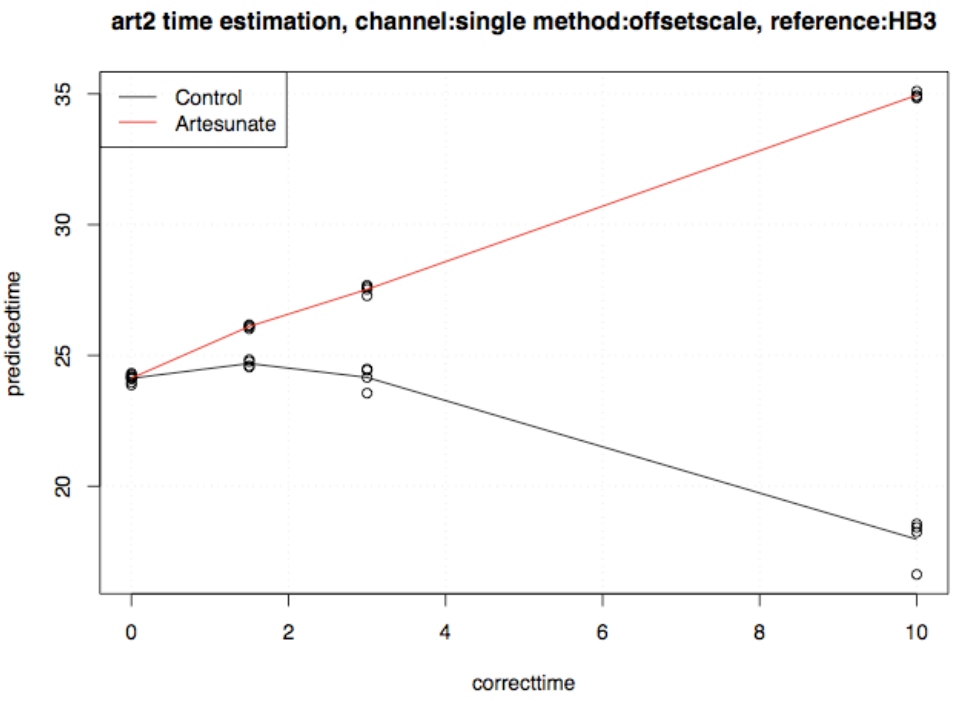
HbAS age estimation with the reference IDC estimated from G-



ART results: age estimation



ART1

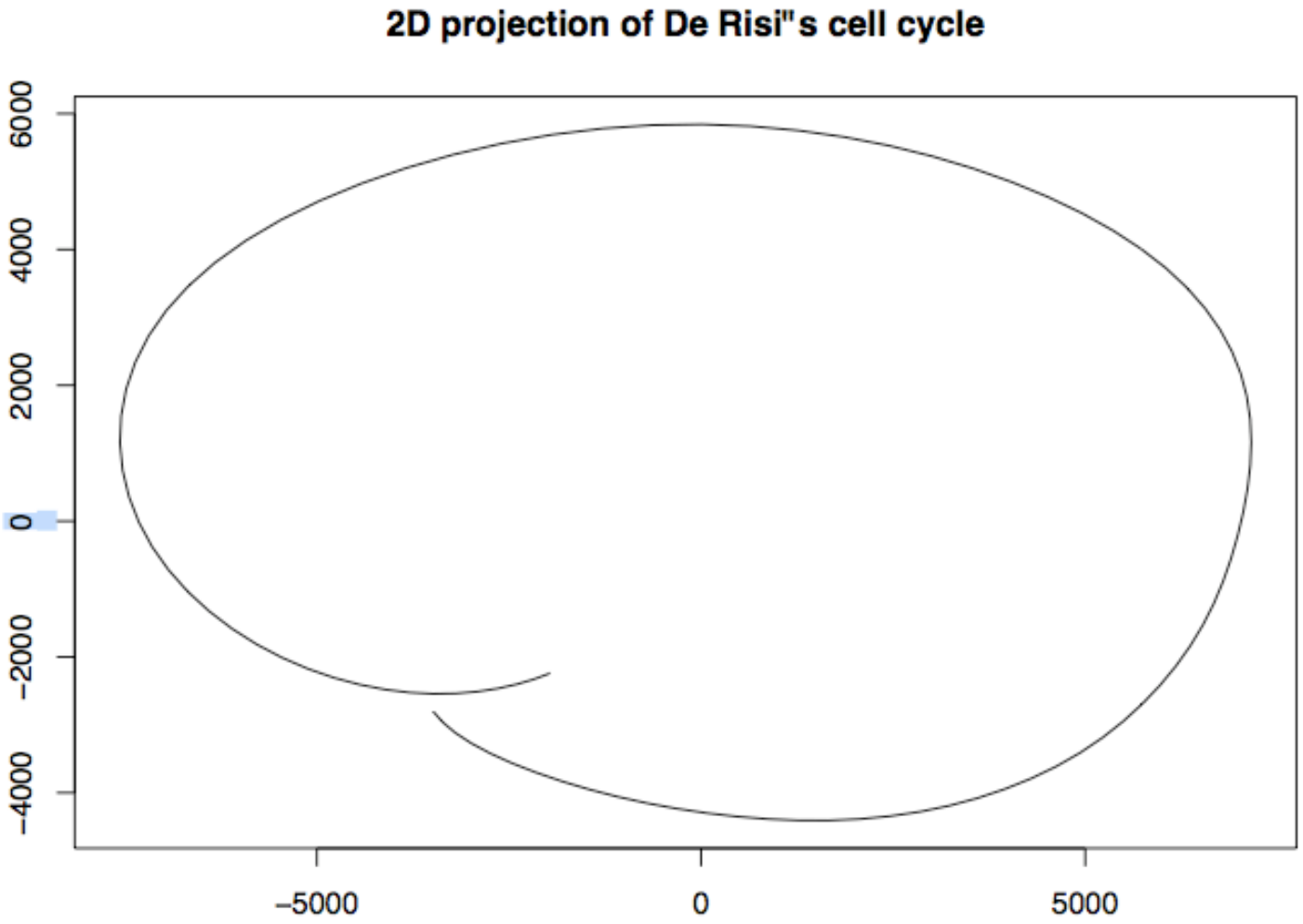


ART2

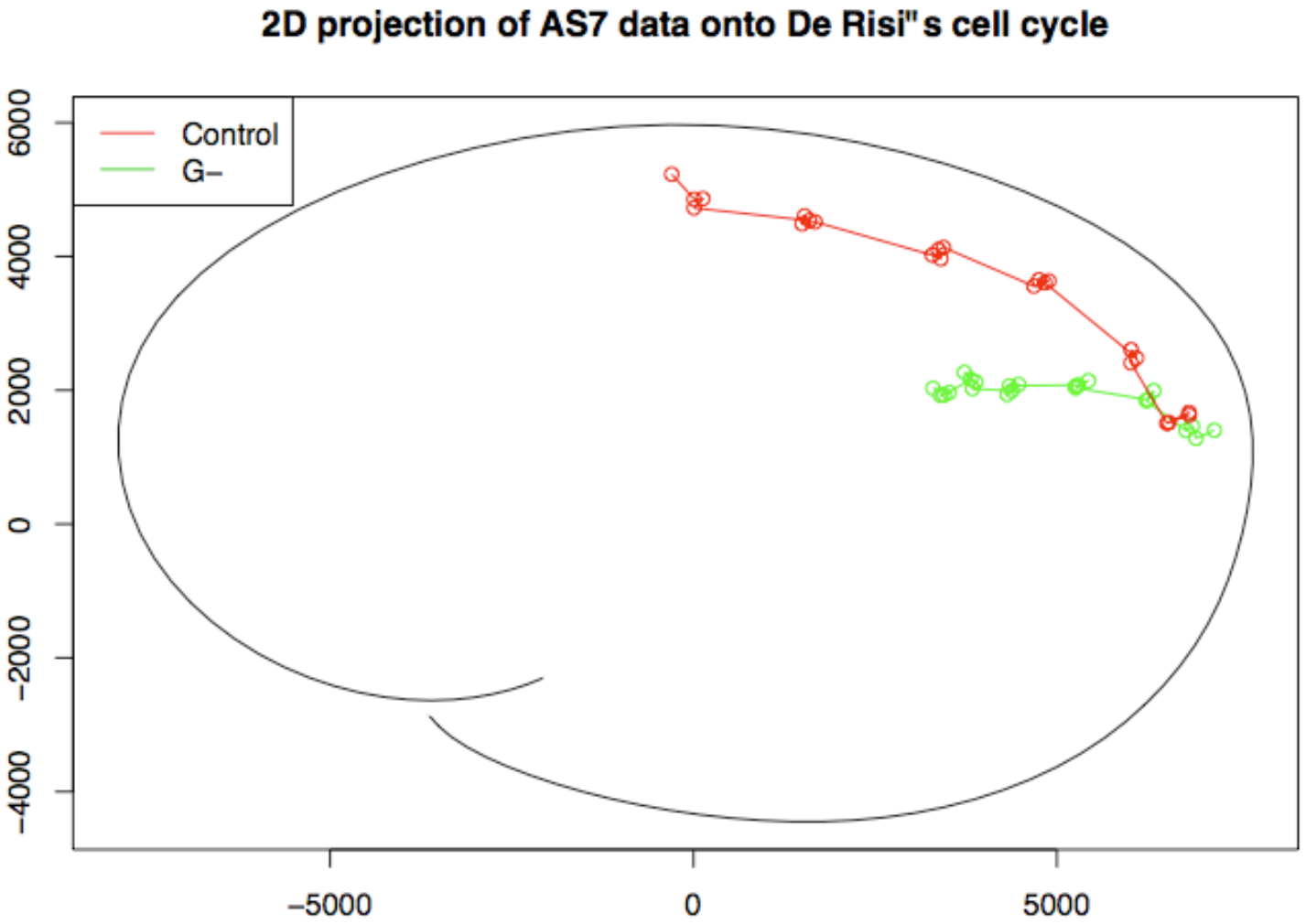
Observation

- Cell cycle seems to slow down in AS (coherent with visual observation)
- Cell cycle seems to go backward in ART1/2 (not coherent with visual observation)
- But are still on the cell cycle trajectory???

De Risi's cell cycle in 2D

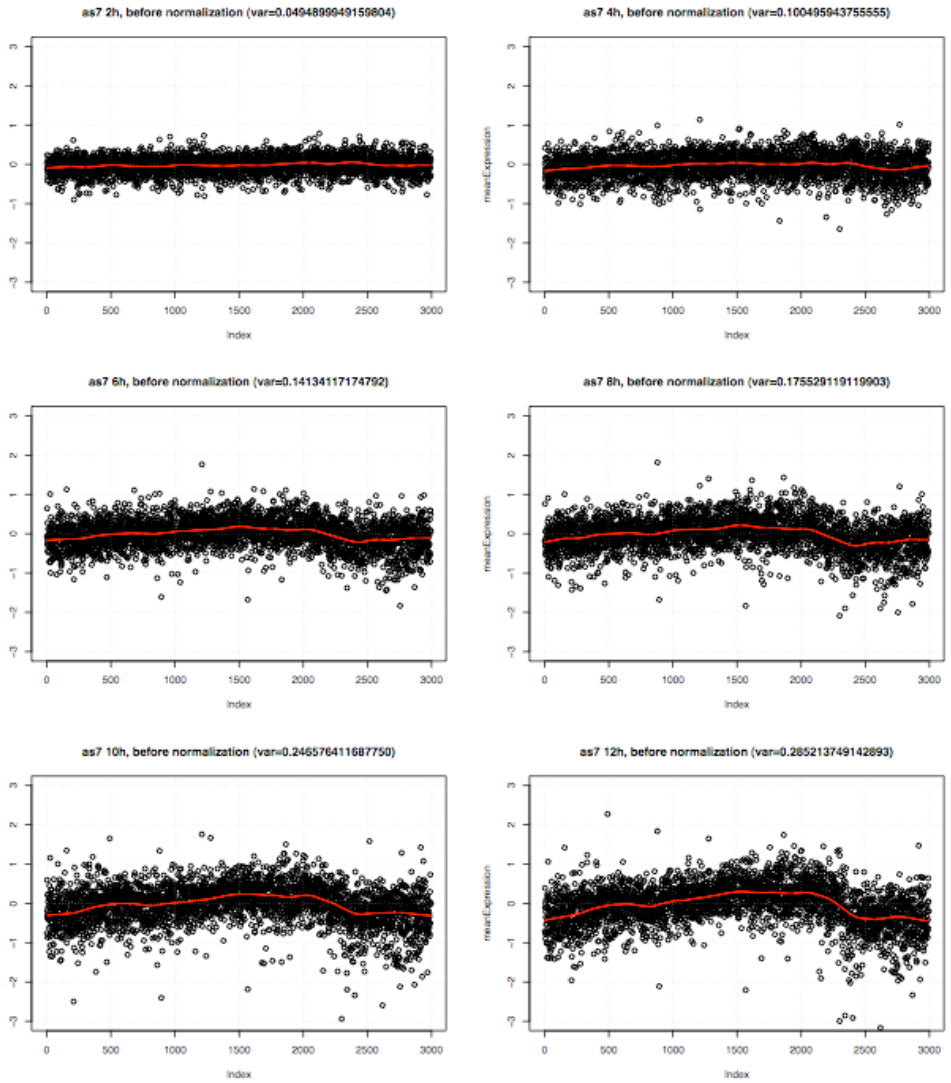


AS experiment in 2D

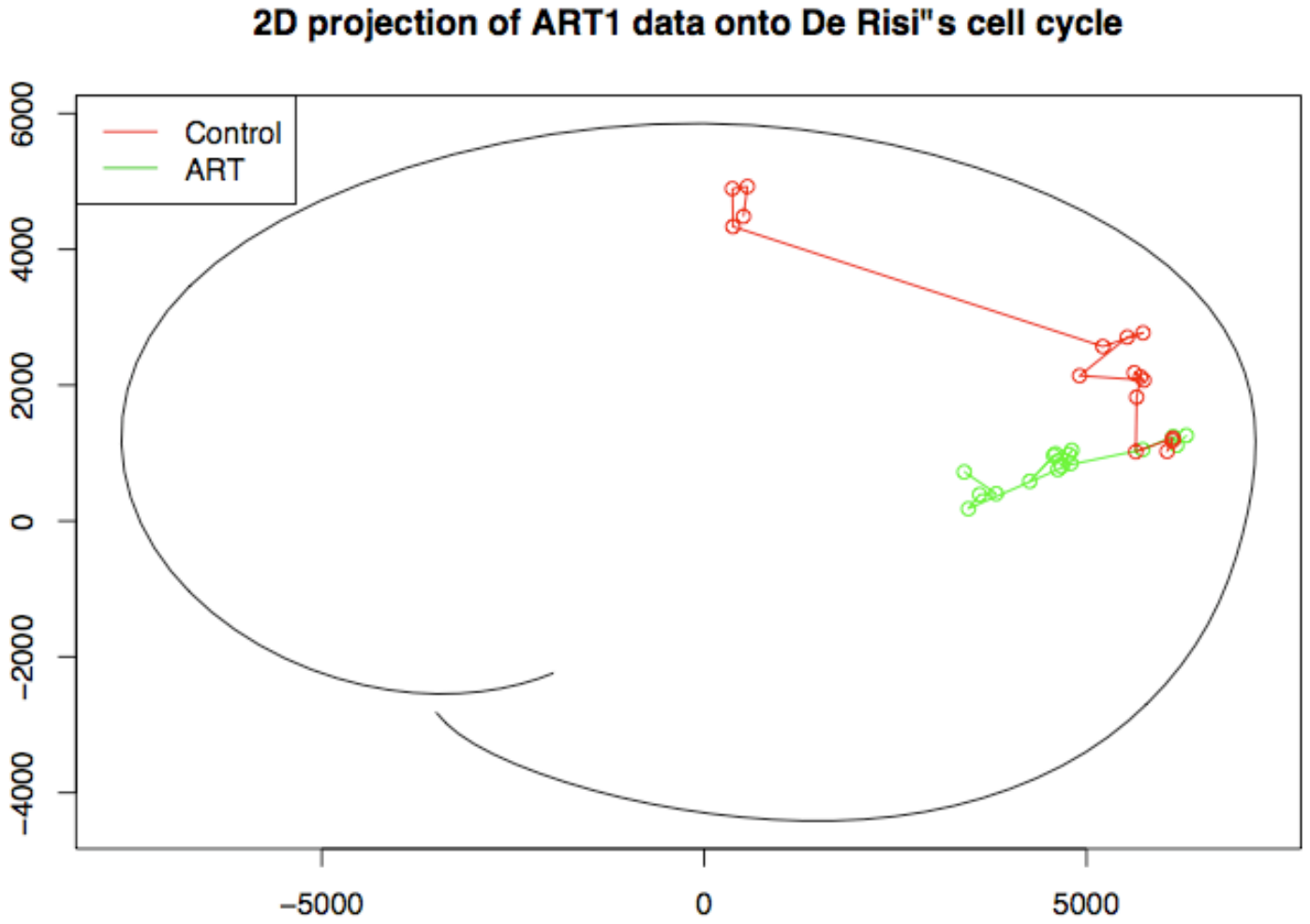


HbAS after age normalization

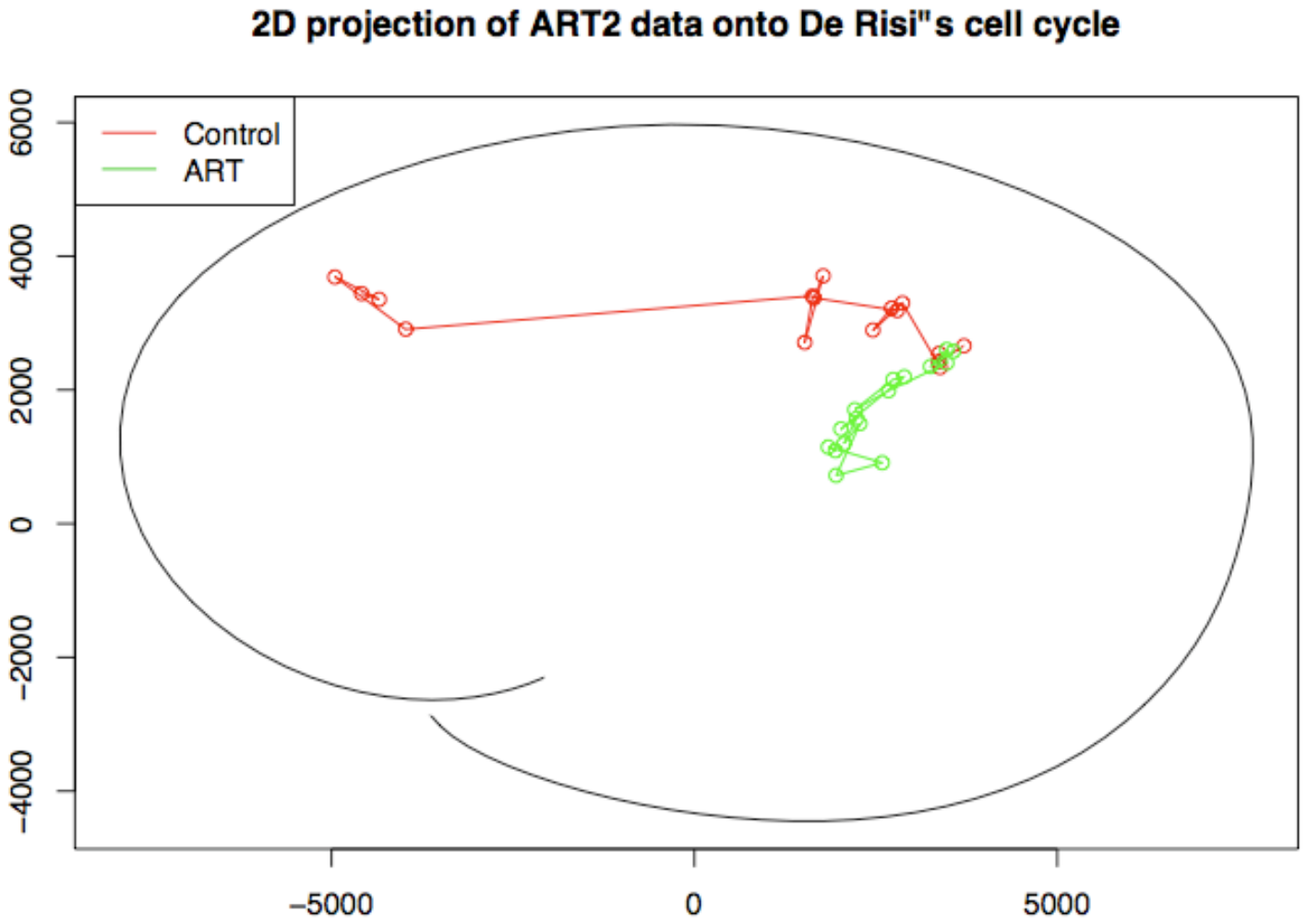
Experiment	Overexpressed	Underexpressed
2h	25	42
4h	41	149
6h	517	651
8h	579	762
10h	602	902
12h	520	794



ART1 experiment in 2D



ART2 experiment in 2D



Where goes the transcriptome?

We observe a **decoupling** of major processes which are usually tightly co-regulated

-Some groups of genes continue to follow the « normal » trend

-Glycolytic pathway, ribonucleotide synthesis *repressed*

-Merozoite genes *induced*

-Some groups of genes stop to follow the « normal » trend

-DNA duplication, proteasome *repressed*

Most disregulated GO terms

UP

Rank	GO ID	GO label	Nb of genes	P-val
1	GO:0004222	metalloendopeptidase activity	98	0.02580335
2	GO:0031072	heat shock protein binding	43	0.0358741
3	GO:0006310	DNA recombination	5	0.03737588
4	GO:0006605	protein targeting	6	0.05129916
5	GO:0004672	protein kinase activity	74	0.07710545
6	GO:0046983	protein dimerization activity	5	0.08376097
7	GO:0008270	zinc ion binding	68	0.09156549
8	GO:0004386	helicase activity	57	0.09322718
9	GO:0000062	acyl-CoA binding	3	0.09616104
10	GO:0006813	potassium ion transport	3	0.0964928

DOWN

Rank	GO ID	GO label	Nb of genes	P-val
1	GO:0004298	threonine endopeptidase activity	14	0.0003070538
2	GO:0005839	proteasome core complex (sensu Eukaryota)	14	0.0003070538
3	GO:0003735	structural constituent of ribosome	121	0.001668131
4	GO:0006511	ubiquitin-dependent protein catabolic process	48	0.001838988
5	GO:0016020	membrane	575	0.002002896
6	GO:0006260	DNA replication	22	0.002537832
7	GO:0020033	antigenic variation	49	0.002757033
8	GO:0007018	microtubule-based movement	22	0.003375036
9	GO:0015031	protein transport	21	0.003968599
10	GO:0007264	small GTPase mediated signal transduction	16	0.003987648

Conclusion

- Precise age estimation is possible, for each channel independently, as long as the transcriptome is « normal »
- Both AS and ART experiments have an strong effect on the transcriptome, which leaves the normal cell cycle trajectory
- The tight regulation between different pathways seems to be lost

Many Thanks to:

- Céline Lacroix
- Emmanuel Bischoff
- Peter David
- Jean-Yves Coppée
- Marie-Agnès Dillies
- Ghislaine Guigon…
- Christian Lajaunie
- Karen Willbrand