Manual Classification of Early Enhancer, Promoter and Transcription Factor Response Dynamics

Kristoffer Vitting-Seerup
(kristoffer.vittingseerup@bio.ku.dk)
Robin Andersson
Albin Sandelin

- This work is heavily inspired by (but not meant to replace):
    - Erik, Robin and Olga’s fuzzy clustering + meta clusters
    - Stuart and Collin’s clustering of kinetic signatures

Definitions

- Response Dynamics = Log2 FC vs t0 (measure of changes compared to t0)
- TFs = Transcription factors
- Feature = A common term for Enhancers, promoters and TFs.

Goal

- Analyse unsupervised clustering of significantly differentially expressed features to find all interesting response dynamics
- To define a set of restrictions that mimics the found response dynamics thereby creating a classification scheme
- Use this classification scheme to analyse response dynamics across many time courses.

Motivation

- This approach have several advantages:
    - By using temporal cutoffs (measured in minutes) different time-courses with different time points can easily be compared
    - By using FC cutoffs we can be certain that all members of the groups
        - Behaves similar (instead of some features just being put in a cluster because it needs to be clustered somewhere)
        - Have defined size response (meaning we filter out background and small changes)
    - It can be used on both enhancers, promoters and TFs
    - It’s simple and easy to understand

Identifying different types of response dynamics

- The response dynamics (log2 FC vs t0) of significant deregulated (amongst all pairwise comparisons) enhancers, promoters and TF were clustered (K-means, k = 16) for all time-course with early time points.

- The time-courses included were:

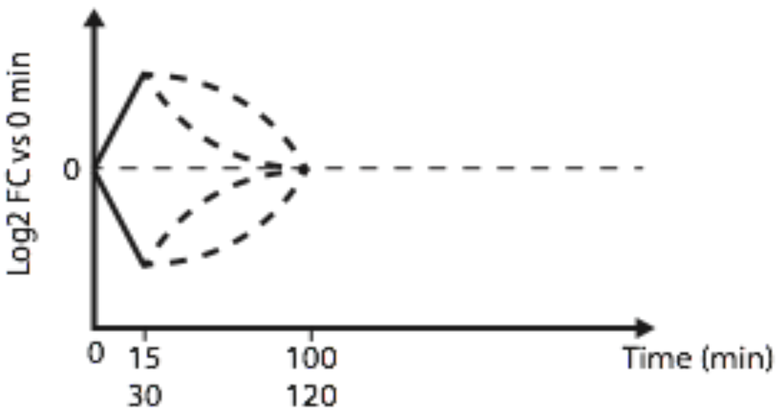
1. "human_Monocyte-derived_macrophages_response_to_LPS",
2. 'human_Lymphatic_Endothelial_cells_response_to_VEGFC',
3. 'human_MCF7_breast_cancer_cell_line_HRG',
4. 'human_Aortic_smooth_muscle_cell_IL1b',
5. 'human_ARPE-19',
6. 'human_MCF7_breast_cancer_cell_line_EGF1',
7. 'human_Saos-2_osteosarcoma',
8. 'human_mesenchymal_stem_cells_adipose_derived',
9. 'human_Aortic_smooth_muscle_cell_FGF2',
10. 'mouse_J2E',
11. 'mouse_ST2_adipocytes',
12. ‘mouse_ST2_osteocytes',
13. ‘mouse_EBF_KO_HPCs_induced_to_T_cell

- Since we are interested in the early dynamics only time points up to (and including) 360 min (6 hrs) were considered.

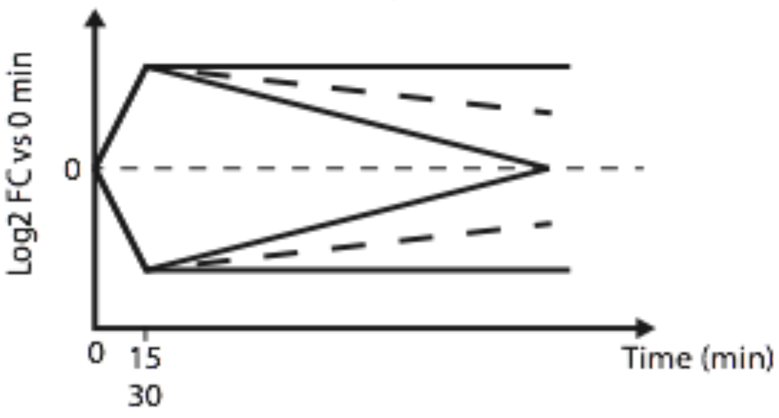
- By manually inspecting all of these clusters of dynamics a set of interesting dynamics were defined

Interesting response dynamics

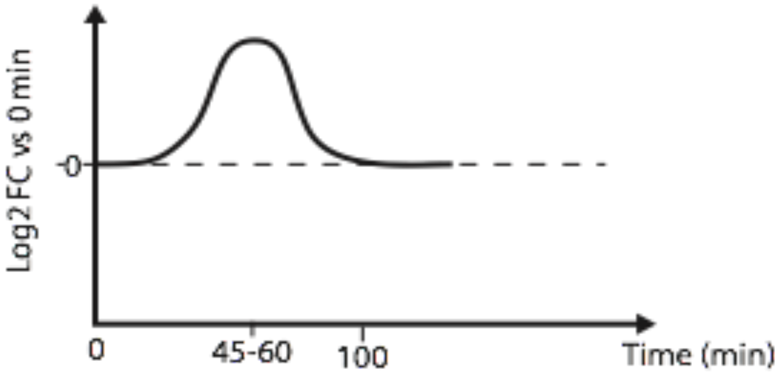
- 6 types of response dynamics (12 if direction is considered) were found
- These are (for now) classified hierarchically as shown by the numbering below - meaning each feature is classified as the first for these classes it matches.



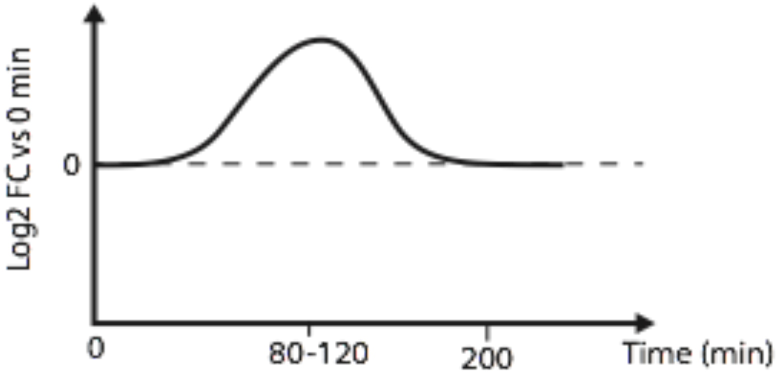
1) Rapid Short Response



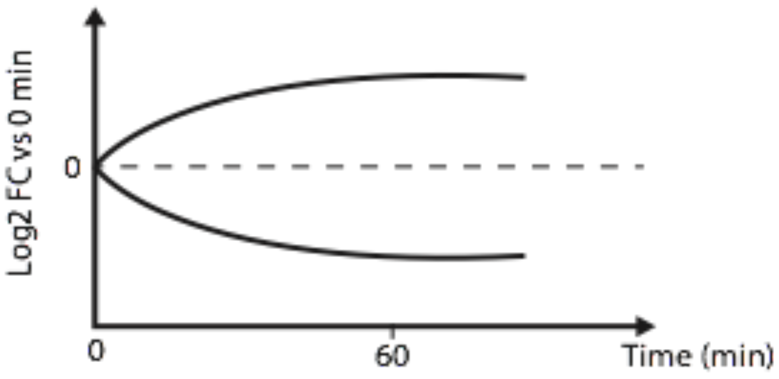
2) Rapid Long Response



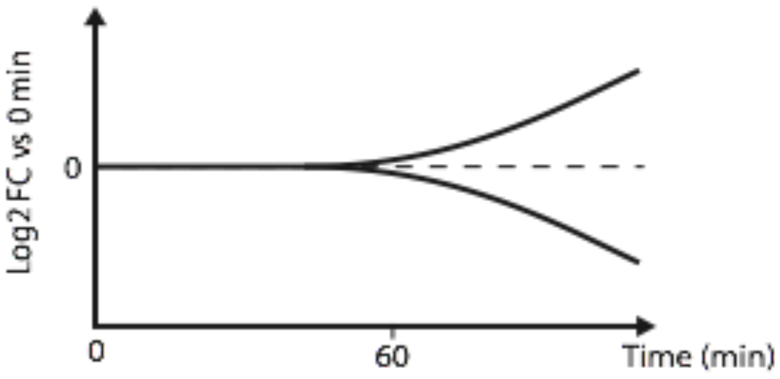
3) Early Standard Response



4) Late Standard Response



5) Slow Long Respons



6) Late Upregulation

Data

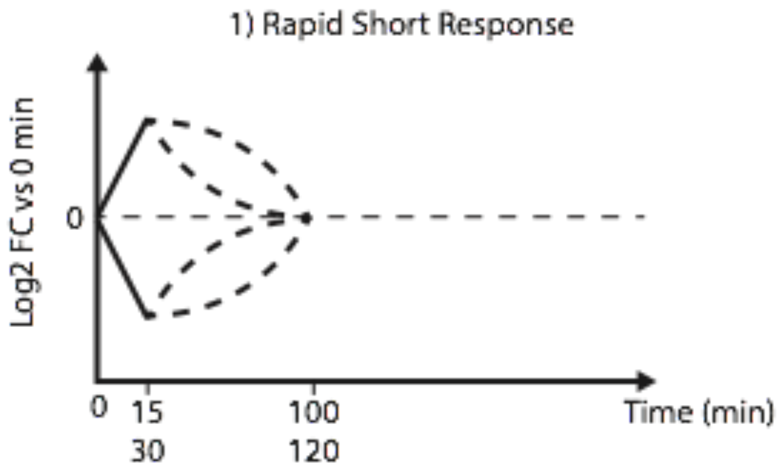
- Only features significantly differentially expressed (amongst all pairwise comparisons) up to and including 360 min were included in the following analysis.

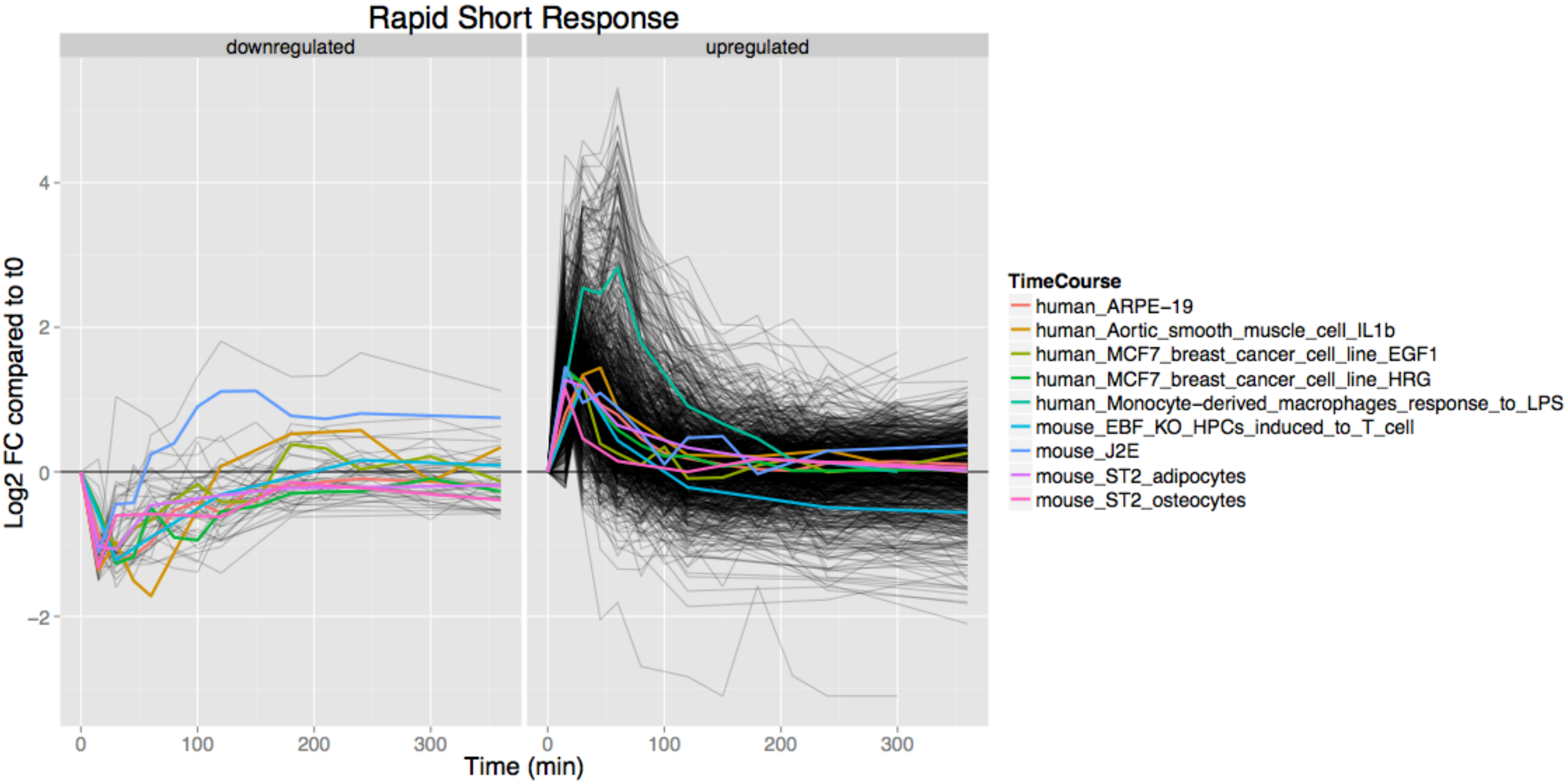
- Only time courses with more than 30 significantly differentially expressed enhancers, promoters and TFs were included, resulting in this list:
- ‘human_Monocyte-derived_macrophages_response_to_LPS',
- 'human_MCF7_breast_cancer_cell_line_HRG',
- 'human_Aortic_smooth_muscle_cell_IL1b',
- 'human_ARPE-19',
- 'human_MCF7_breast_cancer_cell_line_EGF1',
- 'mouse_J2E',
- 'mouse_ST2_adipocytes',
- ‘mouse_ST2_osteocytes',
- ‘mouse_EBF_KO_HPCs_induced_to_T_cell

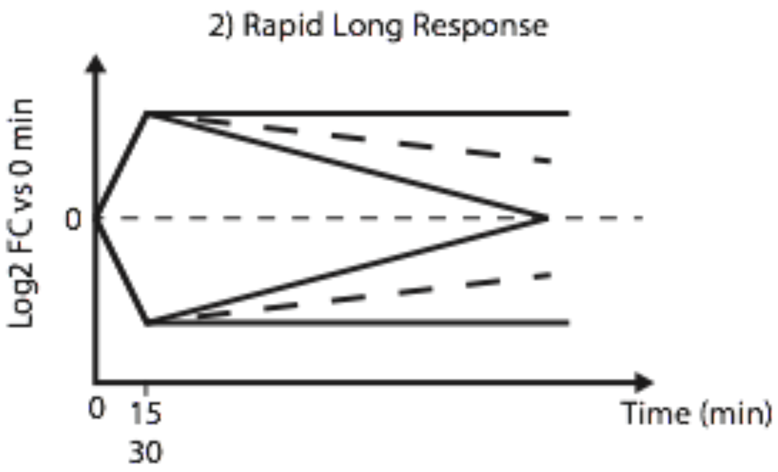
- Expression is measured as the mean of replicates

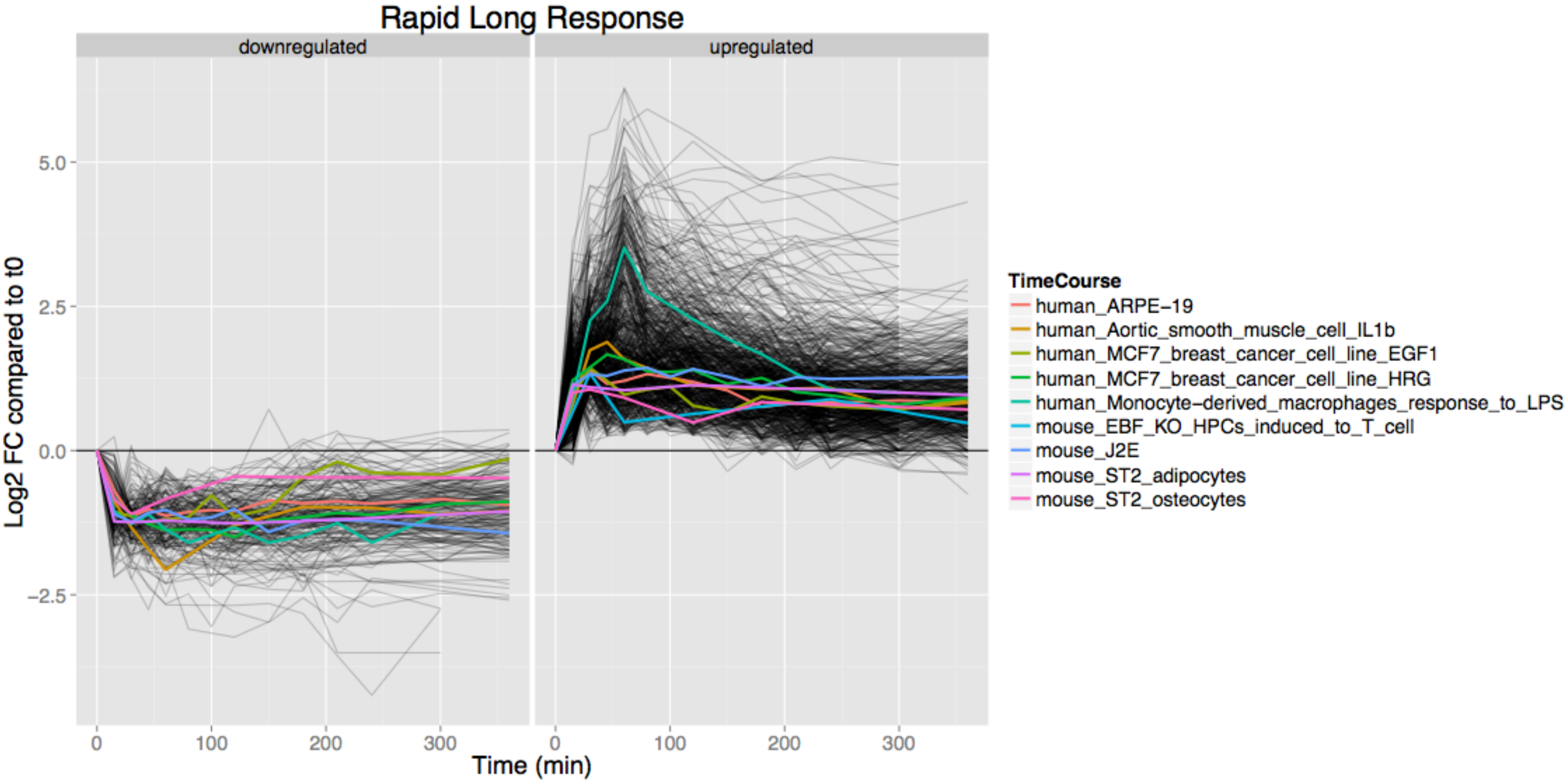
- In the following the promoter dataset does not contain the TFs

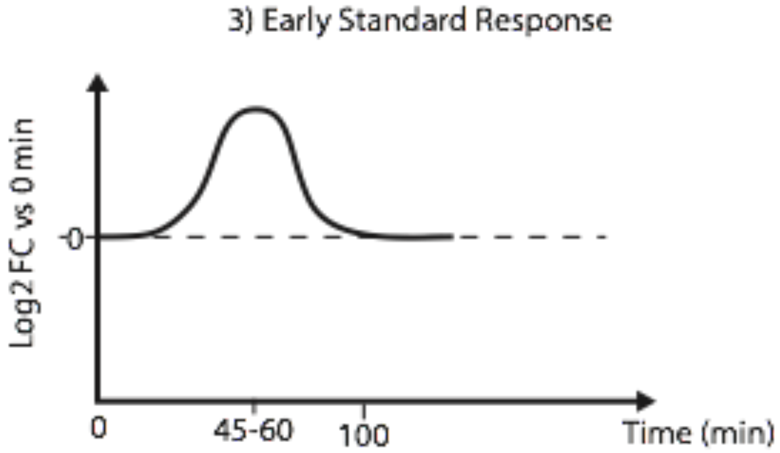
- Note that not all time courses have all classes so if a time course is missing in the following it’s due to no features classified.


1) Rapid Short Response
Log2 FC vs 0 min
0
0
15
30
100
120
Time (min)

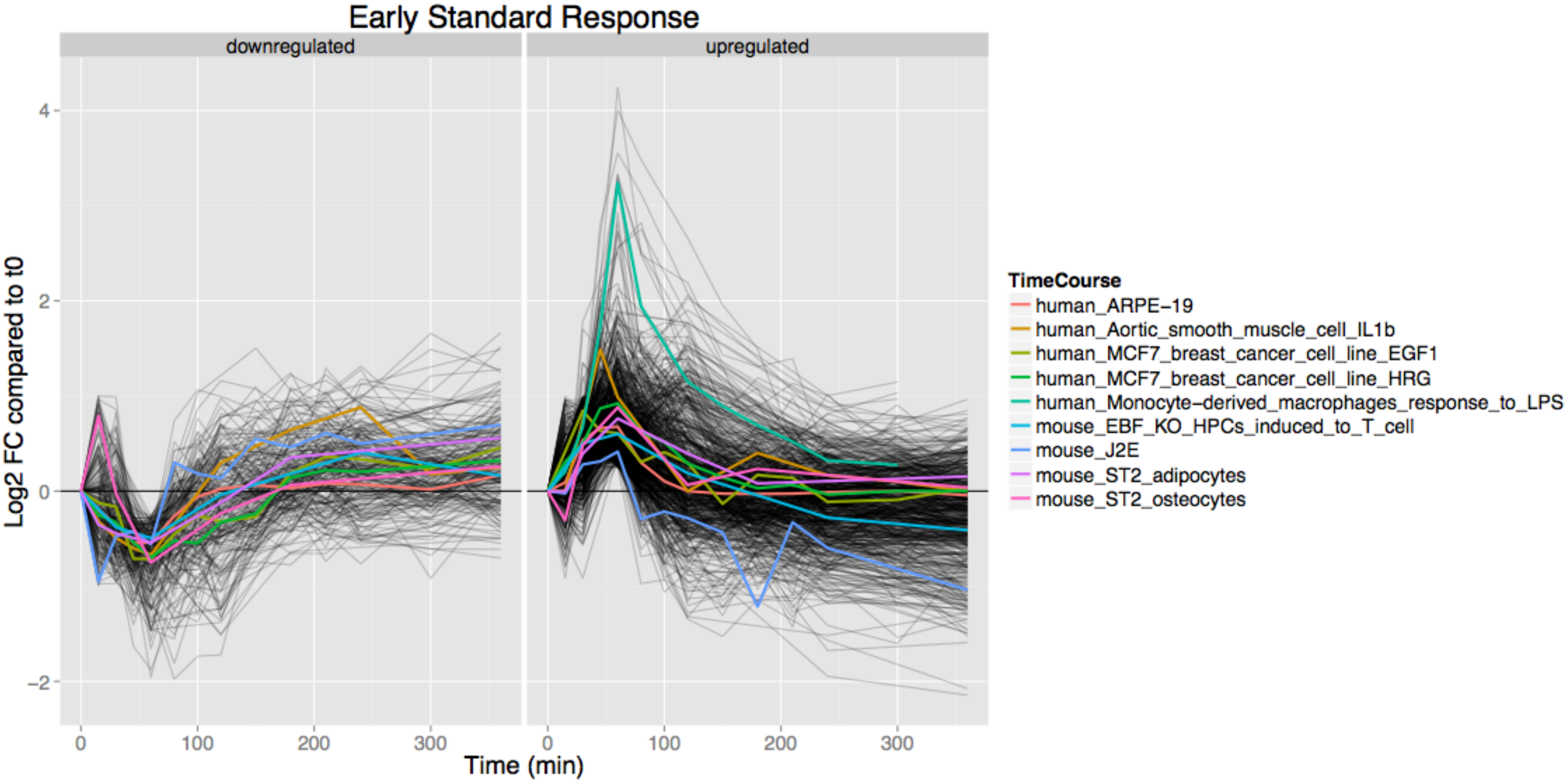

Rapid Short Response
downregulated
upregulated
Log2 FC compared to t0
4
2
0
−2
0
100
200
300
0
100
200
300
Time (min)
TimeCourse
human_ARPE−19
human_Aortic_smooth_muscle_cell_IL1b
human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_HRG
human_Monocyte−derived_macrophages_response_to_LPS
mouse_EBF_KO_HPCs_induced_to_T_cell
mouse_J2E
mouse_ST2_adipocytes
mouse_ST2_osteocytes


2) Rapid Long Response
Log2 FC vs 0 min
0
0
15
30
Time (min)

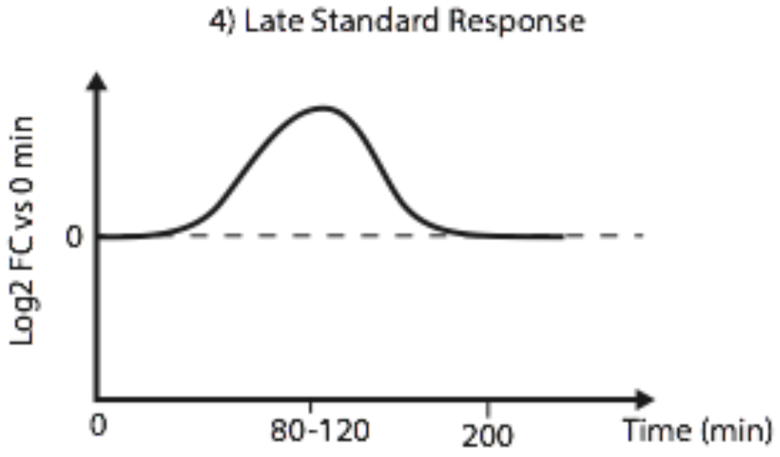

Rapid Long Response
downregulated
upregulated
Log2 FC compared to t0
5.0
2.5
0.0
−2.5
0
100
200
300
0
100
200
300
Time (min)
TimeCourse
human_ARPE−19
human_Aortic_smooth_muscle_cell_IL1b
human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_HRG
human_Monocyte−derived_macrophages_response_to_LPS
mouse_EBF_KO_HPCs_induced_to_T_cell
mouse_J2E
mouse_ST2_adipocytes
mouse_ST2_osteocytes

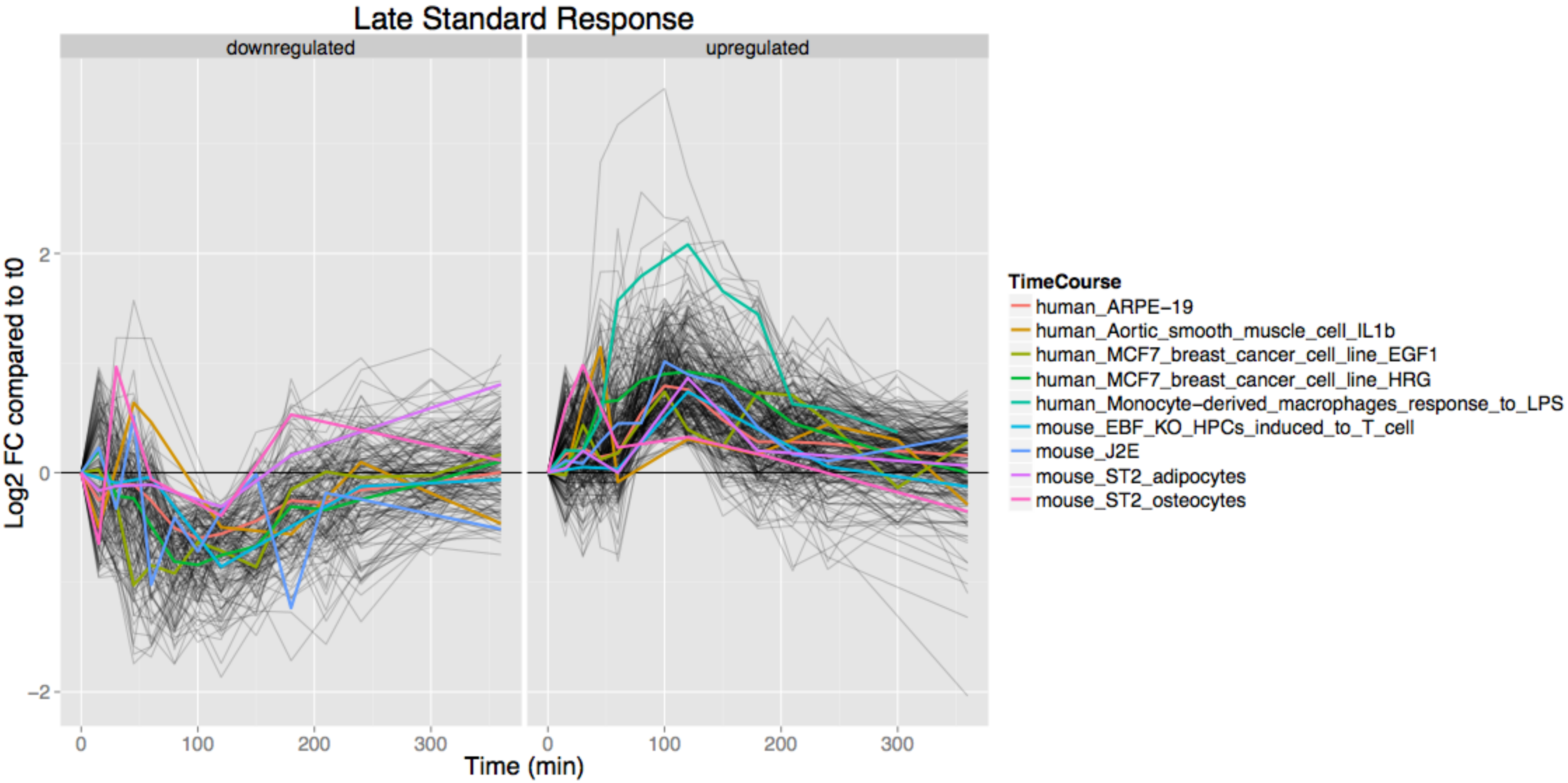


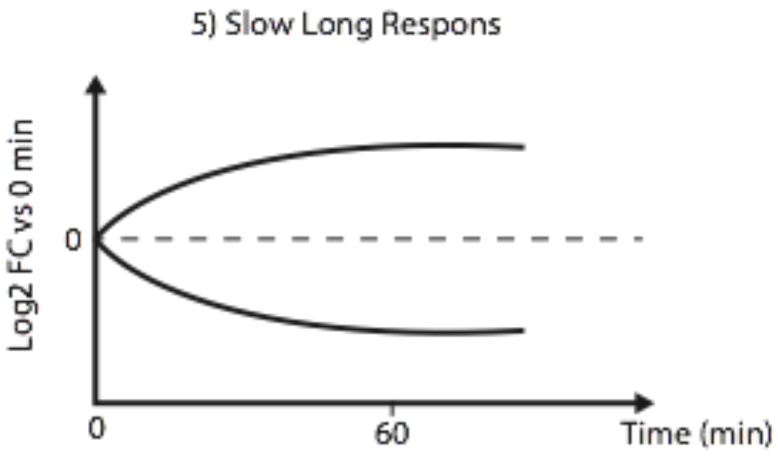
3) Early Standard Response

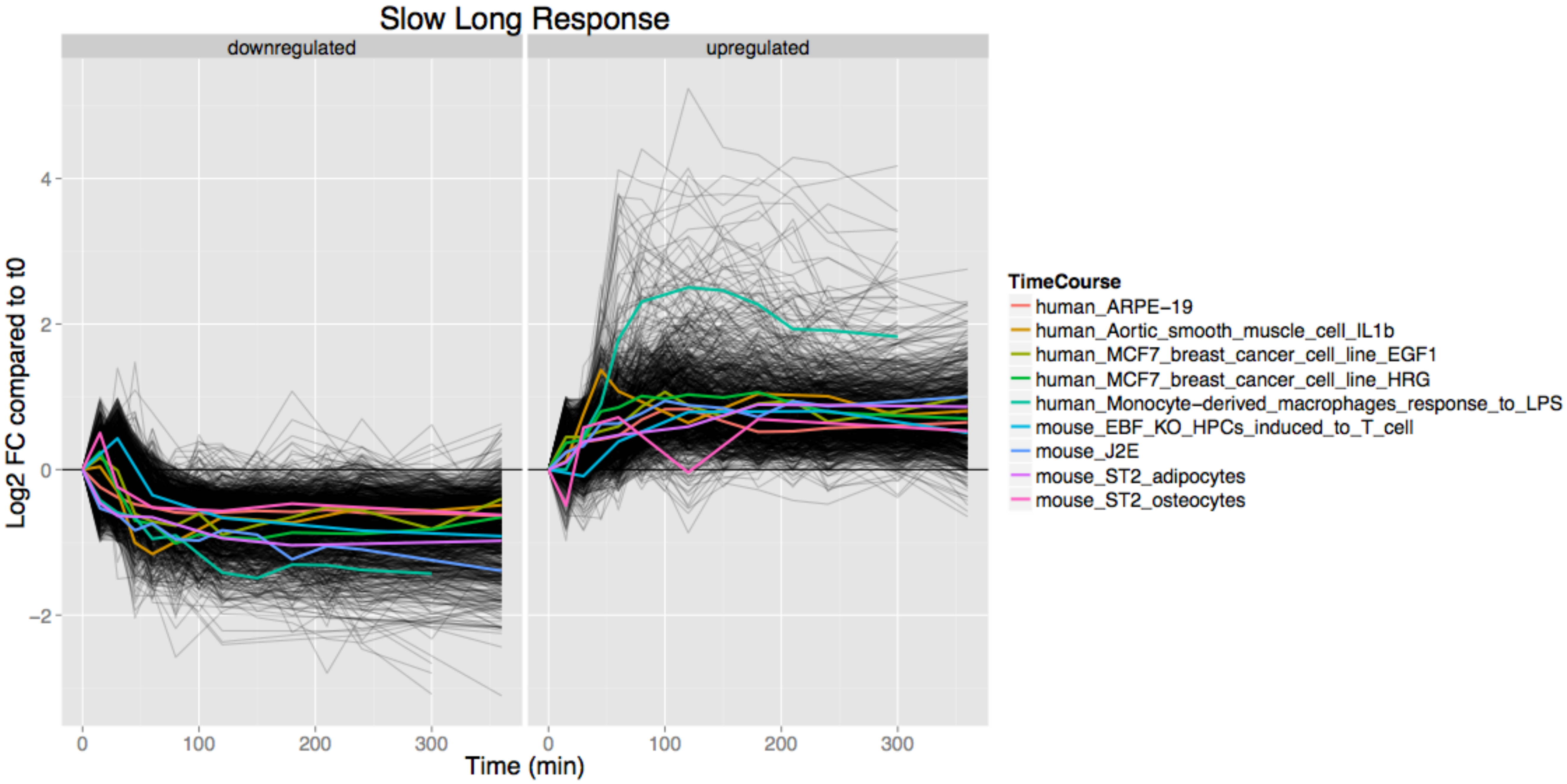


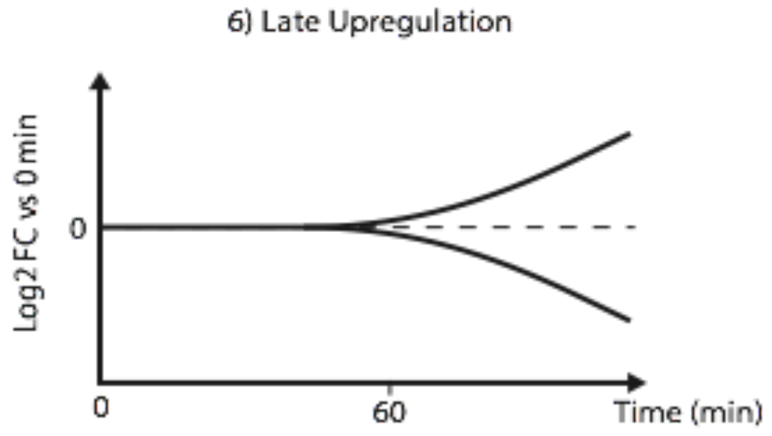
Early Standard Response

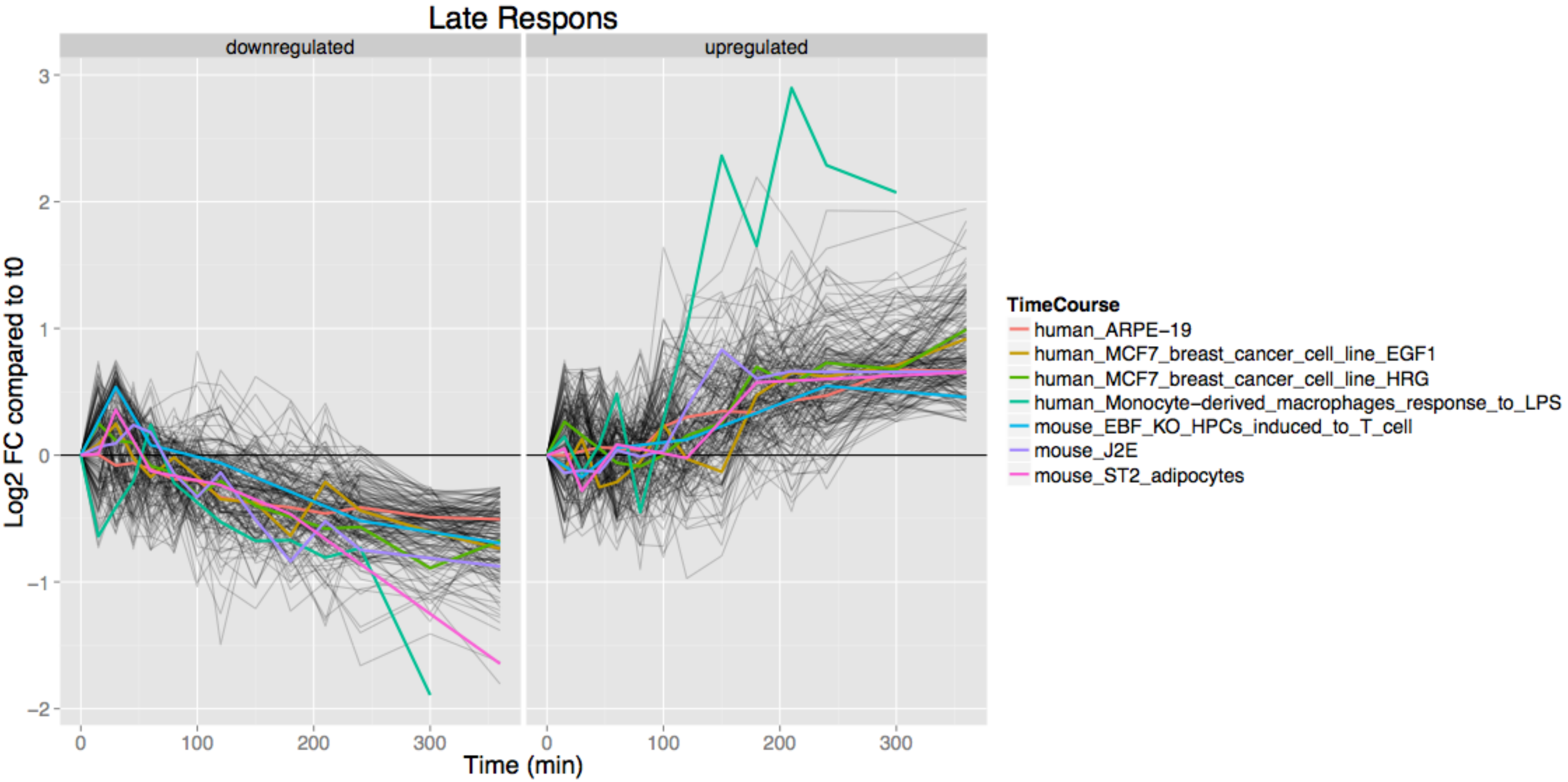

4) Late Standard Response
Log2 FC vs 0 min
0
0
80-120
200
Time (min)


Late Standard Response
downregulated
upregulated
Log2 FC compared to t0
2
0
−2
0
100
200
300
0
100
200
300
Time (min)
TimeCourse
human_ARPE−19
human_Aortic_smooth_muscle_cell_IL1b
human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_HRG
human_Monocyte−derived_macrophages_response_to_LPS
mouse_EBF_KO_HPCs_induced_to_T_cell
mouse_J2E
mouse_ST2_adipocytes
mouse_ST2_osteocytes

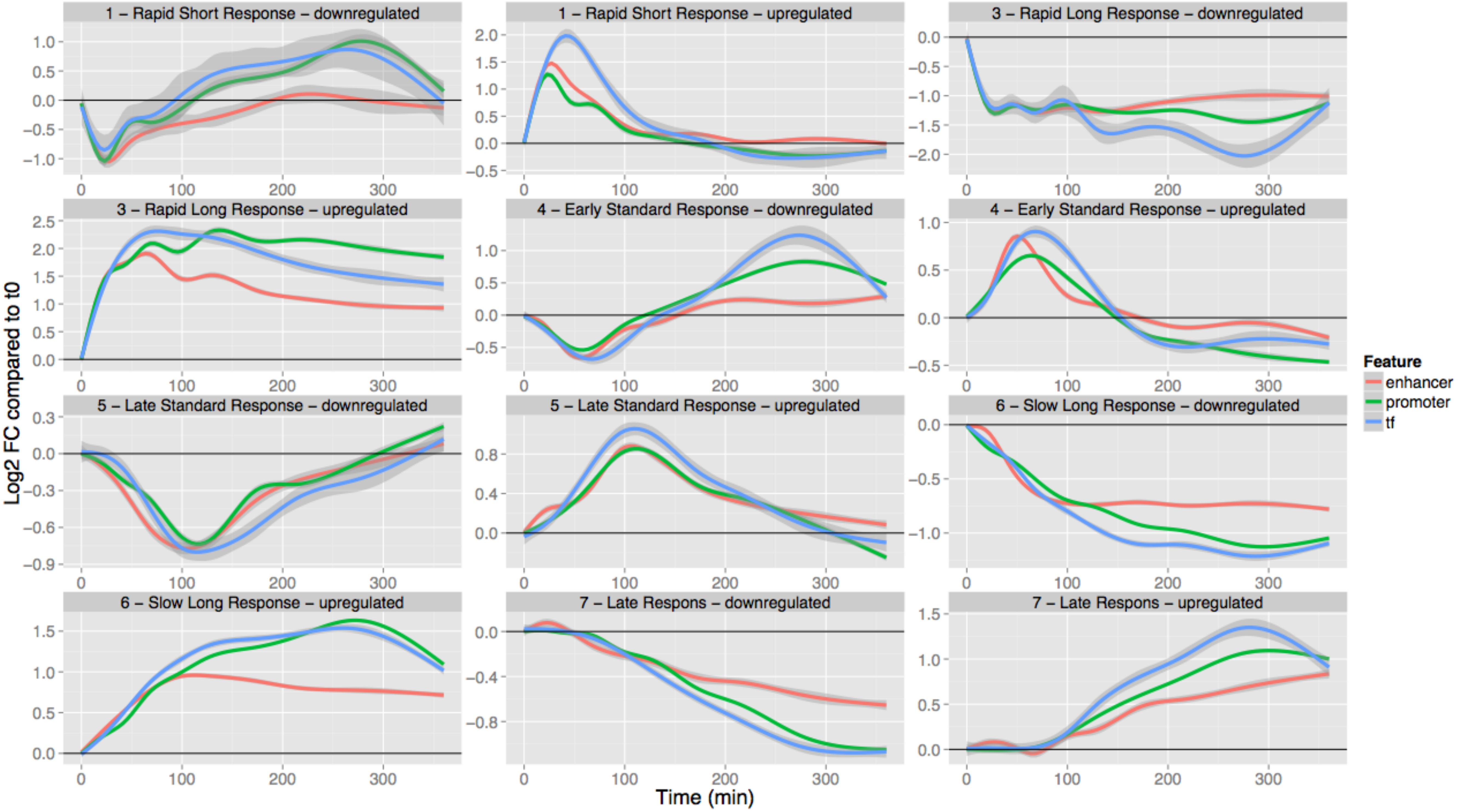

5) Slow Long Respons
Log2 FC vs 0 min
0
0
60
Time (min)


Slow Long Response
downregulated
upregulated
Log2 FC compared to t0
4
2
0
−2
0
100
200
300
0
100
200
300
Time (min)
TimeCourse
human_ARPE−19
human_Aortic_smooth_muscle_cell_IL1b
human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_HRG
human_Monocyte−derived_macrophages_response_to_LPS
mouse_EBF_KO_HPCs_induced_to_T_cell
mouse_J2E
mouse_ST2_adipocytes
mouse_ST2_osteocytes


6) Late Upregulation
Log2 FC vs 0 min
0
0
60
Time (min)


Late Respons
downregulated
upregulated
Log2 FC compared to t0
3
2
1
0
−1
−2
0
100
200
300
0
100
200
300
Time (min)
TimeCourse
human_ARPE−19
human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_HRG
human_Monocyte−derived_macrophages_response_to_LPS
mouse_EBF_KO_HPCs_induced_to_T_cell
mouse_J2E
mouse_ST2_adipocytes

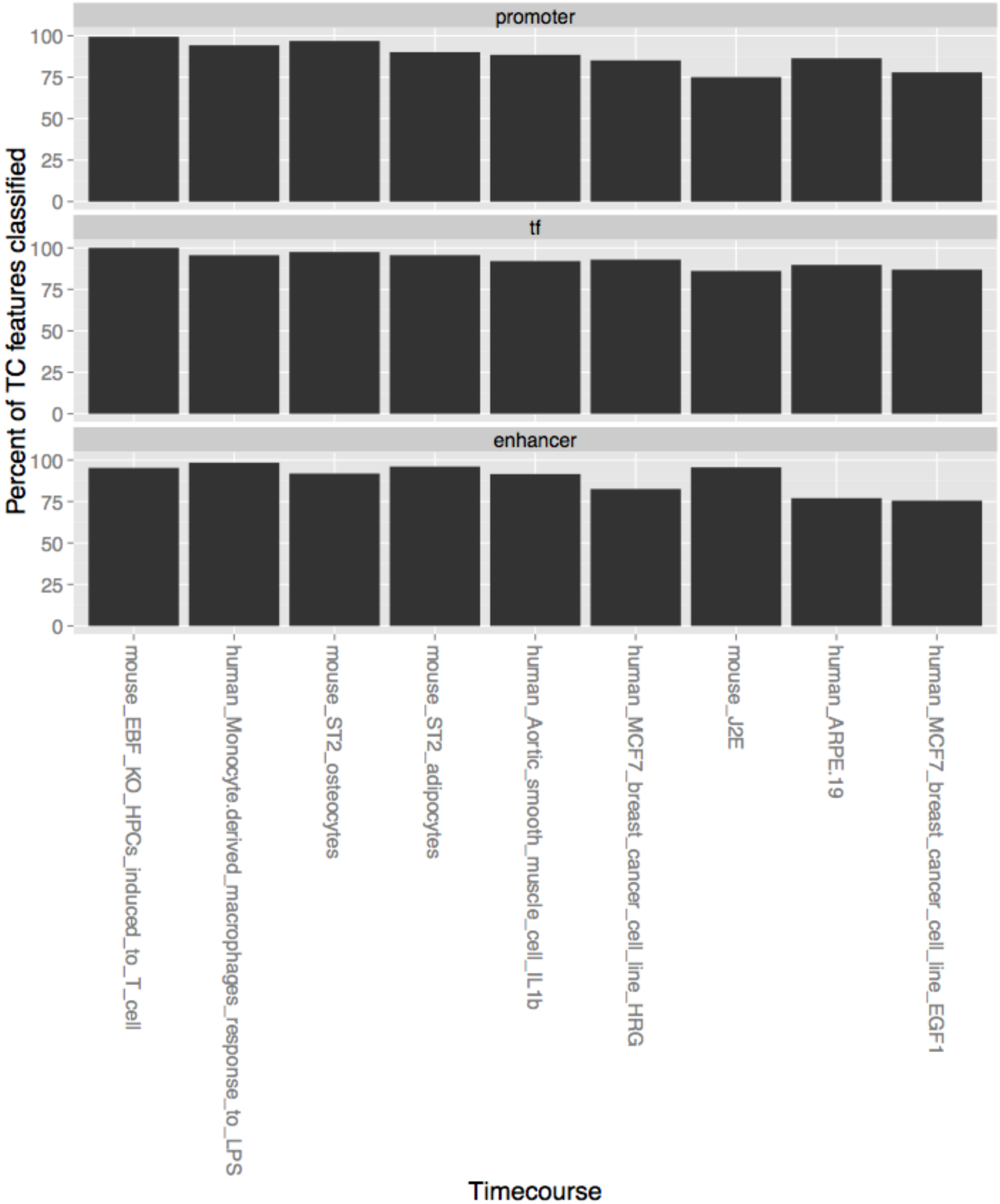
We find the same pattern for promoters and TFs



Sanity check

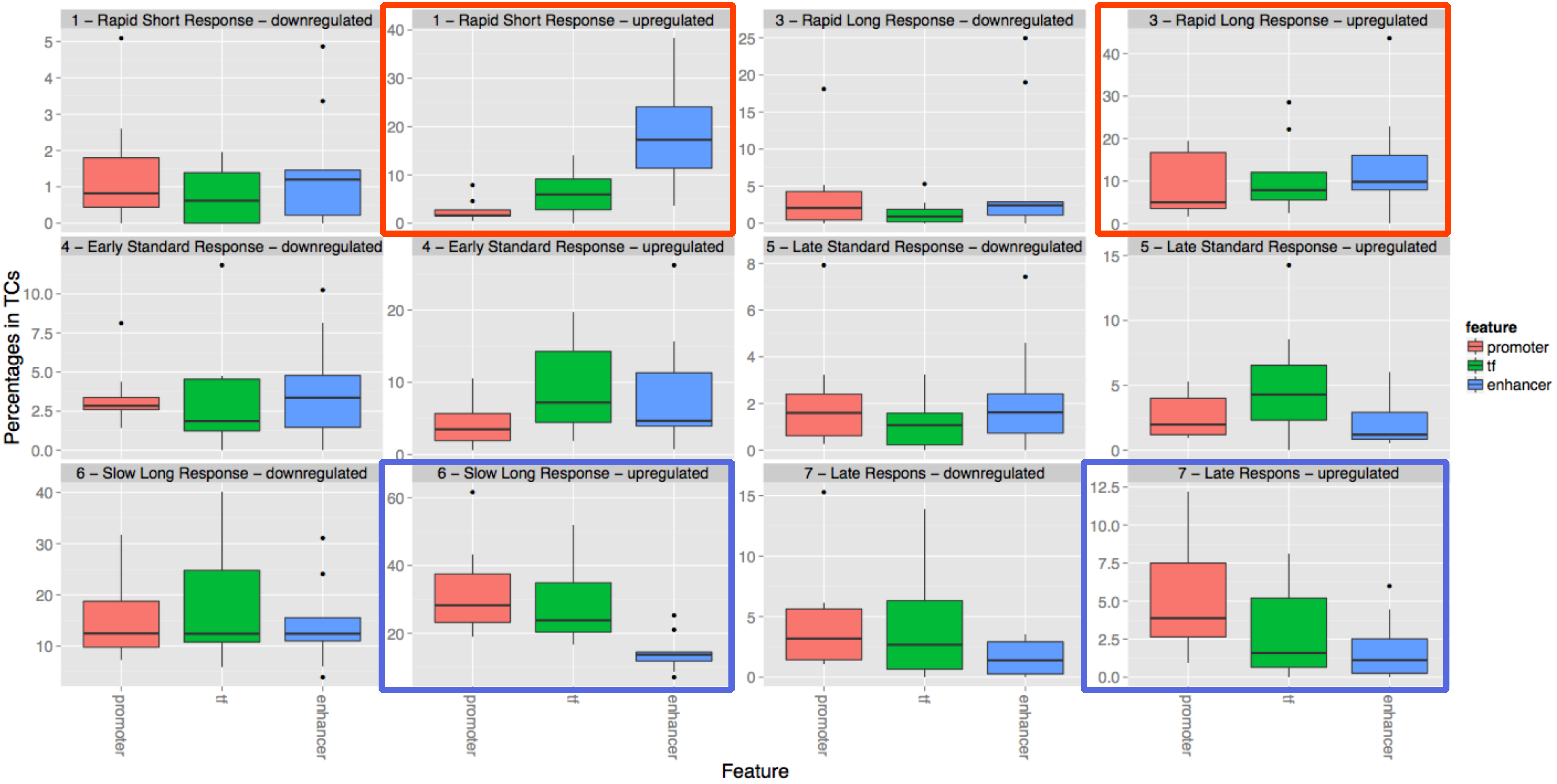
- For each time course and each feature make unsupervised clustering (k-means, k=16) on those dynamics NOT classified by my rules.
- 26 * 16 clusters => I will not show them here.
    - I am working on doing some meta-clustering on these to enable easy visualisation.

Percent classified



	Average percent classified
Promoters	88.18
TFs	92.95
Enhancers	89.33

Comparison percent of features classified into each class



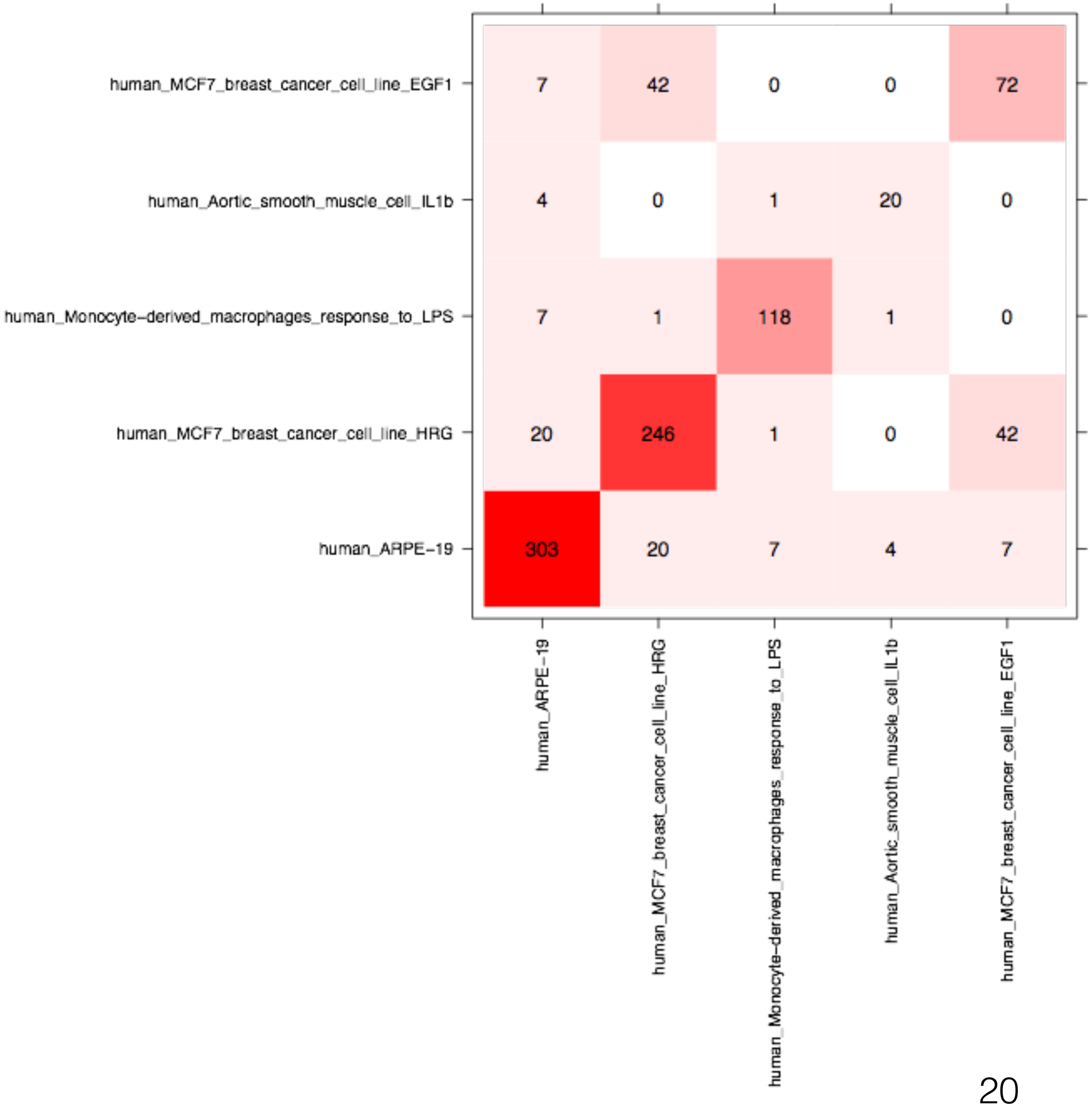
The combined data set

- For human and mouse all time-courses were combined.
- E.g. a feature might in this data-set be represented in multiple classes
- This dataset is used for the rest of the analysis presented here
- If nothing else is stated the analysis presented in the following is for the human time courses

Overlap between classes

Human enhancers: Rapid Short Response - Upregulated

Number of enhancers



	human_ARPE−19	human_MCF7_breast_cancer_cell_line_HRG	human_Monocyte−derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b	human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_EGF1	7	42	0	0	72
human_Aortic_smooth_muscle_cell_IL1b	4	0	1	20	0
human_Monocyte−derived_macrophages_response_to_LPS	7	1	118	1	0
human_MCF7_breast_cancer_cell_line_HRG	20	246	1	0	42
human_ARPE−19	303	20	7	4	7

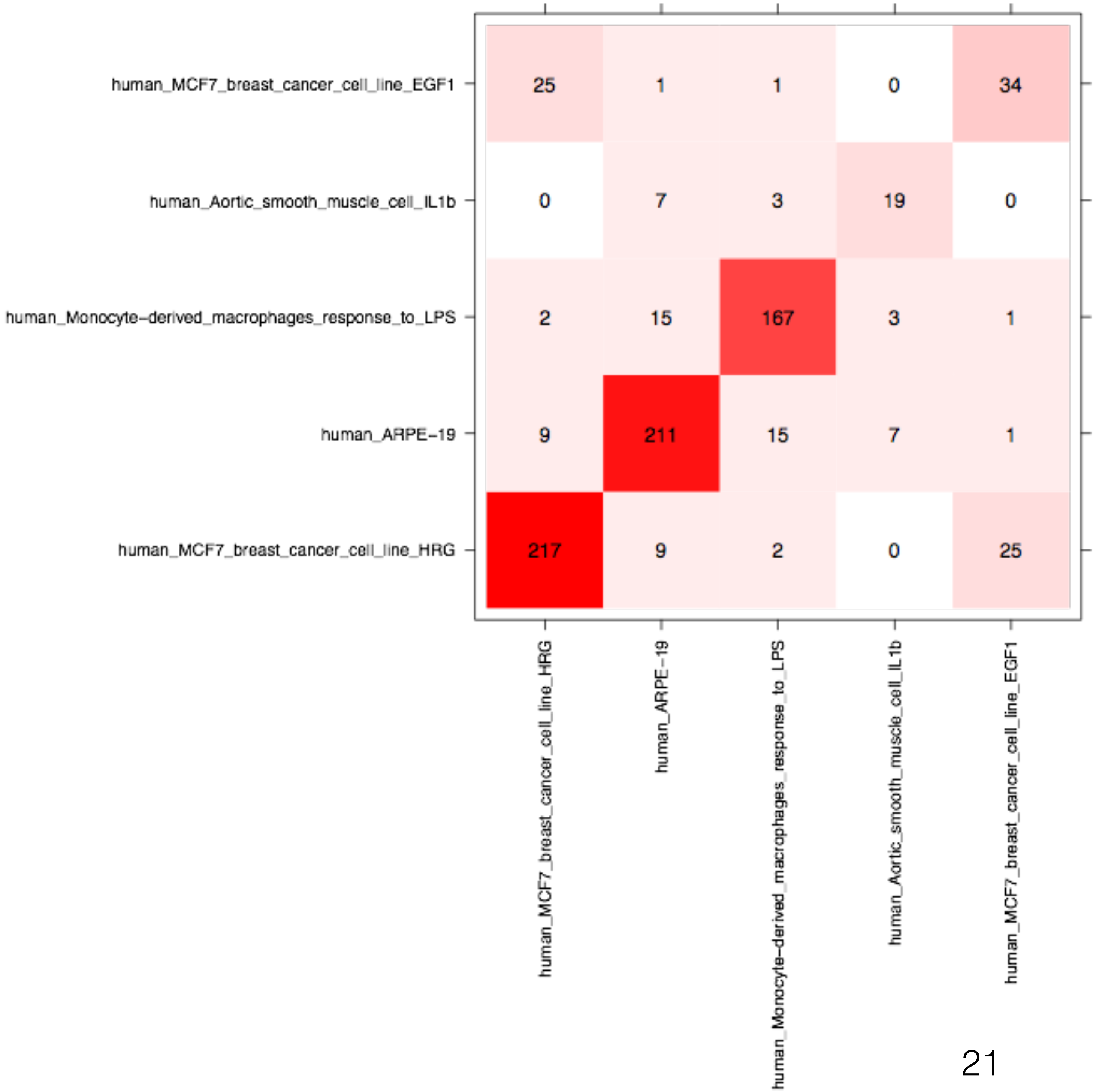
Percent of enhancers

	human_ARPE−19	human_MCF7_breast_cancer_cell_line_HRG	human_Monocyte−derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b	human_MCF7_breast_cancer_cell_line_EGF1
	1.9	15.22	0	0	100
	1.25	0	0.73	100	0
	1.69	0.28	100	0.73	0
	3.78	100	0.28	0	15.22
	100	3.78	1.69	1.25	1.9

Overlap between classes

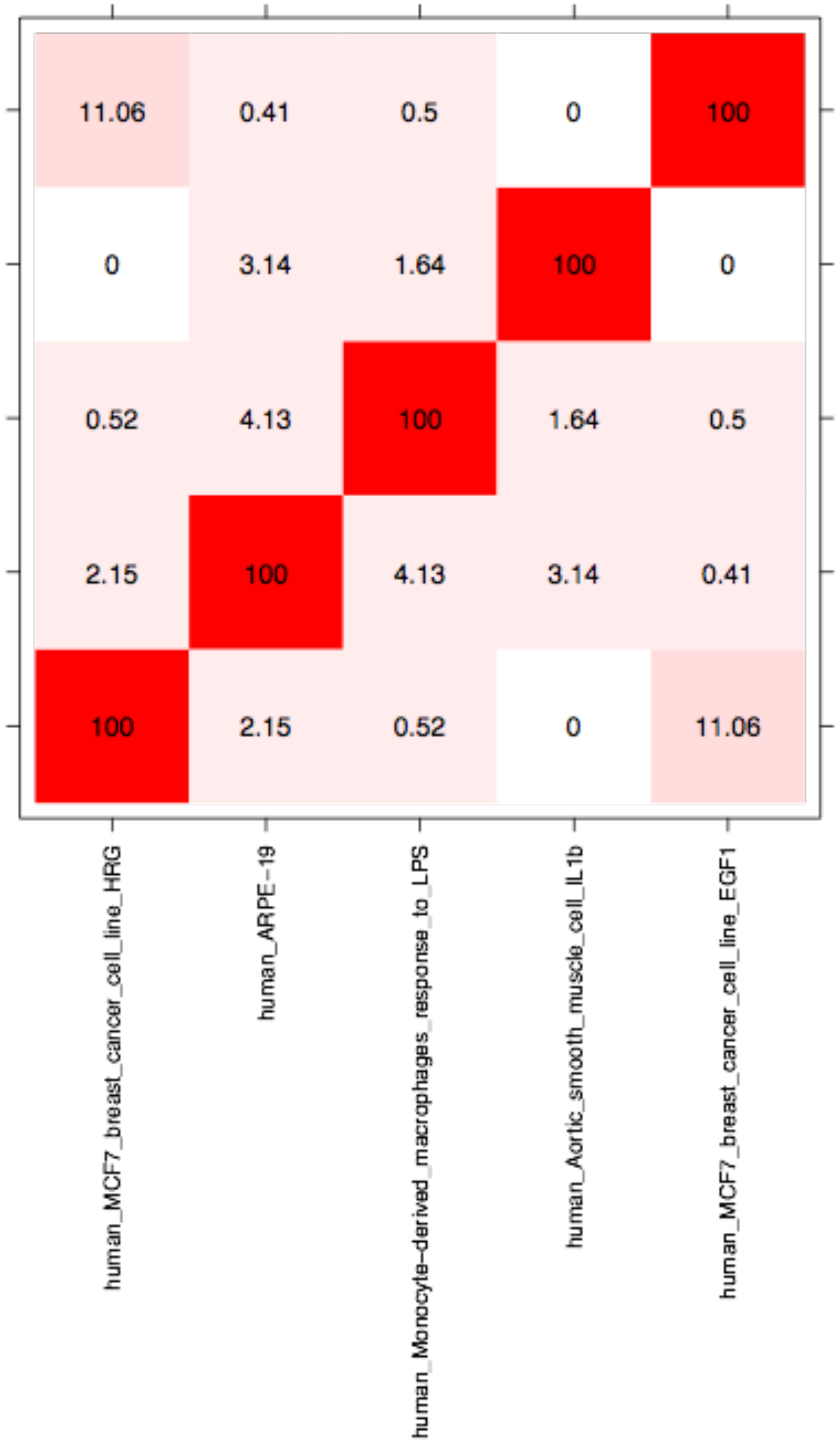
Human enhancers: Rapid Long Response - Upregulated

Number of enhancers



	human_MCF7_breast_cancer_cell_line_HRG	human_ARPE-19	human_Monocyte-derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b	human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_EGF1	25	1	1	0	34
human_Aortic_smooth_muscle_cell_IL1b	0	7	3	19	0
human_Monocyte-derived_macrophages_response_to_LPS	2	15	167	3	1
human_ARPE-19	9	211	15	7	1
human_MCF7_breast_cancer_cell_line_HRG	217	9	2	0	25

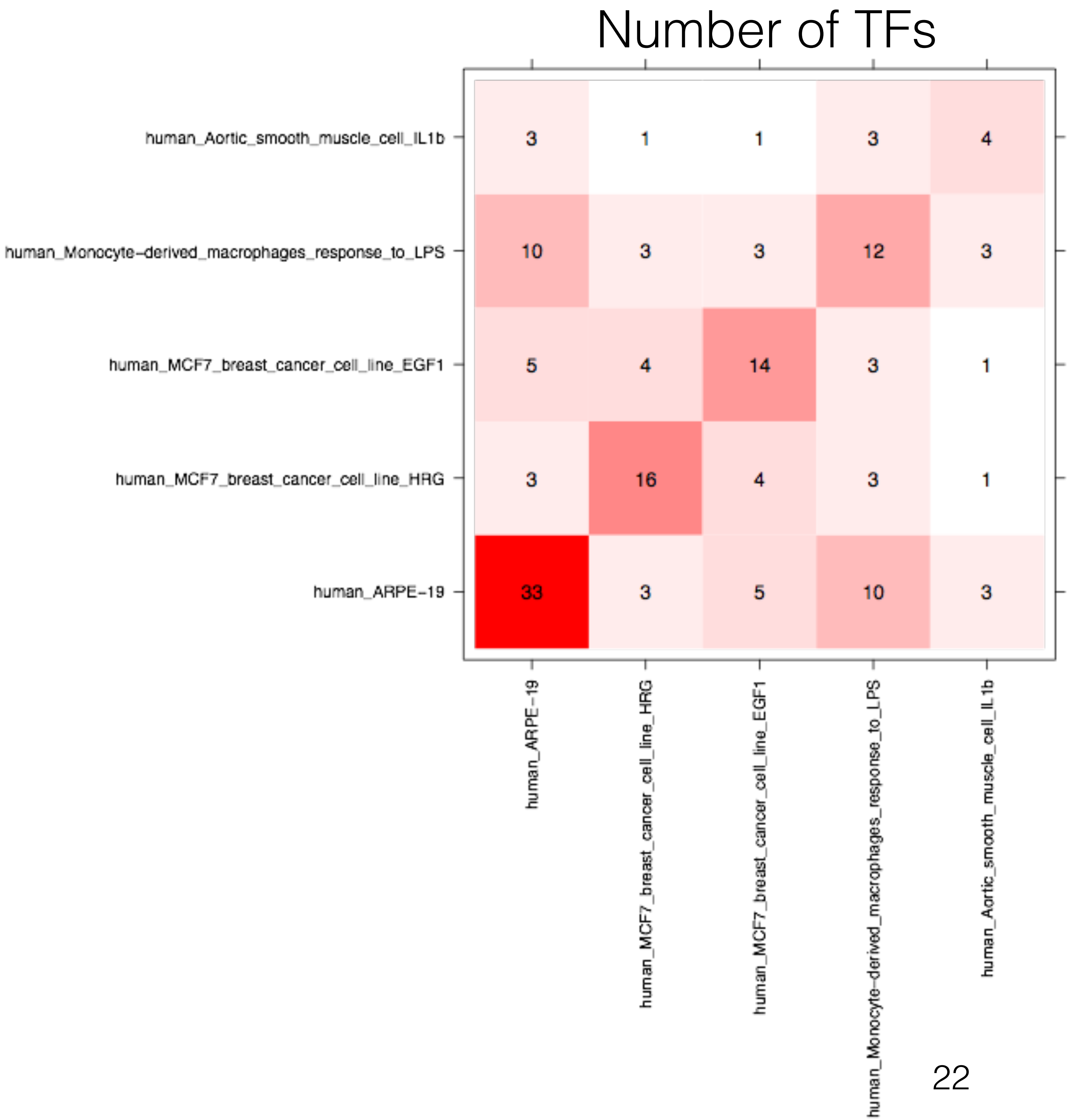
Percent of enhancers



	human_MCF7_breast_cancer_cell_line_HRG	human_ARPE-19	human_Monocyte-derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b	human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_EGF1	11.06	0.41	0.5	0	100
human_Aortic_smooth_muscle_cell_IL1b	0	3.14	1.64	100	0
human_Monocyte-derived_macrophages_response_to_LPS	0.52	4.13	100	1.64	0.5
human_ARPE-19	2.15	100	4.13	3.14	0.41
human_MCF7_breast_cancer_cell_line_HRG	100	2.15	0.52	0	11.06

Overlap between classes

Human TFs: Rapid Short Response - Upregulated



Number of TFs

	human_ARPE−19	human_MCF7_breast_cancer_cell_line_HRG	human_MCF7_breast_cancer_cell_line_EGF1	human_Monocyte−derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b
human_Aortic_smooth_muscle_cell_IL1b	3	1	1	3	4
human_Monocyte−derived_macrophages_response_to_LPS	10	3	3	12	3
human_MCF7_breast_cancer_cell_line_EGF1	5	4	14	3	1
human_MCF7_breast_cancer_cell_line_HRG	3	16	4	3	1
human_ARPE−19	33	3	5	10	3

Percent of TFs

	human_ARPE−19	human_MCF7_breast_cancer_cell_line_HRG	human_MCF7_breast_cancer_cell_line_EGF1	human_Monocyte−derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b
human_Aortic_smooth_muscle_cell_IL1b	8.82	5.26	5.88	23.08	100
human_Monocyte−derived_macrophages_response_to_LPS	28.57	12	13.04	100	23.08
human_MCF7_breast_cancer_cell_line_EGF1	11.9	15.38	100	13.04	5.88
human_MCF7_breast_cancer_cell_line_HRG	6.52	100	15.38	12	5.26
human_ARPE−19	100	6.52	11.9	28.57	8.82

Overlap between classes

Human TFs: Rapid Long Response - Upregulated

Number of TFs

	human_MCF7_breast_cancer_cell_line_HRG	human_ARPE-19	human_Monocyte-derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b	human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_EGF1	31	6	1	1	34
human_Aortic_smooth_muscle_cell_IL1b	2	2	1	2	1
human_Monocyte-derived_macrophages_response_to_LPS	8	21	52	1	1
human_ARPE-19	16	81	21	2	6
human_MCF7_breast_cancer_cell_line_HRG	77	16	8	2	31

Percent of TFs

	human_MCF7_breast_cancer_cell_line_HRG	human_ARPE-19	human_Monocyte-derived_macrophages_response_to_LPS	human_Aortic_smooth_muscle_cell_IL1b	human_MCF7_breast_cancer_cell_line_EGF1
human_MCF7_breast_cancer_cell_line_EGF1	38.75	5.5	1.18	2.86	100
human_Aortic_smooth_muscle_cell_IL1b	2.6	2.47	1.89	100	2.86
human_Monocyte-derived_macrophages_response_to_LPS	6.61	18.75	100	1.89	1.18
human_ARPE-19	11.27	100	18.75	2.47	5.5
human_MCF7_breast_cancer_cell_line_HRG	100	11.27	6.61	2.6	38.75

Overlap between classes

Human: Number of TCs enhancers are found in (out of 5)

	1	2	3	4
Rapid short response - upregulated	616	61	7	-
Rapid long response - upregulated	536	51	2	1
Rapid long response - downregulated	70	5	-	-

Within 20 kb:
KDM6B
LSMD1

Human: Number of TCs TFs are found in (out of 5)

	1	2	3	4	5
Rapid short response - upregulated	42	8	3	1	1
Rapid long response - upregulated	117	44	11	1	-
Rapid long response - downregulated	30	1	-	-	-

p1@FOS

p2@FOS

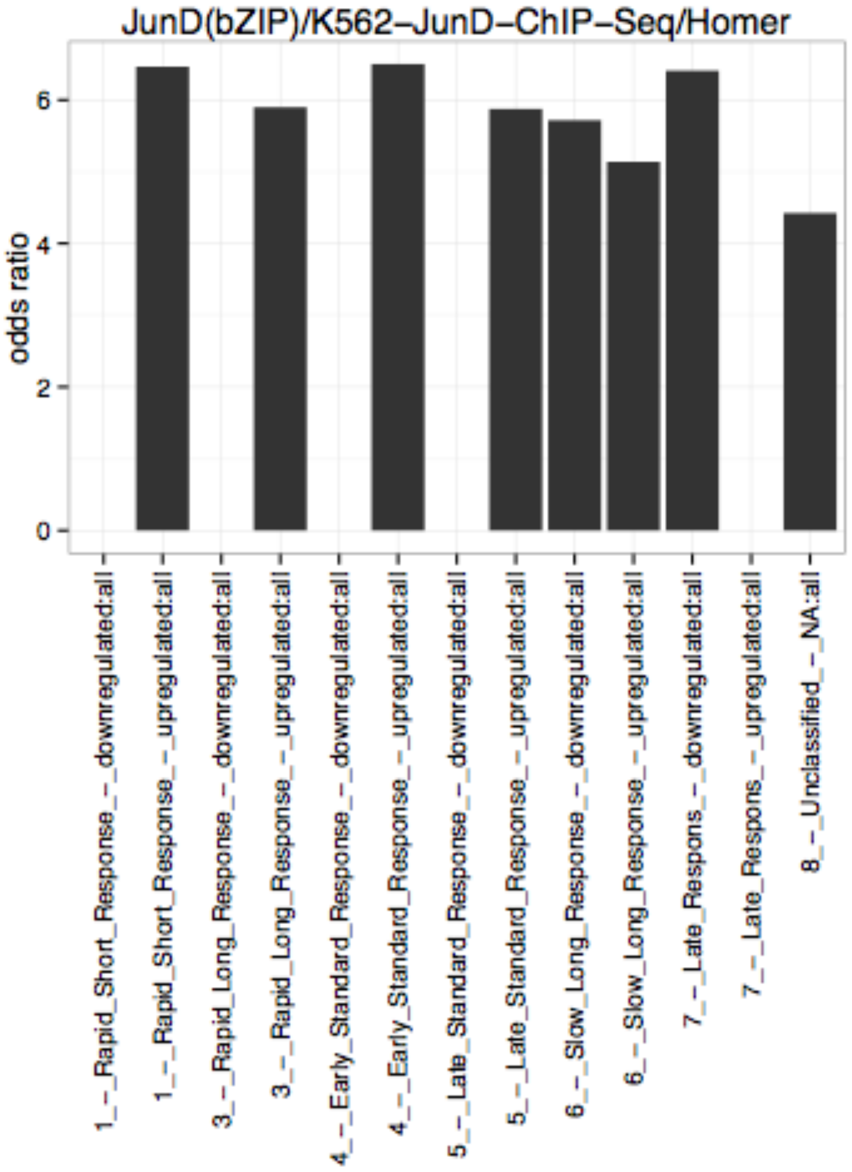
p1@FOSB

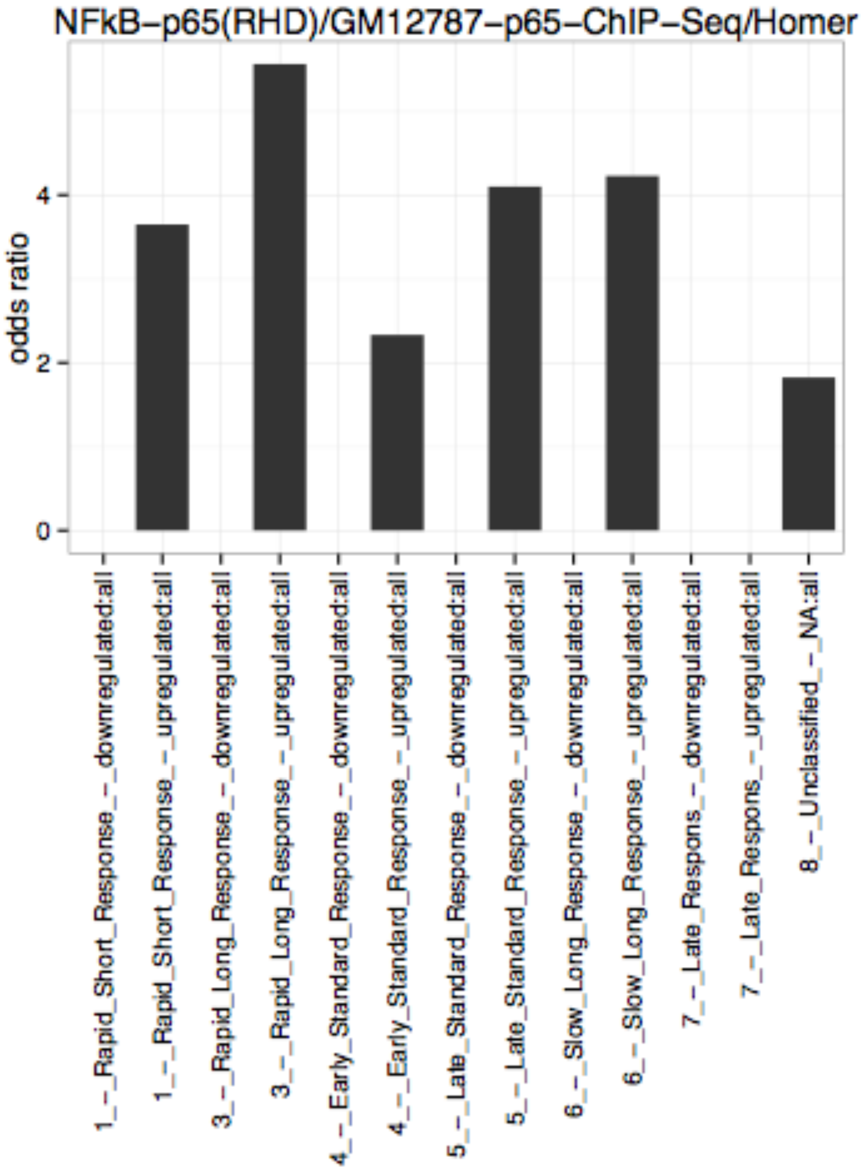
Motif analysis

- The motif analysis was done with Homer2 on known motifs
- The motif analysis was done for the combined dataset - meaning the motifs found are not time course specific.
- Only significant results are shown (no bar = motif not significant)

Motif analysis

All downregulated classes


JunD(bZIP)/K562−JunD−ChIP−Seq/Homer

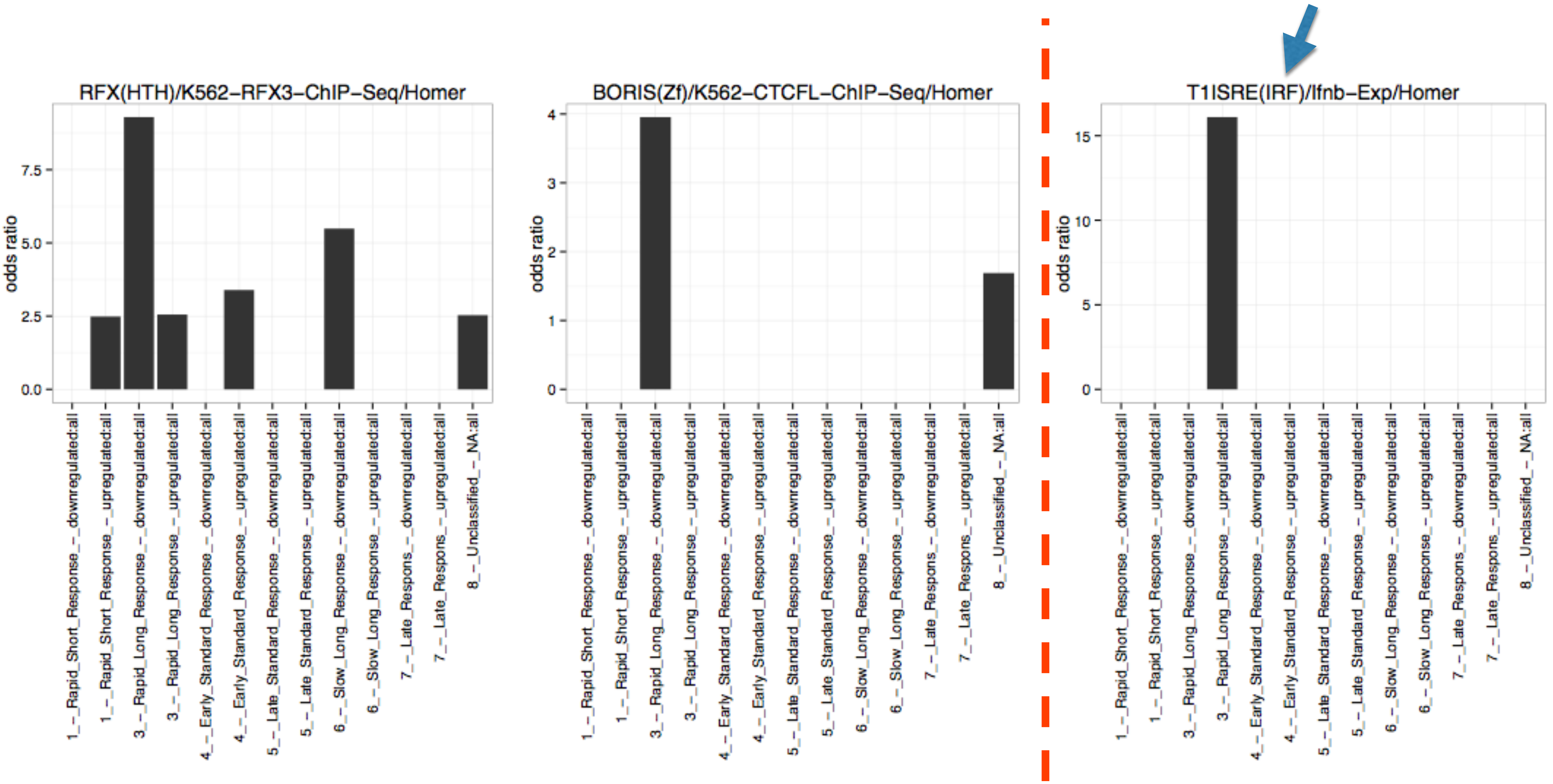

NFkB−p65(RHD)/GM12787−p65−ChIP−Seq/Homer

The downregulated classes are not enriched for JunD and NFkB
(But the upregulated classes are)

Motif analysis

Rapid long response

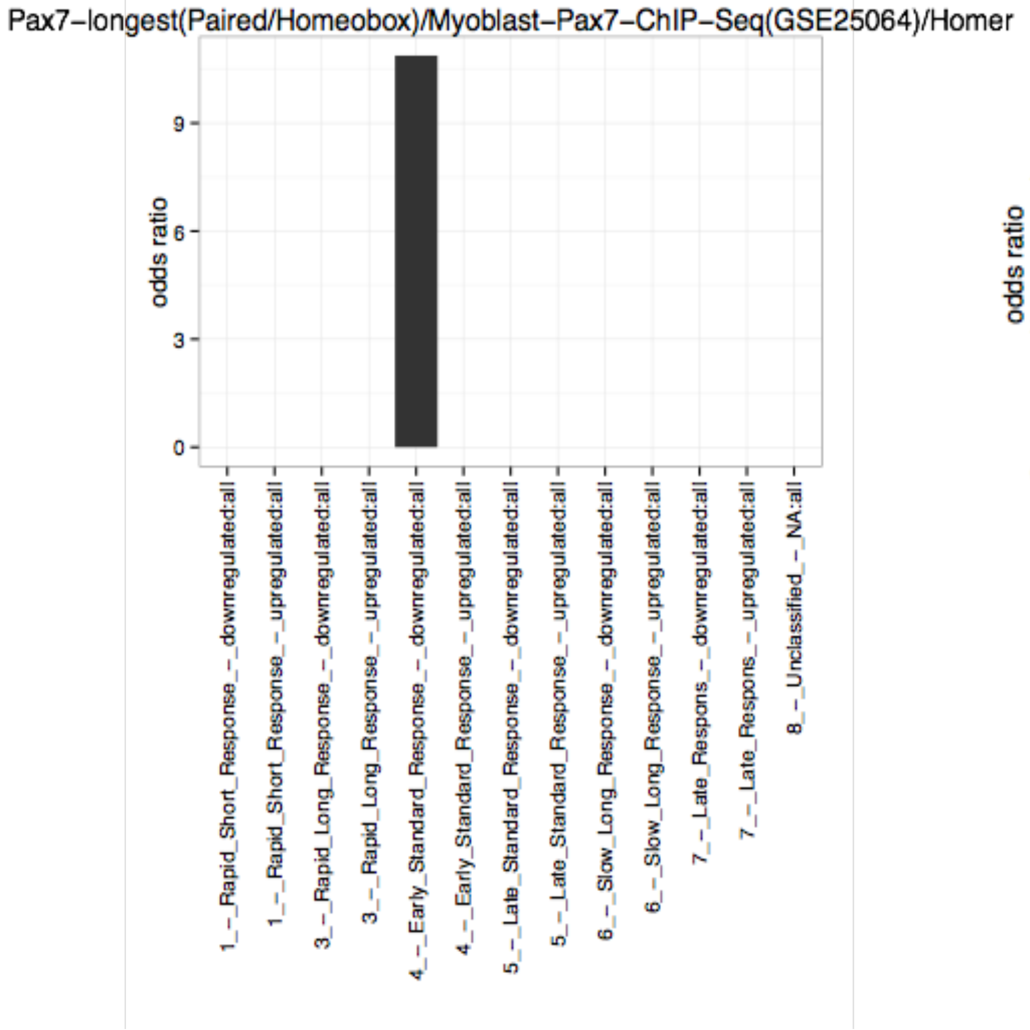
Interferon regulatory factor family

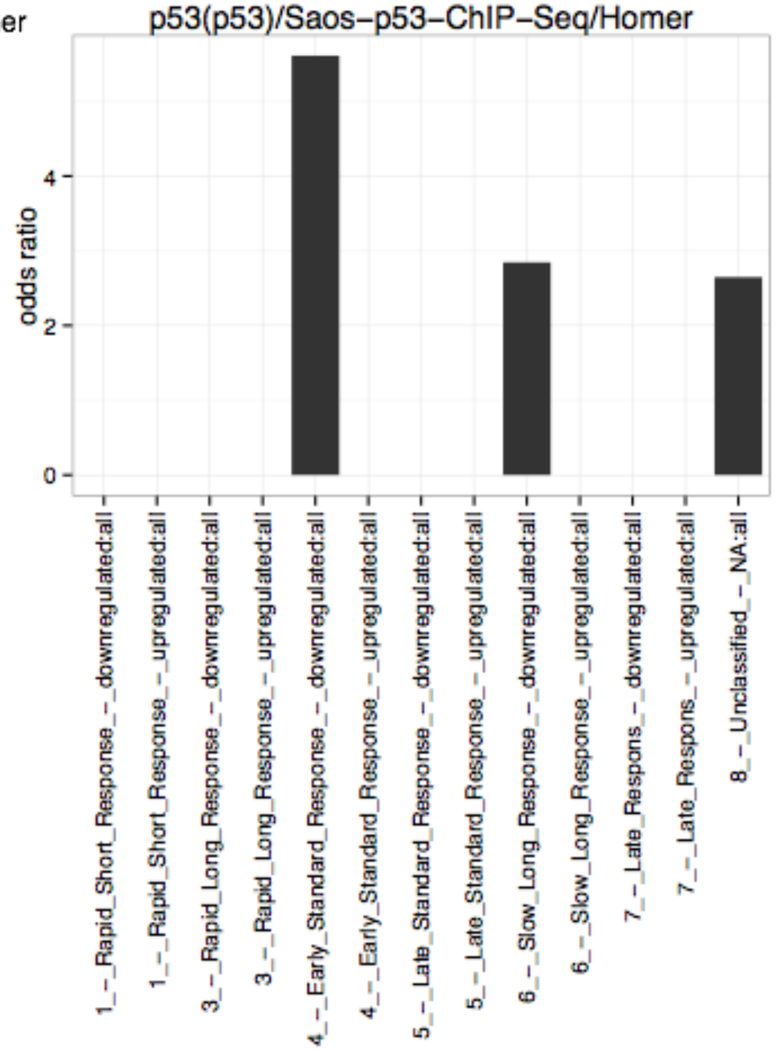


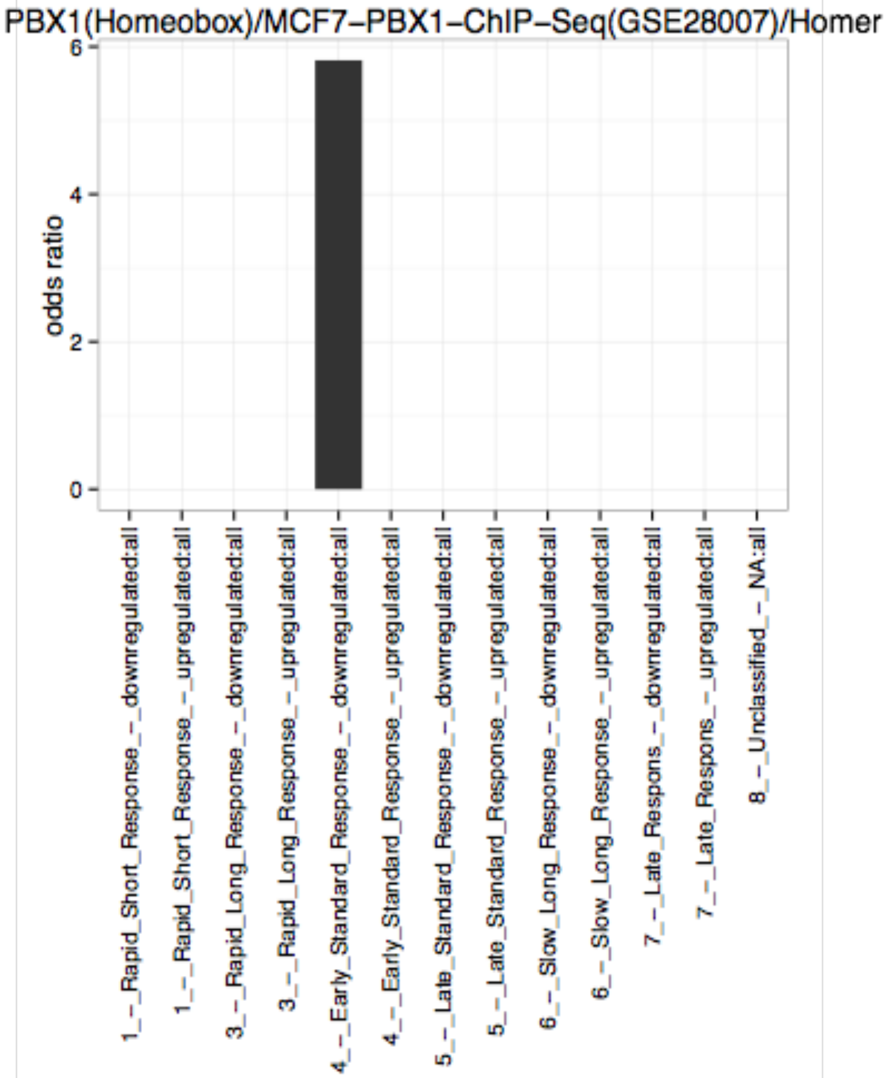
The upregulated rapid long response is enriched for RFX and BORIS while the downregulated is enriched for T1ISRE

Motif analysis

Downregulated Early Standard Response



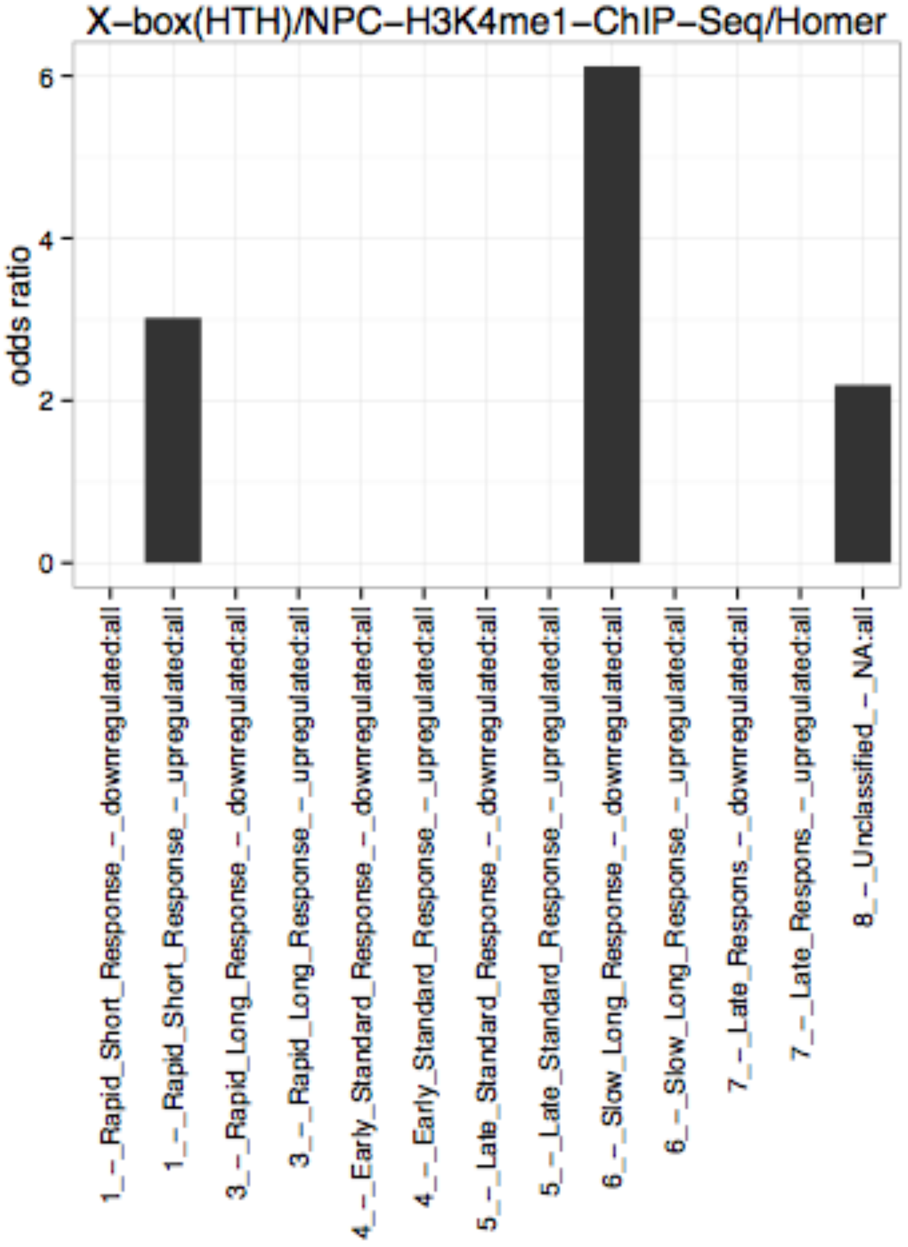




The Downregulated Early Standard Response is enriched for pax7, p53 and PBX1

Motif analysis

X-box



The X-box motif is enriched in the rapid short upregulated and afterwards in the log slow downregulated - first activation then inactivation?

GO-terms of promoters in the vicinity of enhancers in classes

- For each class the promoters within 20 kb of enhancers were extracted

- Only GO-terms from level 5 of the GO-tree (where level 2 consists of “Cellular component”, “Biological process” and “Molecular function”) were used

- Only“Biological Process” GO-terms were used

GO-term analysis

Number of classes containing the a GO-terms	1	2	3	4	5	6	7	8	9	10
Number of GO-terms	260	124	61	33	29	23	15	15	11	16

Class specific GO-terms

Common GO-terms

Note: There is a total of 13 classes

GO-term analysis

Common GO-terms

Number of times go-terms found	1	2	3	4	5	6	7	8	9	10
Number of GO-terms	260	124	61	33	29	23	15	15	11	16

- GO_RNA_biosynthetic_process
- GO_RNA_metabolic_process
- GO_apoptotic_process
- GO_cell_surface_receptor_signaling_pathway
- GO_cellular_protein_modification_process
- GO_negative_regulation_of_macromolecule_metabolic_process
- GO_positive_regulation_of_macromolecule_biosynthetic_process
- GO_positive_regulation_of_nitrogen_compound_metabolic_process
- GO_regulation_of_RNA_metabolic_process
- GO_regulation_of_cellular_biosynthetic_process
- GO_regulation_of_cellular_macromolecule_biosynthetic_process
- GO_regulation_of_cellular_protein_metabolic_process
- GO_regulation_of_macromolecule_biosynthetic_process
- GO_regulation_of_nucleobase
- GO_regulation_of_protein_metabolic_process
- GO_transcription

GO-term analysis

Class specific GO-terms - top 10 (after q-value)

Number of classes containing the a GO-terms	1	2	3	4	5	6	7	8	9	10
Number of GO-terms	260	124	61	33	29	23	15	15	11	16

1 - Rapid Short Response - upregulated
GO skeletal muscle organ development
GO lysosomal transport
GO synaptic transmission
GO vacuolar transport
GO stem cell maintenance
GO transmission of nerve impulse
GO artery morphogenesis
GO positive regulation of autophagy
GO Notch signalling pathway
GO fusion of virus membrane with host plasma membrane

GO-term analysis

Class specific GO-terms - top 10 (after q-value)

Number of classes containing the a GO-terms	1	2	3	4	5	6	7	8	9	10
Number of GO-terms	260	124	61	33	29	23	15	15	11	16

3 - Rapid Long Response - upregulated
GO positive regulation of hydrolase activity
GO regulation of nitric oxide biosynthetic process
GO natural killer cell chemotaxis
GO modification-dependent macromolecule catabolic process
GO nucleoside phosphate biosynthetic process
GO regulation of granulocyte chemotaxis
GO regulation of organelle assembly
GO proteolysis involved in cellular protein catabolic process
GO proteasomal protein catabolic process
GO positive regulation of nitric oxide biosynthetic process

GO-term analysis

Class specific GO-terms - top 10 (after q-value)

Number of classes containing the a GO-terms	1	2	3	4	5	6	7	8	9	10
Number of GO-terms	260	124	61	33	29	23	15	15	11	16

3 - Rapid Long Response - downregulated
GO branching involved in prostate gland morphogenesis
GO prostate gland morphogenesis
GO prostate gland development
GO cellular response to steroid hormone stimulus
GO xenobiotic metabolic process
GO prostate glandular acinus development
GO negative regulation of cell cycle process
GO regulation of cell cycle arrest
GO cellular response to ketone

Future Plans

- Make non-hierarchical classifier - meaning each response dynamics can be classified into several classes
- Meta-clustering of those dynamics that are not classified by my algorithm (sanity check)
- Motif analysis on individual TCs
- Overlap with the IER promoters/enhancers that Erik and Robin have identified
- Look into how promoters close to each class of enhancer responds
- Analyse the relative distance between enhancer and promoter groups using GenometriCorr (Favorov et al., PLoS Comput Biol. 2012)

Acknowledgement

- Robin Andersson
- Albin Sandelin