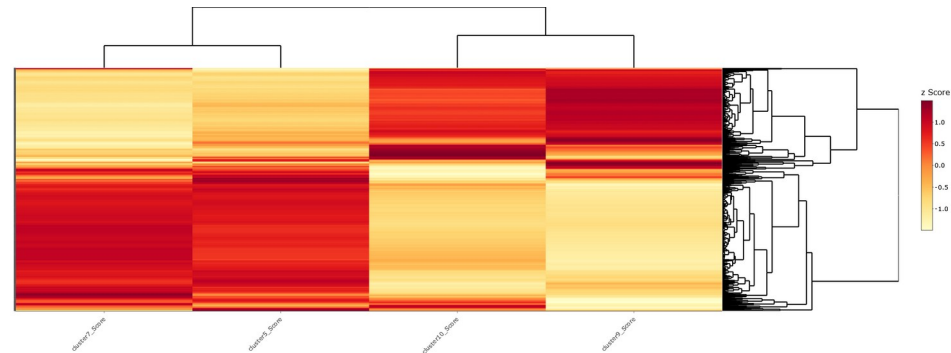# WP3 – 21.01.2022

Tasks done since last week

- Snakemake pipeline mostly running
  - Limited by RAM
  - Testing with reduced comparisons for BINDetect
  - Quite slow (more cores might speed it up)
- Simple script to extract the defining TFs for each cluster

cluster5_Score	cluster7_Score	cluster9_Score	cluster10_Score
ATOH1(var.2)	ATF7	Ar	CLOCK
BARHL2	BACH2(var.2)	CENPB	CREM
EBF1	BHLHE22	DPRX	EOMES
Ebf2	CEBPB	Dux	HOXD12
EN2	CEBPE	GATA1::TAL1	IRF5
FOXA1	CEBPG	GATA6	MGA
FOXC1	CEBPG(var.2)	Gmeb1	MYF5
FOXC2	Creb5	Hic1	MYF6
FOXD2	DBP	HMBOX1	NKX2-3
FOXE1	DMRT3	LEF1	NKX2-5
Foxo1	FOS::JUN(var.2)	NFKB1	Pax2
GATA1	FOSB::JUN	NFKB2	PAX3(var.2)
HAND2	FOSB::JUNB(var.2)	NR1I2	PAX9
HNF4A(var.2)	FOSL1::JUN(var.2)	NR2C2	PKNOX2
HOXA13	FOSL1::JUND(var.2)	ONECUT2	Rhox11
HOXC13	FOSL2::JUN(var.2)	OSR1	RUNX1
ISL2	FOSL2::JUNB(var.2)	PBX1	RUNX3
NEUROD1	FOSL2::JUND(var.2)	PROX1	SMAD2::SMAD3::SMAD4
NEUROG2(var.2)	HOXA9	RFX1	SREBF1
Nkx3-1	HOXC10	RFX4	SREBF2
NR1I3	HOXC12	RXRA::VDR	TBR1
Nr2e3	HOXC8	SNAI1	TBX1
ONECUT3	HOXC9	SNAI3	TBX15
OSR2	HOXD10	SOX10	TBX18
Ptf1a	HOXD11	SOX12	TBX2
RARA::RXRA	JDP2(var.2)	Spz1	TBX20
RARB(var.3)	JUN::JUNB(var.2)	TCF4	TBX21
RELA	MZF1	TCF7	TBX3
RXRG	NFATC1	TCF7L1	TBX4
SOX14	NFATC2	TCF7L2	TBX5
SRF	NFATC3	THAP11	TBX6
TAL1::TCF3	PAX1	TP73	TBXT
TBP	PITX3	VEZF1	Tcf21
TCF21(var.2)	PKNOX1	ZBTB6	TFEC
TFCP2	Six3	ZEB1	TGIF1
THRB	TEF	ZNF135	TGIF2
VENTX	TFEB	ZNF143	THRB(var.2)
ZBTB18	ZNF652	ZNF354C	ZNF382

Examples:

Cluster 10:
High expression of TFs associated with the circadian clock, i.e. CLOCK, CREM and muscle-specific genes, i.e. MYF5, MYF6, TBX5 (heart), but other organ develoment too

Cluster 7:
High expression of Tfs associated with immune responses, i.e. ATF7, CEBPB, CEBPG, CEBPE (bone marrow),

ToDos

- Test snakemake pipeline again with more RAM (Monday)

- Improve extraction of defining Tfs

- Finishing interfaces