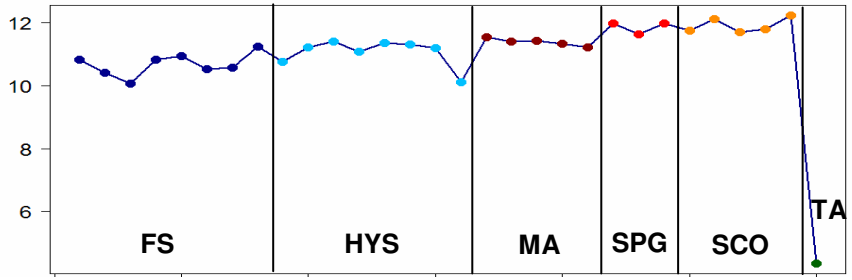
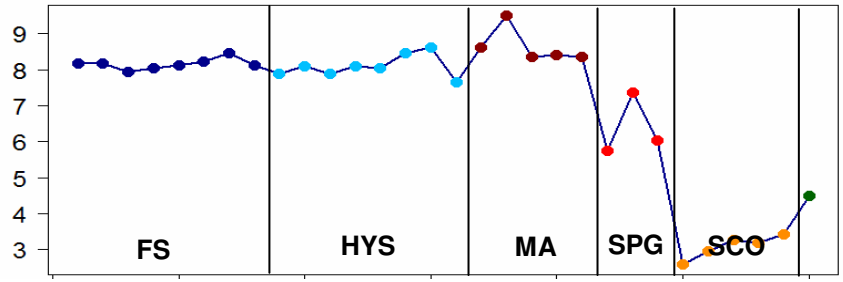
Feasibility of Correlation Filtering
=> Another Example !

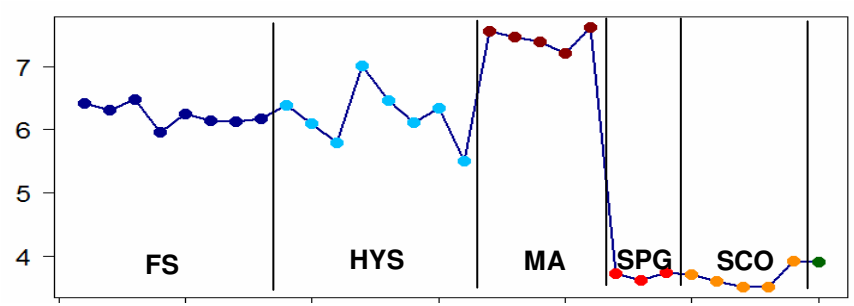
The usual approach to filtering cell type-specific gene expression in human testis: Look at two different states of spermatogenic impairment…



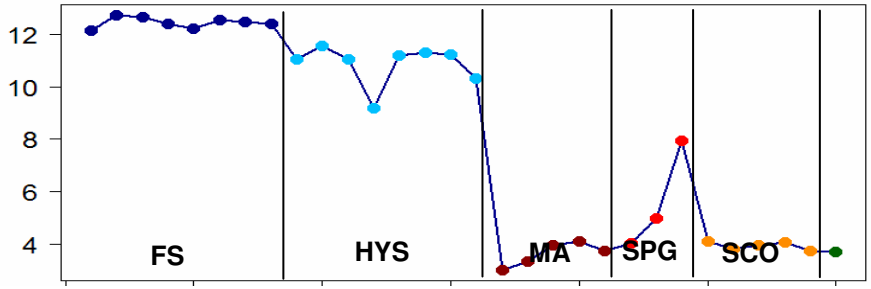
DEFB119 => Sertoli cell specific



FGFR3 => Spermatogonia specific



DMC1 => Spermatocyte specific



HMGB4 => Spermatid specific

Feig et al, Mol Hum Reprod (2007)
Spiess et al, Hum Reprod (2007)
Von Kopylow et al, Hum Reprod (2010)
Cappallo-Obermann et al, Hum Genet (2010)
Cappallo-Obermann et al, Hum Reprod (2013)

Feasibility of Correlation Filtering
=> Another Example !

But how to proceed if there is no differential setup available?

• **Leydig cells:** There is no human testis pathology in which the only difference are Leydig/Peritubular Myoid/testicular macrophages/mast cells etc. For example: No phenotype where Leydig cells are missing and germ cells are present !

Unlike to rat models: EDS (ablates Leydig cells) +
T (preserves spermatogenesis)

=> See O'Shaughnessy et al. (2014)

Consequence: We need another readout for cell specificity other than a classical +/- situation !

And that is…

Feasibility of Correlation Filtering
=> Another Example !

… the expression levels over a large cohort of samples.

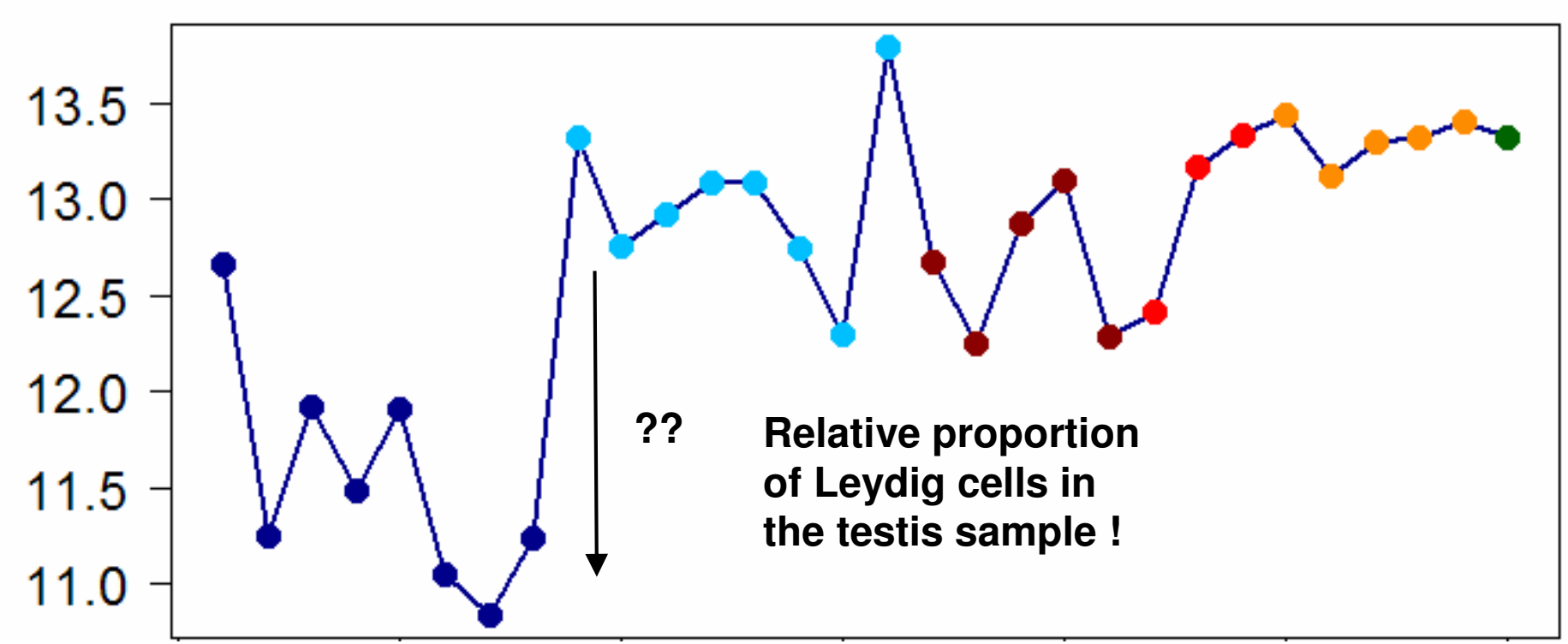
Other approaches:

- **Transcriptomics on isolated single cells.**
  Advantage: Very pure data
  Disadvantage: Amplification may introduce bias

- **Enrichment of a cell population**
  Disadvantage: Contamination, Isolation effects

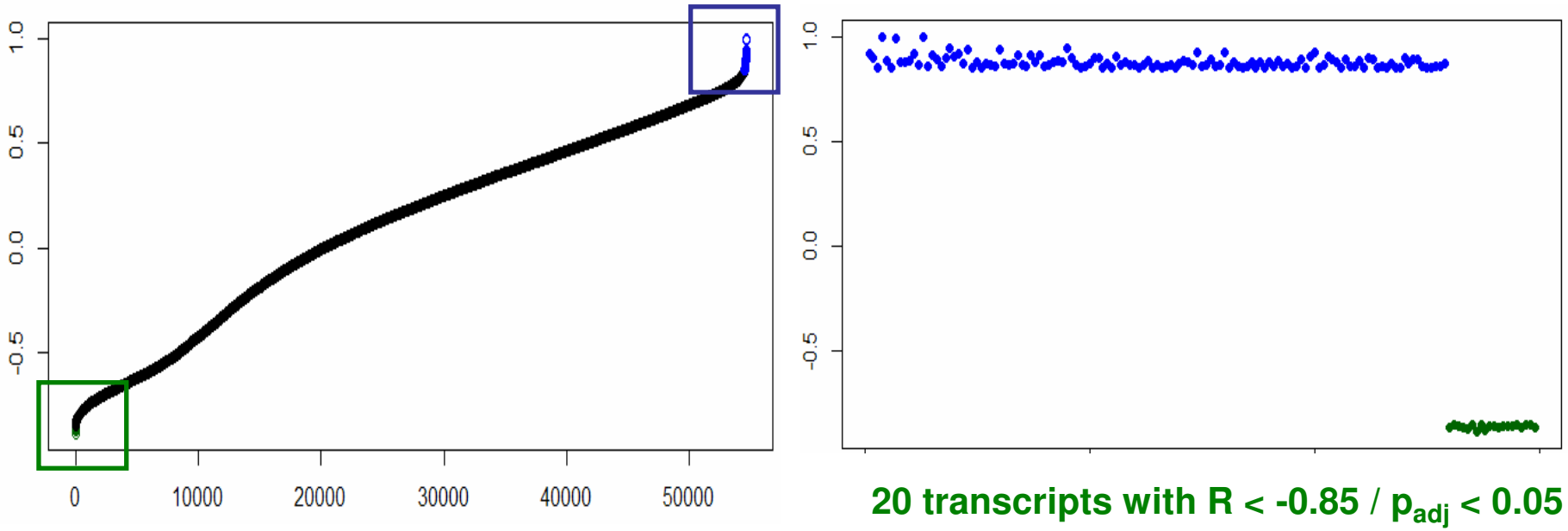
Paradigm: A high similarity of expression to the „template gene“ over all samples is an indicator for expression in the same cell type !

Step 1: Select cell-specific gene as a template, i.e. INSL3



Step 2: Calculate Pearson/Spearman correlation to all 55000 transcripts/Probesets on the array

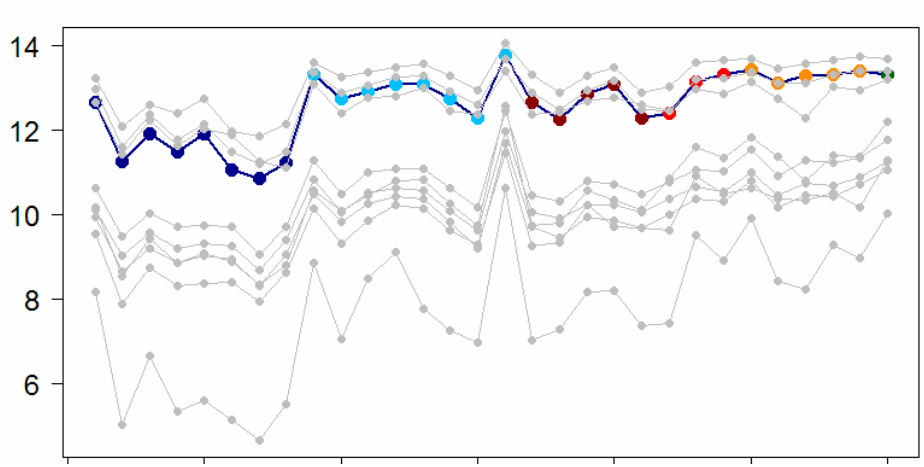
⇒ **70 sec on a standard laptop using the *R* programming language**



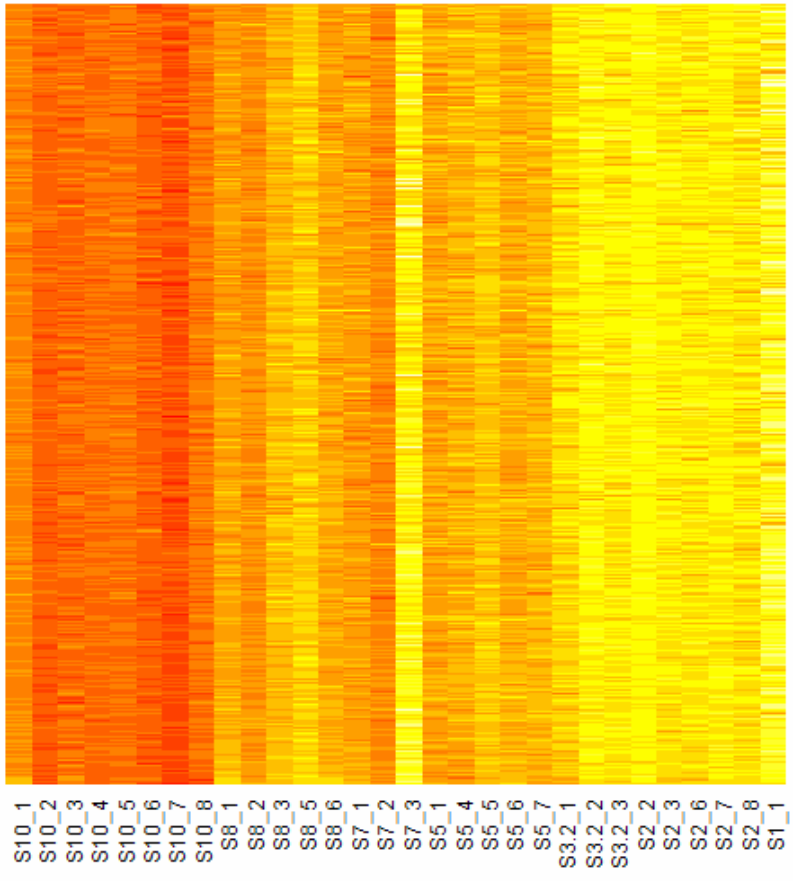
Månsson R, Tsapogas P, Akerlund M, Lagergren A, Gisler R, Sigvardsson M.
Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor.
J Biol Chem. 2004 Apr 23;279(17):17905-13.

Step 3: Inspect genes with highest positive/negative correlation

Profile plot



Heatmap



Step 4: Downstream analysis, i.e. GO-Terms

Term	RT	Genes	Count	%	P-Value	Benjamini
steroid biosynthetic process	RT		9	8,3	8,0E-8	4,0E-5
steroid metabolic process	RT		12	11,0	8,0E-8	2,0E-5
cellular hormone metabolic process	RT		7	6,4	2,3E-6	3,8E-4
response to hydrogen peroxide	RT		6	5,5	3,2E-5	4,0E-3
hydrogen peroxide catabolic process	RT		4	3,7	1,8E-4	1,7E-2
sterol metabolic process	RT		6	5,5	5,3E-4	4,3E-2
cellular response to reactive oxygen species	RT		4	3,7	9,8E-4	6,8E-2
hormone biosynthetic process	RT		4	3,7	1,1E-3	6,5E-2
androgen metabolic process	RT		3	2,8	2,3E-3	1,2E-1
coenzyme catabolic process	RT		3	2,8	1,3E-2	4,7E-1
androgen catabolic process	RT		2	1,8	1,3E-2	4,5E-1
acetyl-CoA metabolic process	RT		3	2,8	1,8E-2	5,2E-1
cellular respiration	RT		4	3,7	2,6E-2	6,4E-1
positive regulation of cell division	RT		3	2,8	2,7E-2	6,2E-1
androgen biosynthetic process	RT		2	1,8	3,3E-2	6,7E-1
peptide cross-linking via chondroitin 4-sulfate glycosaminoglycan	RT		2	1,8	3,9E-2	7,1E-1
development of primary sexual characteristics	RT		4	3,7	5,1E-2	7,9E-1
hormone catabolic process	RT		2	1,8	5,8E-2	8,1E-1
glucocorticoid metabolic process	RT		2	1,8	6,4E-2	8,2E-1
cell redox homeostasis	RT		3	2,8	6,5E-2	8,1E-1
development of primary male sexual characteristics	RT		3	2,8	6,8E-2	8,1E-1
carboxylic acid metabolic process	RT		8	7,3	7,3E-2	8,2E-1
male sex differentiation	RT		3	2,8	8,4E-2	8,5E-1
chondroitin sulfate metabolic process	RT		2	1,8	8,8E-2	8,5E-1