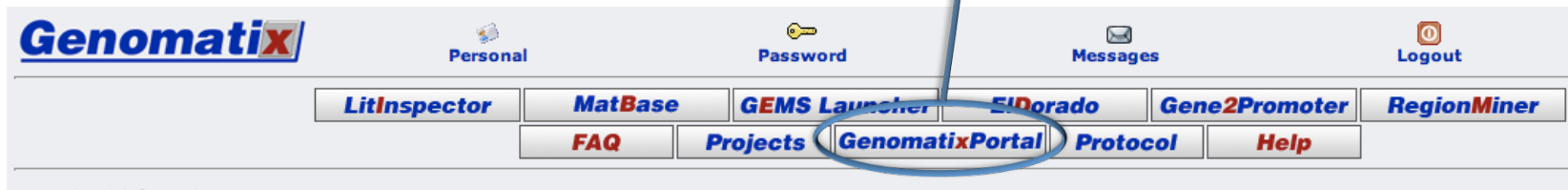
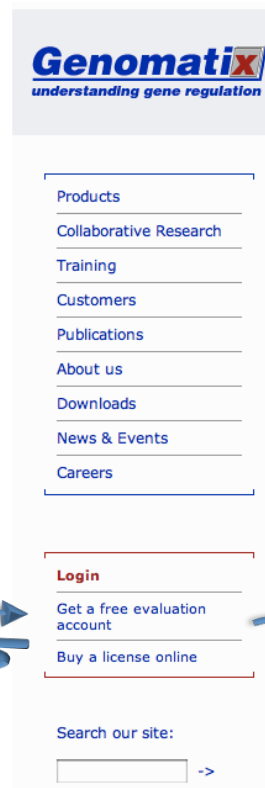


MatInspector

www.genomatix.de

Only 1000 € / mo.



FAQ

GenomatixPortal

Help

[Release Notes](#)

GeneRanker

Characterization of Gene Sets

LitInspector

Literature & Pathway Mining

MatInspector

Search Transcription Factor Binding Sites

MatBase

Transcription Factor Knowledge Base

Unlimited

10 queries/mo

Projects

Protocol

EIDorado

Extended Genome Annotation

Gene2Promoter

Retrieve & Analyze Promoters

GEMS Launcher

Sequence Analysis & Modeling Software

RegionMiner

Analyze Genomic Regions

Gene Name Input

☐ Search promoters by gene

Start typing a gene name and select a gene from the drop-down list to analyze its promoter
enter gene

This feature is only available for licensed users of EIDorado/Gene2Promoter

or Sequence Input

☐ Choose from your previously uploaded sequences

CMV.seq
clump.seq
dscoles.seq
dscoles_1.seq
dscoles_2.seq

☐ or enter the correctly formatted DNA sequence(s)

Supported formats: plain, EMBL, FASTA, GCG/RSF, GenBank, IG

Name for your sequence-file: dscoles_7.seq

☐ or upload a file containing sequence(s) (max. 100 MB)

(Choose file) no file selected

Optional name for your sequence-file on the server:

☐ or enter accession number(s)

(separated by spaces or commas)

[Search corresponding promoters for your sequence\(s\)](#)

☐ Check your sequence(s) against our genome database to find corresponding promoters for analysis

Homo sapiens

This feature is only available for licensed users of EIDorado/Gene2Promoter

or Database Input

Library Selection

Please select one of the following libraries:

Transcription factor binding sites (Weight matrices)
User-defined IUPAC string
Plant IUPAC library (based on PLACE)
Restriction Sites: 4-cutters

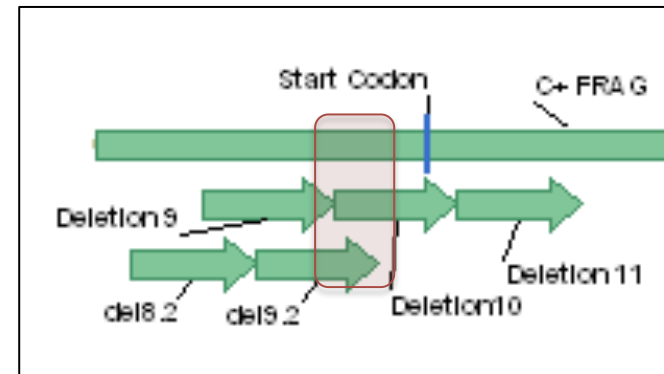
- See the [list](#) of available weight matrices
- See the [list](#) of available IUPAC library strings

[Continue](#) [Reset Form](#)

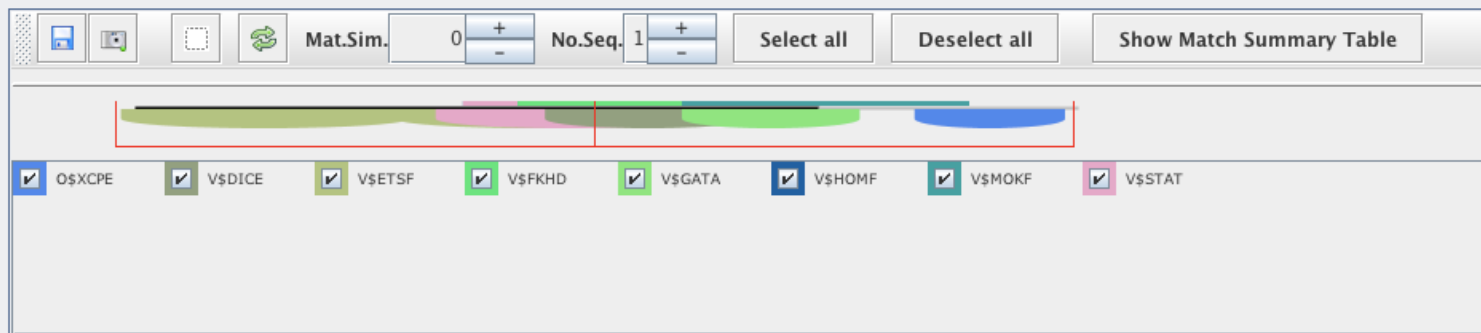
Note: The output is restricted to ca. 5000 matches!

MatInspector analysis

Right side of 9



Graphical overview:



Problems displaying the Java graphics? See our [Technical FAQ](#)

Inspecting sequence dscoles_3 (1 - 70):

[Show/hide columns](#) **Detailed Family Information** [Show all columns](#)

[Need help on the interactive features](#)

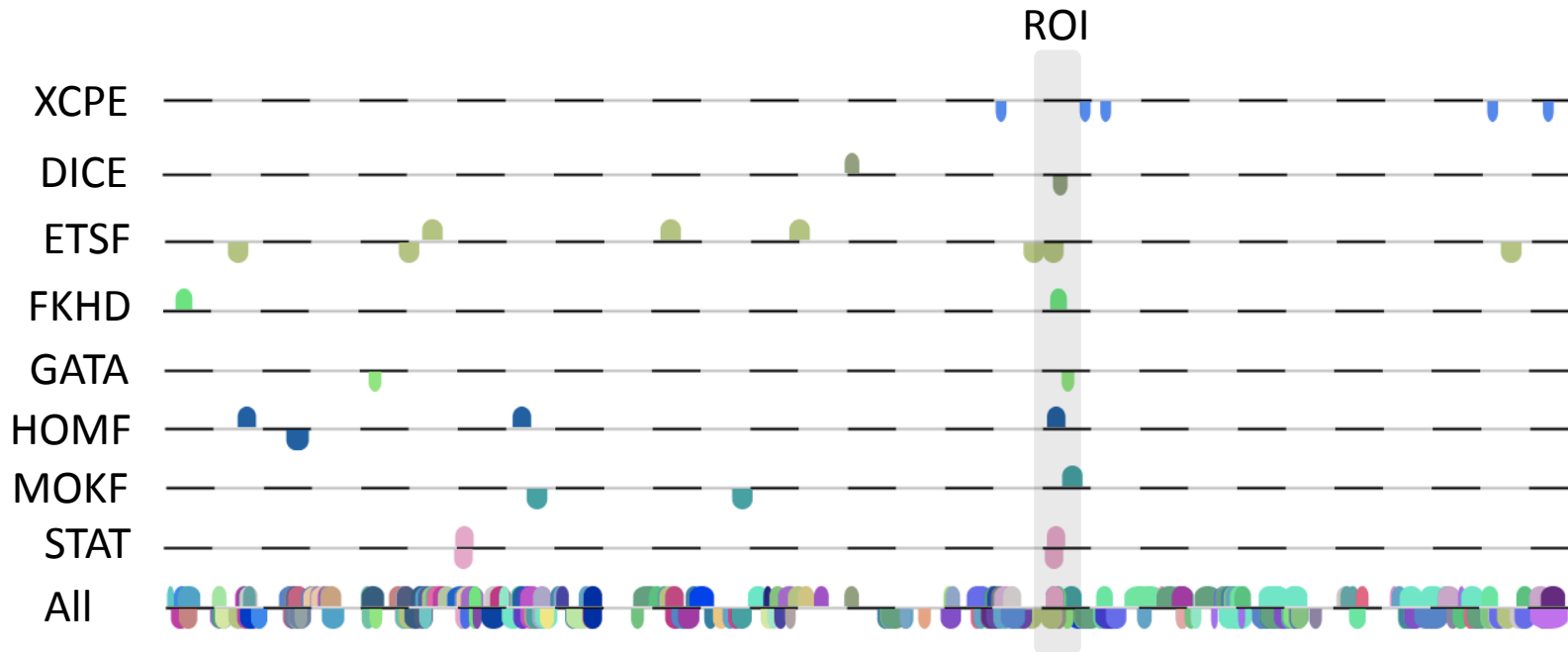
10 matches found in this sequence

Displayed matches: 10

Check transcription factor <-> matrix family assignment										Additional lines of evidence		Sequence (red: ci-value > 60 CAPITALS: core sequence)
Matrix Family	Detailed Family Information	Matrix	from	to	anchor	Strand	Matrix sim.	#	Evidence			
V\$ETSF	Human and murine ETS1 factors	V\$CETS1P54.01	21	41	31	(-)	0.995	1	part of model match of IRFF_ETSF_02_STAT_ETSF_03		tctttaCGGaaagtcggaggg	ctctttaccGGAagtcgga ggtcagacGGAagcagaacgt cgacttcgGTAaagagtc ggGTGGgagcg cgacttcCGGTaaagagtc ttccggTAAAgagtcgcc gggaCTCTtaccgg gtccctatccgcaCCTCgct
V\$GATA	GATA binding factors	V\$GATA1.06	42	54	48	(-)	0.966	0			tgcgGATAgggac	
V\$STAT	Signal transducer and activator of transcription	V\$STAT.01	24	42	33	(-)	0.948	1	part of model match of STAT_ETSF_03		ctctttaccGGAagtcgga	
V\$ETSF	Human and murine ETS1 factors	V\$ELK1.02	1	21	11	(-)	0.920	0			ggtcagacGGAagcagaacgt	
V\$STAT	Signal transducer and activator of transcription	V\$STAT.01	26	44	35	(+)	0.910	0			cgacttcgGTAaagagtc	ggGTGGgagcg cgacttcCGGTaaagagtc ttccggTAAAgagtcgcc gggaCTCTtaccgg gtccctatccgcaCCTCgct
O\$XCPE	Activator-, mediator- and TBP-dependent core promoter element for RNA polymerase II transcription from TATA-less promoters	O\$XCPE1.01	59	69	64	(-)	0.885	0			ggGTGGgagcg	
V\$HOMF	Homeodomain transcription factors	V\$TLX1.01	26	44	35	(+)	0.878	0			cgacttcCGGTaaagagtc	
V\$FKHD	Fork head domain factors	V\$FREAC2.01	30	46	38	(+)	0.850	0			ttccggTAAAgagtcgcc	
V\$DICE	Downstream Immunoglobulin Control Element, critical for B cell activity and specificity	V\$DICE.01	32	46	39	(-)	0.831	0			gggaCTCTtaccgg	
V\$MOKF	Mouse Krueppel like factor	V\$MOK2.01	42	62	52	(+)	0.750	0			gtccctatccgcaCCTCgct	

A total of 10 matches were found in 1 sequence!

Right side of 9, cont.



	Ets Core	TTCC		MS
ROI	gaccctccgacttccggtaaagagtcctatccgca			
DICE (-)	acgttctgcttccgtctgaccctccgacttccggtaaAGAGtccctatccgcacctccgctcccaccggc			0.831
ETSF (-)	acgttctgcttccgtctgaccctccgacttCCGGTaaagagtcctatccgcacctccgctcccaccggc			0.920
FKHD	acgttctgcttccgtctgaccctccgacttccggtAAAgagtcctatccgcacctccgctcccaccggc			0.850
GATA (-)	acgttctgcttccgtctgaccctccgacttccggtaaagagtcctTATCcgacactccgctcccaccggc			0.966
HOMF	acgttctgcttccgtctgaccctccgacttccggtaaagagtcctatccgcacctccgctcccaccggc			0.878
MOKF	acgttctgcttccgtctgaccctccgacttccggtaaagagtcctatccgcaCCTCcgctcccaccggc			0.750
STAT	acgttctgcttccgtctgaccctccgactTTCcggtaaagagtcctatccgcaCCTCcgctcccaccggc			0.910
STAT (-)	acgttctgcttccgtctgaccctccgacttccgGTAAagagtcctatccgcaCCTCcgctcccaccggc			0.948

Caps = Core Sequence, the four highest conserved consecutive positions

Red = *Ci* vector (consensus index vector) > 60, ranges from 0-100

MS = Matrix similarity. Good match > 0.8, perfect match = 1.0

Site of CGA mutation

ROI	Ets Core	TTCC	MS
	gacccctccgacttccggtaaagagtccctatccgca		
DICE (-)	acgttctgcttccgtctgacccctccgacttccggtaaAGAGtccctatccgcacctccgctcccaccgcgc		0.831
ETSF (-)	acgttctgcttccgtctgacccctccgacttCCGGtaaagagtccctatccgcacctccgctcccaccgcgc		0.920
FKHD	acgttctgcttccgtctgacccctccgacttccggTAAAgagtccctatccgcacctccgctcccaccgcgc		0.850
GATA (-)	acgttctgcttccgtctgacccctccgacttccggtaaagagtcccTATCcgacctccgctcccaccgcgc		0.966
HOMF	acgttctgcttccgtctgacccctccgacttccGGTaaagagtccctatccgcacctccgctcccaccgcgc		0.878
MOKF	acgttctgcttccgtctgacccctccgacttccggtaaagagtccctatccgcaCCTCcgctcccaccgcgc		0.750
STAT	acgttctgcttccgtctgacccctccgacTTCCggtaaagagtccctatccgcacctccgctcccaccgcgc		0.910
STAT (-)	acgttctgcttccgtctgacccctccgacttccgGTAAagagtccctatccgcacctccgctcccaccgcgc		0.948
Primer MutR1	5' -ccgtctgacccctccgactcgaaggtaaagagtccc-3'		Cloned
Primer MutR2	5' -ccgtctgacccctc-----gagtcctatccgc-3'		Cloned

Knocking out STAT

These changes add XhoI

Quikchange Primer MutR1

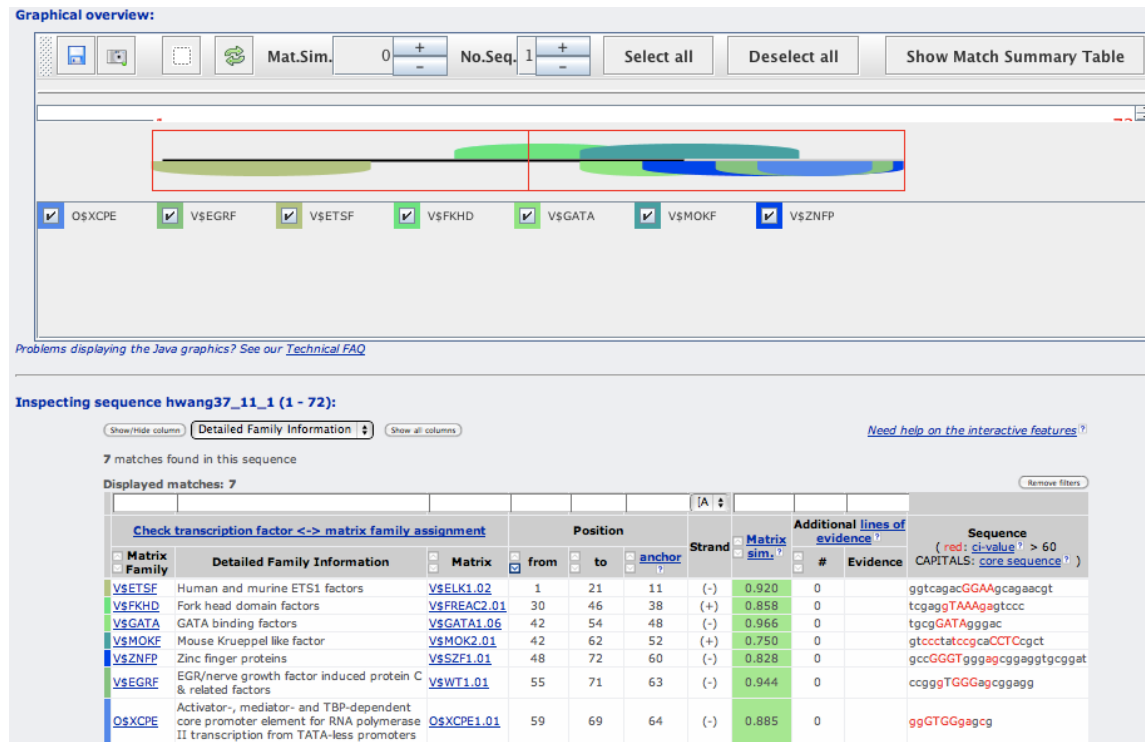
STAT acgttctgcttccgtctgacccctccgac**TTCC**ggtaaagagtcctatccgcaCCTCcgctcccacccggc 0.910
 Changes acgttctgcttccgtctgacccctccgactcgaggtaaagagtcctatccgcaCCTCcgctcccacccggc
 Primer MutR1 5'-ccgtctgacccctccgactcgaggtaaagagtcct-3'

Start

DICE
 ETSF
 FKHD
 GATA
 HOMF
 MOKF
 STAT

Finish

ETSF
 FKHD
 GATA
 ZNFP
 MOKF



This change adds XhoI

Quikchange Primer MutR2

acgtttctgcttccgtctgacccctccgacttccggtaaagagtcacctatccgcacctccgctcccacccggc

Changes acgtttctgcttccgtctgacccctc-----gagtcacctatccgcacctccgctcccacccggc

Primer MutR2 5'-ccgtctgacccctcgcagtcacctatccgc-3'

Start

DICE

ETSF

FKHD

GATA

HOMF

MOKF

STAT

Finish

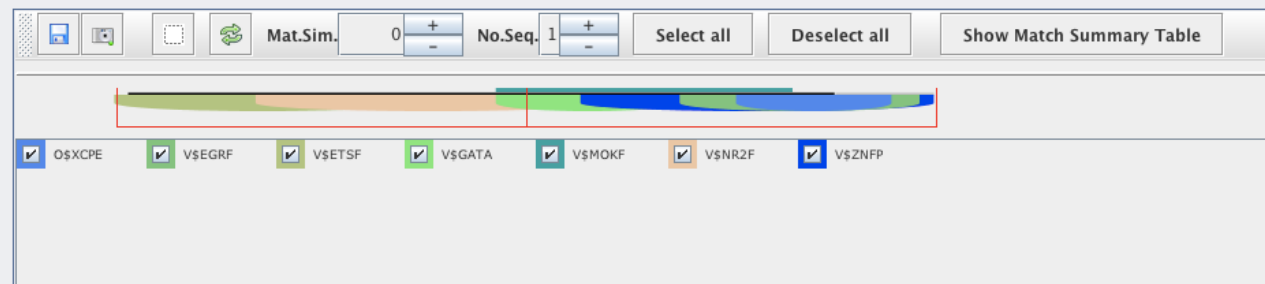
GATA

NR2F (MS<0.8)

ZNFP

MOKF

Graphical overview:



Problems displaying the Java graphics? See our [Technical FAQ](#)

Inspecting sequence hwang37_5 (1 - 58):

[Show/Hide return](#) [Detailed Family Information](#) [Show all columns](#)

[Need help on the interactive features?](#)

7 matches found in this sequence

Displayed matches: 7

Check transcription factor <-> matrix family assignment												[A]					
Matrix Family	Detailed Family Information	Matrix	Position			Strand	Matrix sim.	Additional lines of evidence		Evidence	Sequence (red: ci-value > 60 CAPITALS: core sequence:)						
			from	to	anchor			#									
V\$ETSF	Human and murine ETS1 factors	V\$ELK1.02	1	21	11	(-)	0.920	0			ggtcagacGGAAgcagaacgt						
V\$NR2F	Nuclear receptor subfamily 2 factors	V\$TR2.01	11	35	23	(-)	0.792	0			atagggaactcgaggGGTCagacgga						
V\$GATA	GATA binding factors	V\$GATA1.06	28	40	34	(-)	0.966	0			tcgcGATAaggac						
V\$MOKF	Mouse Krueppel like factor	V\$MOK2.01	28	48	38	(+)	0.750	0			gtccctatccgcaCCTCcgct						
V\$ZNFP	Zinc finger proteins	V\$SZF1.01	34	58	46	(-)	0.828	0			gccGGGTgggagcggaggctcgga						
V\$EGRF	EGF/nerve growth factor induced protein C & related factors	V\$WT1.01	41	57	49	(-)	0.944	0			ccgggTGGGagcggagg						
O\$XCPE	Activator-, mediator- and TBP-dependent core promoter element for RNA polymerase II transcription from TATA-less promoters	O\$XCPE1.01	45	55	50	(-)	0.885	0			ggGTGGgagcg						

These changes result in loss of STAT ... adds Eagl

Quikchange Primer MutR3

acgttctgtcttccgtctgacccctccgacttccggtaaagagtcctatccgcacctccgctcccaccgcc

Changes acgttctgtcttccgtctgacccctccgacttccg-----ccgcacctccgctcccaccgcc
Primer MutR3 5'-ctccgacttccggccgcacctccgc-3'

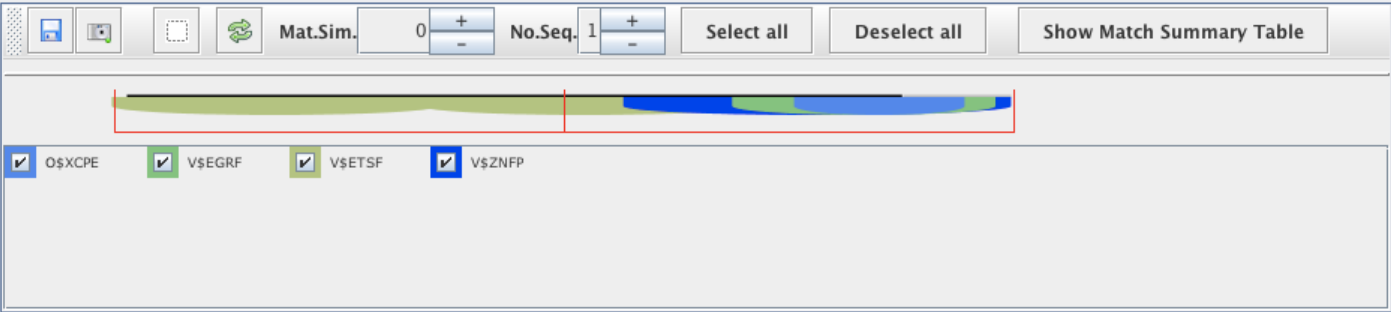
Start

DICE
ETSF
FKHD
GATA
HOMF
MOKF
STAT

Finish

ETSF
ZNFP

Graphical overview:



Problems displaying the Java graphics? See our [Technical FAQ](#)

Inspecting sequence hwang37_6 (1 - 58):

[Show/Hide column](#) [Detailed Family Information](#) [Show all columns](#) [Need help on the interactive features](#)

5 matches found in this sequence

Displayed matches: 5

Matrix Family	Detailed Family Information	Matrix	from	to	anchor	Strand	Matrix sim.	#	Evidence	Sequence (red: ci-value > 60 CAPITALS: core sequence)
V\$ETSF	Human and murine ETS1 factors	V\$ELK1.02	1	21	11	(-)	0.920	0		ggtcagacGGAAgcagaacgt
V\$ETSF	Human and murine ETS1 factors	V\$ELK1.02	21	41	31	(-)	0.983	1	part of model match of SP1F_ETSF_05	gtgcggccGGAAgtcggagg
V\$ZNF	Zinc finger proteins	V\$SZF1.01	34	58	46	(-)	0.828	0		gccGGTggagcggaggtcggcc
V\$EGRF	EGR/nerve growth factor induced protein C & related factors	V\$WT1.01	41	57	49	(-)	0.944	0		ccgggTGGGagcggagg
O\$XCPE	Activator-, mediator- and TBP-dependent core promoter element for RNA polymerase II transcription from TATA-less promoters	O\$XCPE1.01	45	55	50	(-)	0.885	0		ggGTGGagcg

Summary “right of 9”

Start

DICE
ETSF
FKHD
GATA
HOMF
MOKF
STAT

1

Finish

ETSF
FKHD
GATA
ZNFP
MOKF

2

Finish

GATA
NR2F (MS<0.8)
ZNFP
MOKF

3

Finish

ETSF
ZNFP

Left side transcription factors

SPZ1 = Testis-specific bHLH-Zip transcription factors, TSP1, Spermatogenic leucine zipper 1.

EGRF = EGR/nerve growth factor induced protein C & related factors

SP1F = GC-Box factors SP1/GC

DMTF = Cyclin D binding myb-like transcription factor

CTCF = CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains

KLFS = Krueppel like transcription factors

Right side transcription factors

DICE = Downstream Immunoglobulin Control Element, critical for B cell activity and specificity

ETSF = Human and murine ETS1 factors

FKHD = Fork head domain factors

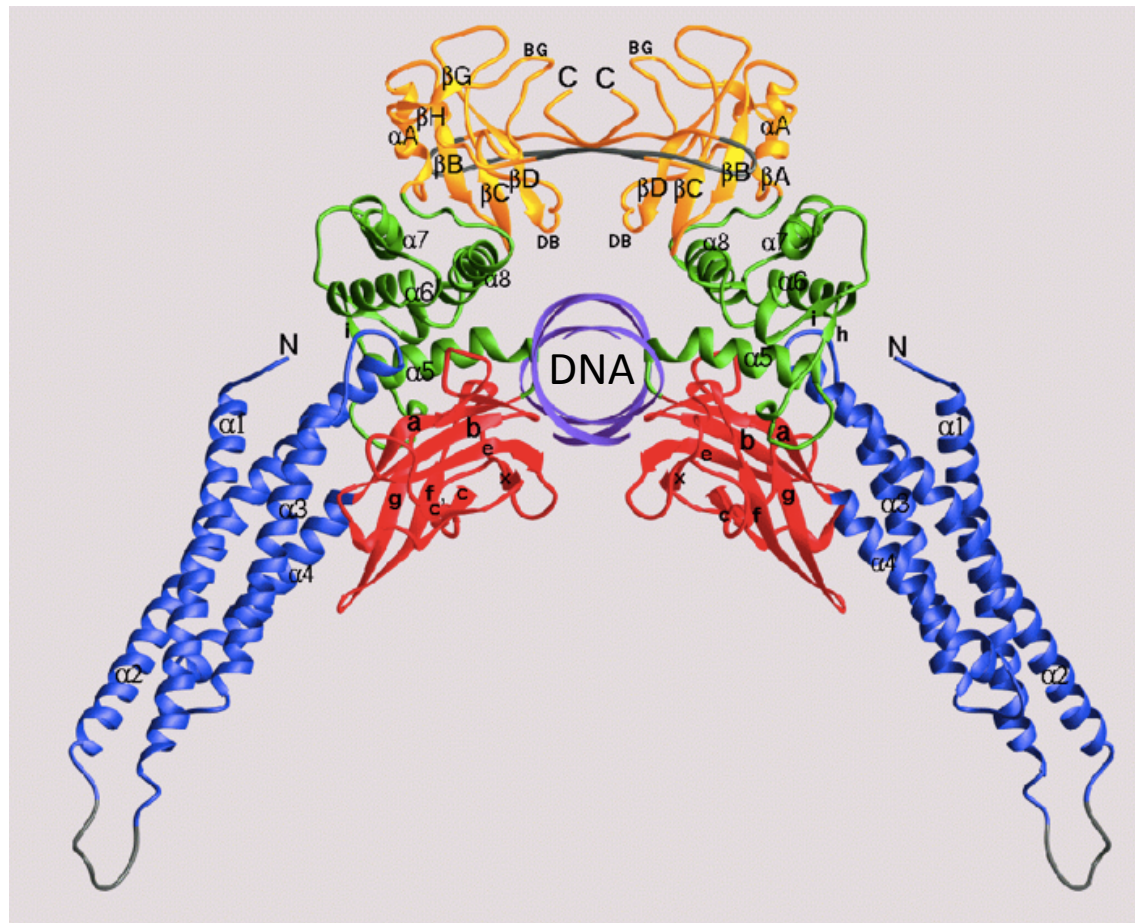
GATA = GATA binding factors

HOMF = Homeodomain transcription factors

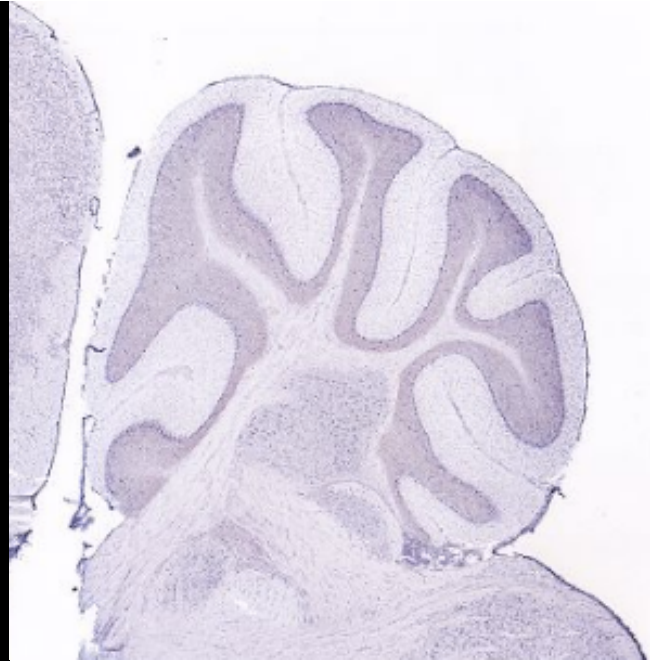
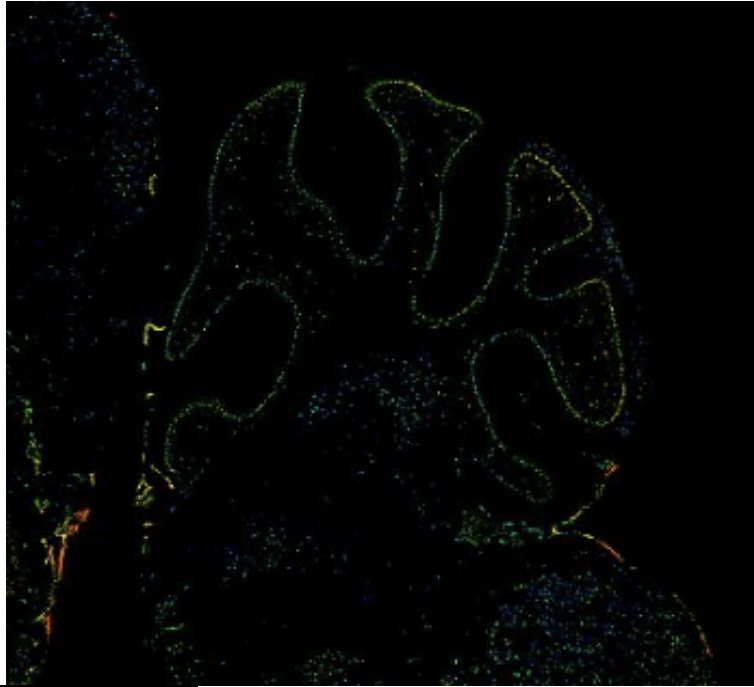
MOKF = Mouse Krueppel like factor

STAT = Signal transducer and activator of transcription

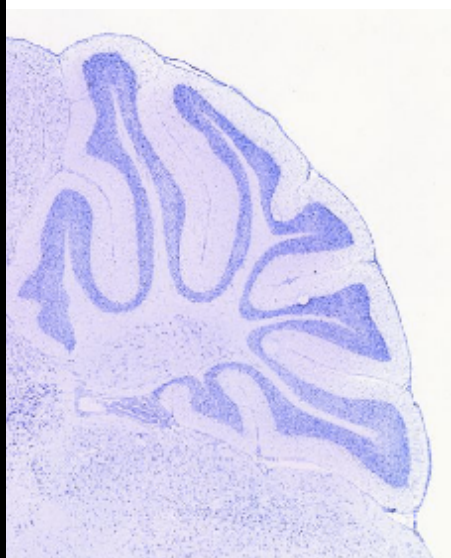
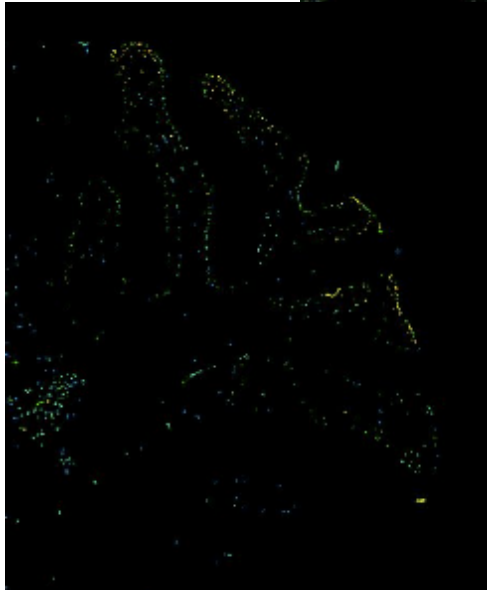
Structure of STAT on DNA



STAT1



STAT5



No expression in PC

STAT2

STAT3

STAT4

STAT6

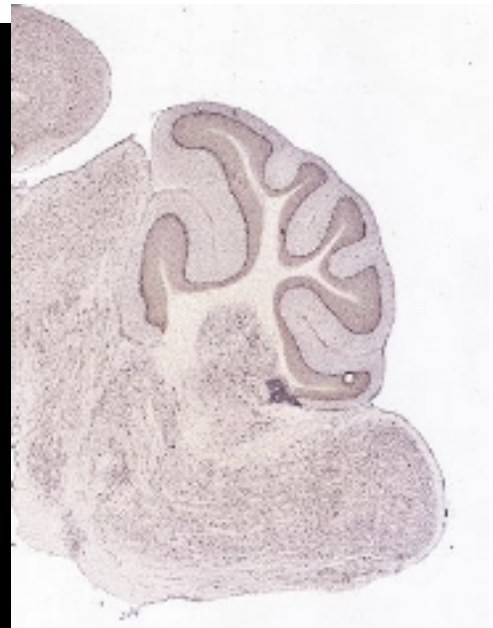
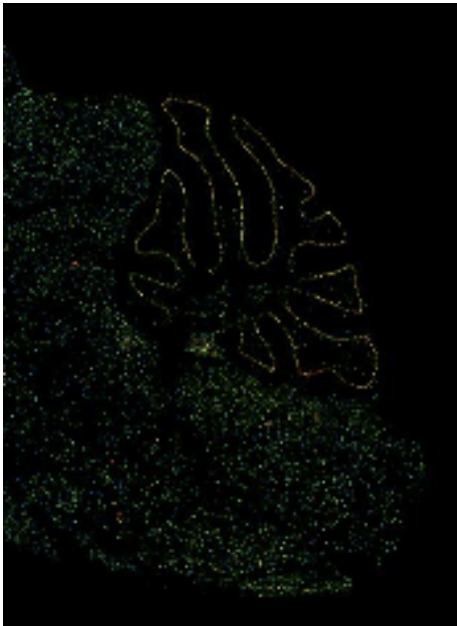
DICE

The transcription factor is GTF2I REPEAT DOMAIN-CONTAINING PROTEIN 1

Encoded by *GTF2IRD1*

“general transcription factor”

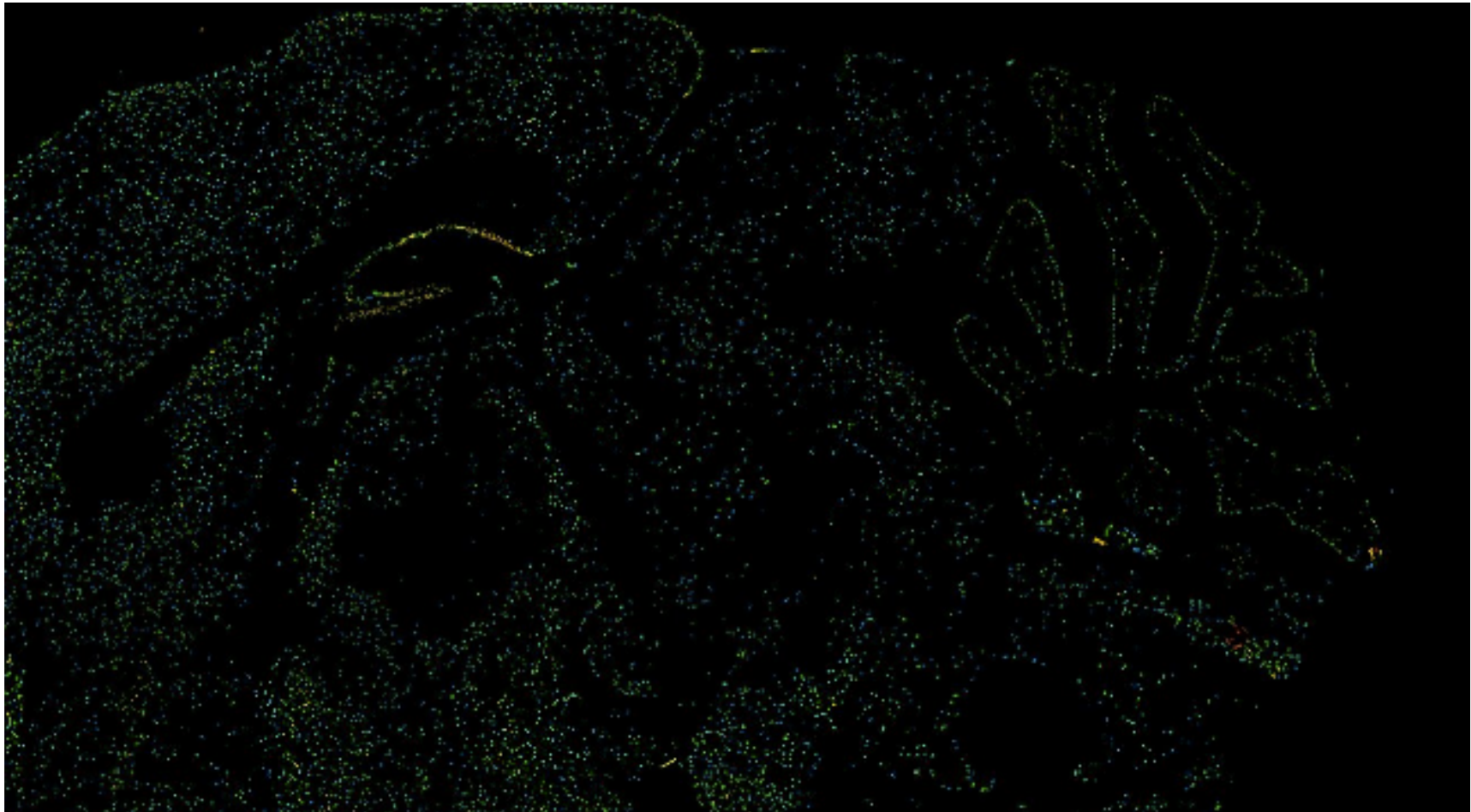
Deleted in classic Williams syndrome



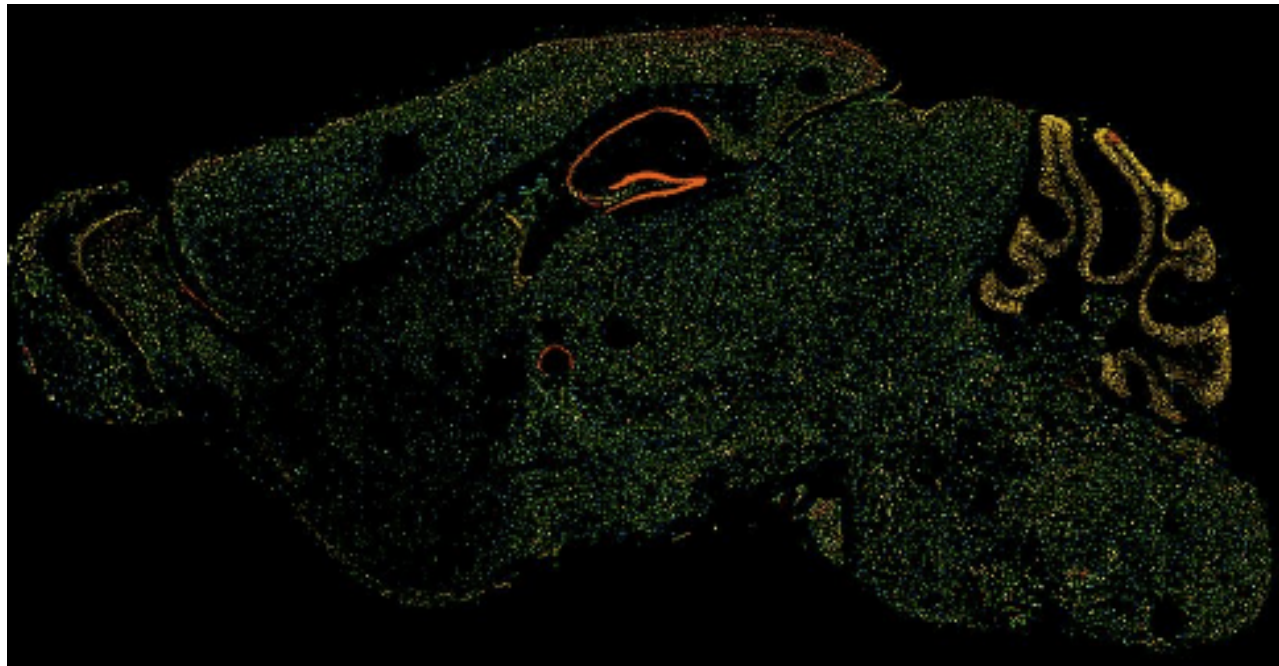
DMTF1

CYCLIN D-BINDING MYB-LIKE TRANSCRIPTION FACTOR 1; DMTF1

Staining is weak in PC



CCCTC-BINDING FACTOR; CTCF
Transcriptional repressor
Binds unmethylated regions



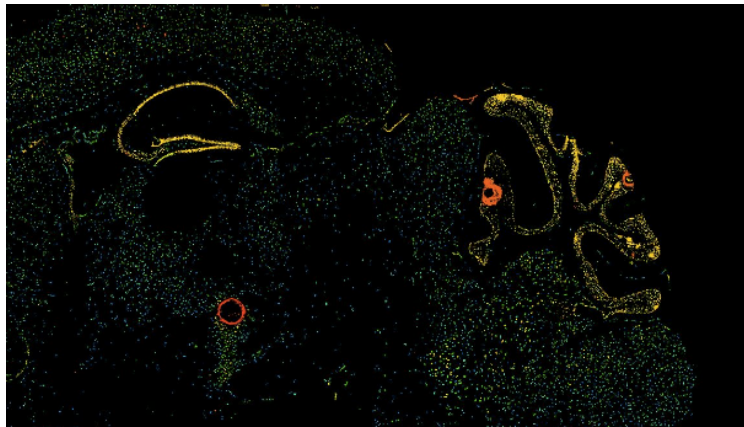
KLFS Kruppel like transcription factors

No cerebellar expression by Allen Atlas: KLF1,2,4,5,6,8,10,12,13,14,15,17

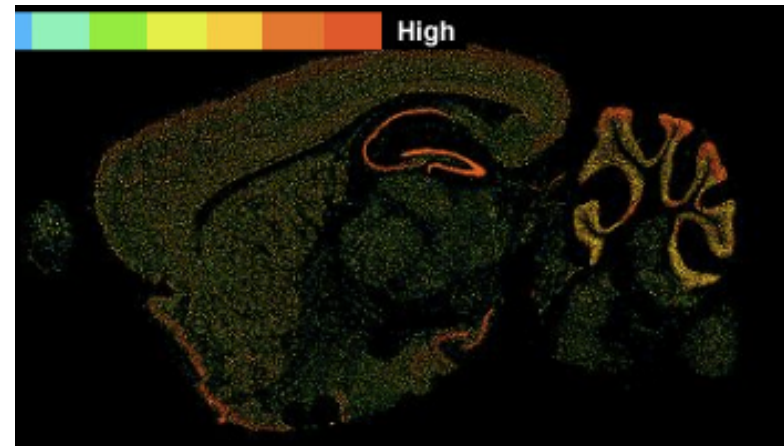
Maybe weak expression for KLF16

There is no KLF11

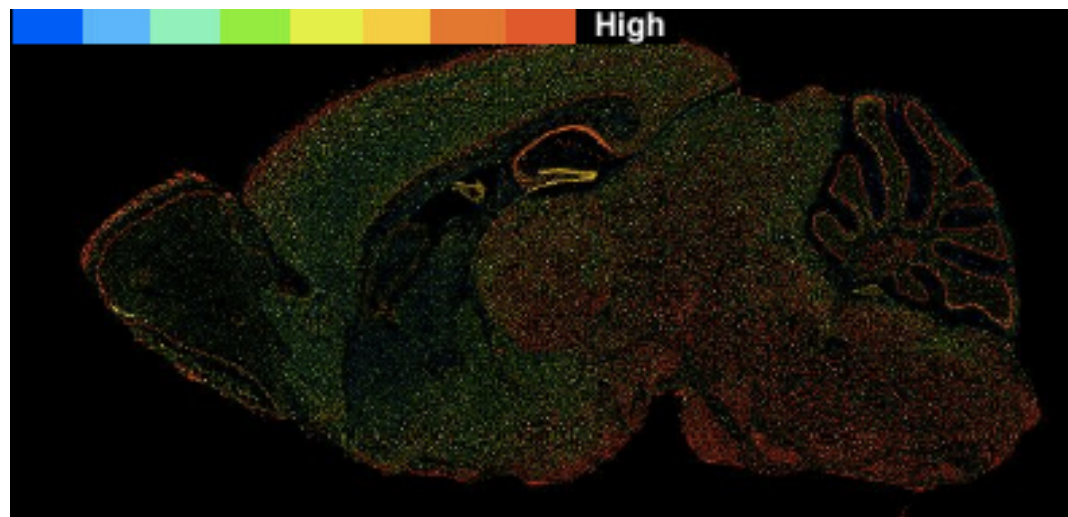
KLF3



KLF9



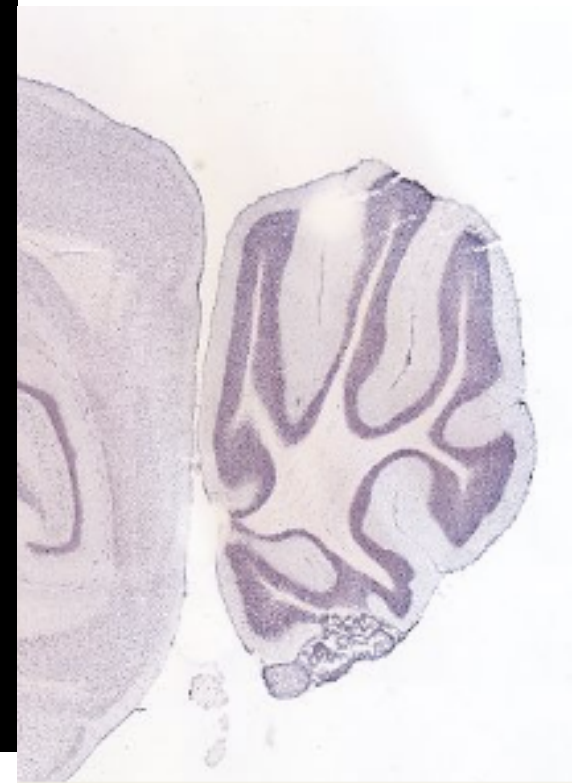
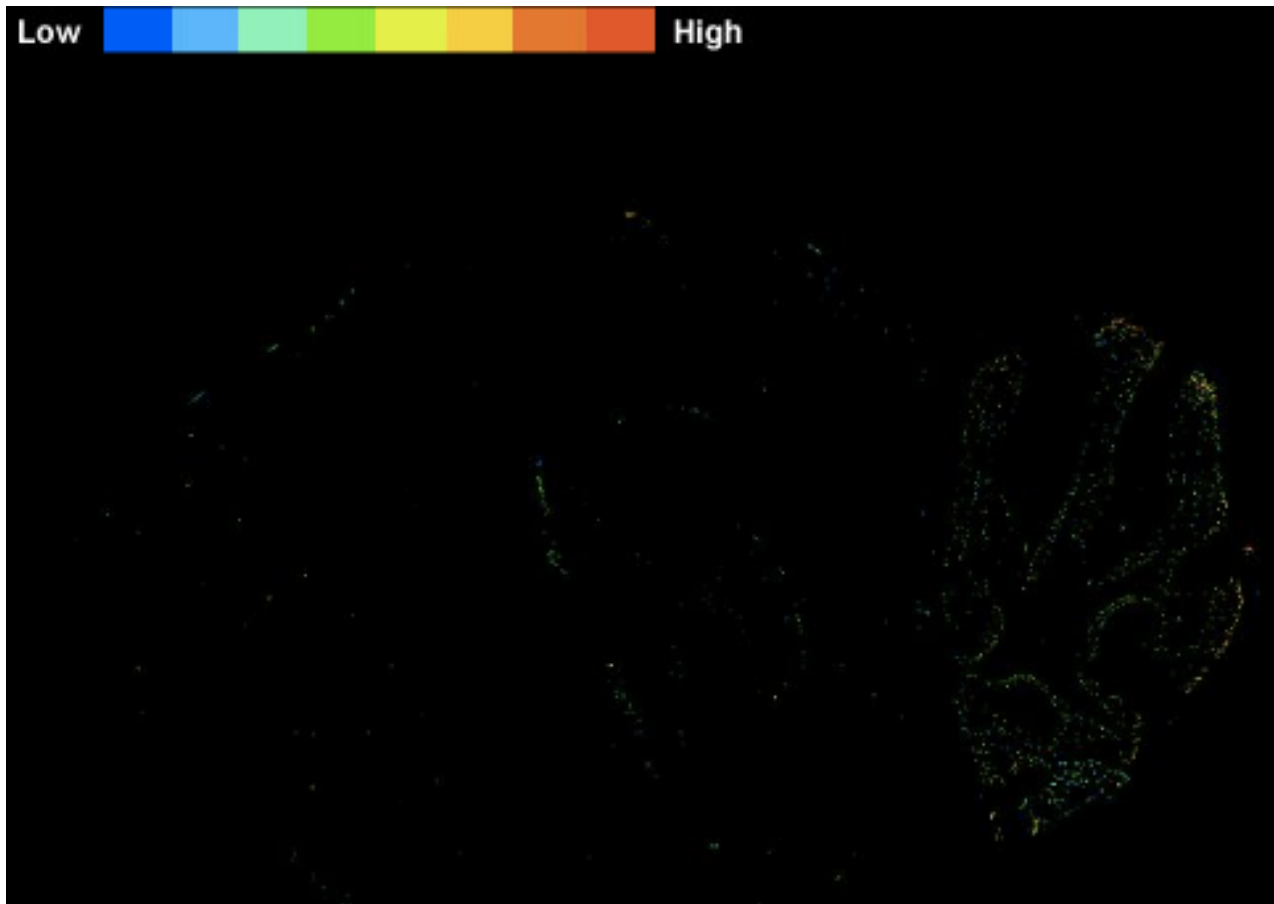
KLF7
“ubiquitous”



SP1F

Specificity Protein 1

Belongs to the Sp/KLF family of transcription factors.

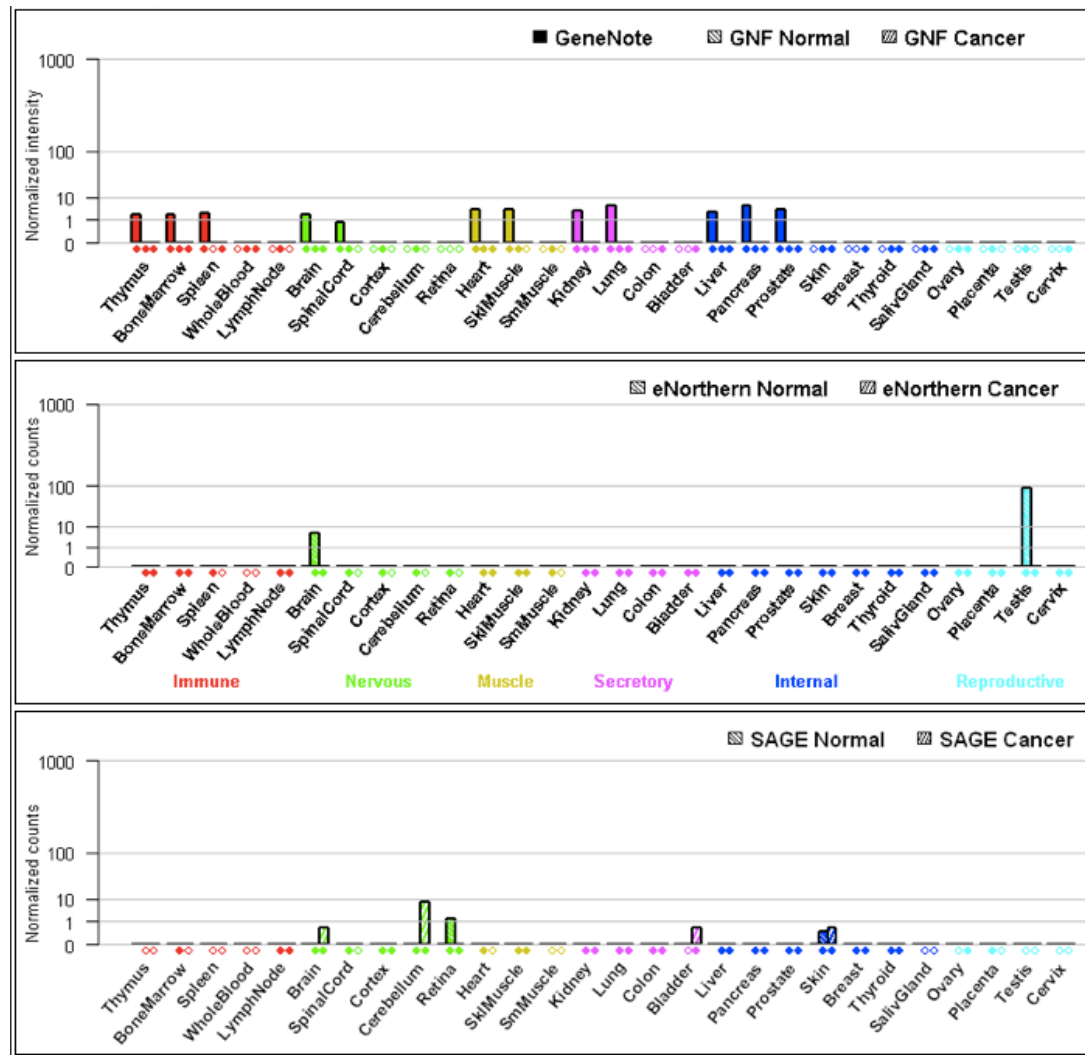


Expression very weak in cerebellum

SPZ1 in GeneCards

Testis-specific bHLH-Zip transcription factors, TSP1, Spermatogenic leucine zipper 1.

Nothing evident in Allen Atlas



ETSF

ETS (E-twenty six) family

The largest transcription factor family in mammalian cells

29 genes in humans

Can be activators or repressors of transcription

Subfamily	Mammalian family members	Invertebrate orthologs
ELF	ELF1, ELF2 (NERF), ELF4 (MEF)	
ELG	GABPα	ELG
ERG	ERG, FLI1, FEV	
ERF	ERF (PE2), ETV3 (PE1)	
ESE	ELF3 (ESE1/ESX), ELF5 (ESE2), ESE3 (EHF)	
ETS	ETS1, ETS2	POINTED
PDEF	SPDEF (PDEF/PSE)	
PEA3	ETV4 (PEA3/E1AF), ETV5 (ERM), ETV1 (ER81)	
ER71	ETV2 (ER71)	
SPI	SPI1 (PU.1), SPIB, SPIC	
TCF	ELK1, ELK4 (SAP1), ELK3 (NET/SAP2)	LIN
TEL	ETV6 (TEL), ETV7 (TEL2)	YAN

ETS family transcription factors in PCs:

ER81, ETS1, ETS2 (not comprehensive)

HOMF

Homeodomain transcription factors

Homeodomain = Helix-turn-Helix

I counted 91 of these here:

<http://www.arabidopsis.org/browse/genefamily/Homeobox.jsp>

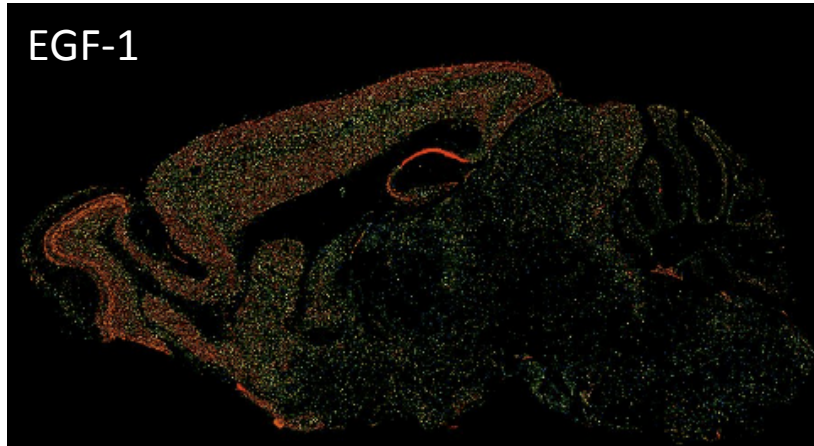
EGFR

EGR/nerve growth factor induced protein C & related factors

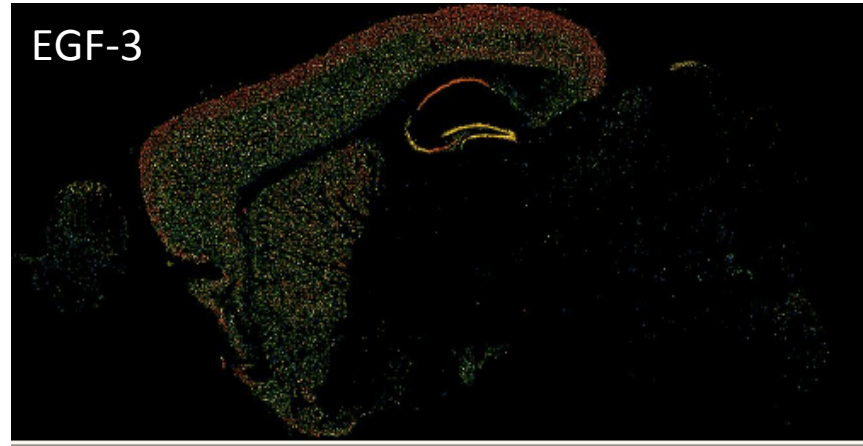
EGR=Early Growth Response

In the nervous system the main ones are

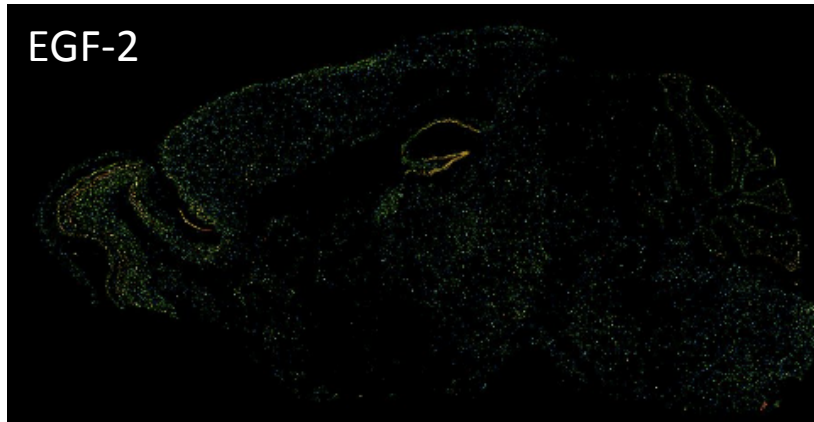
EGF-1



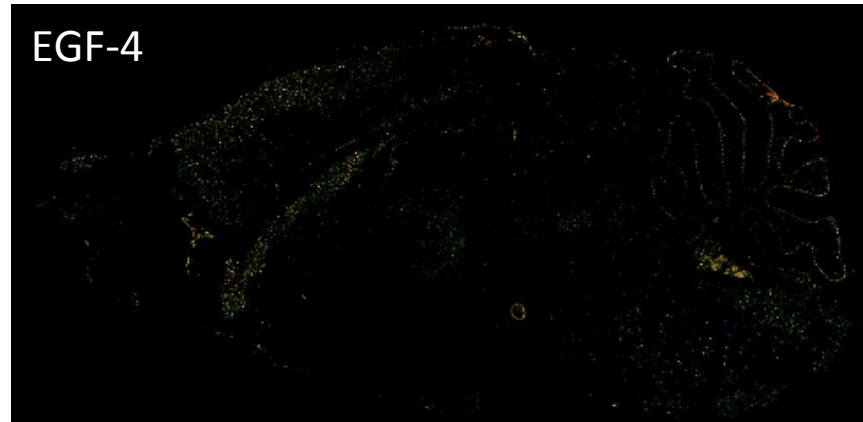
EGF-3



EGF-2



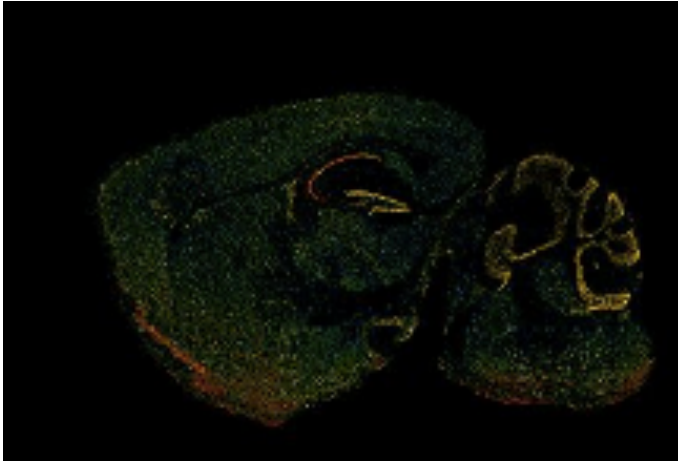
EGF-4



GATA

Transcription factors are GATA binding proteins GATA1-GATA6

GATA3



No cerebellar expression for
GATA1,2,5,6



GATA4
(differences among
Sections)

FKHD

Forkhead proteins, FOX
Helix-Turn-Helix TFs

Genes

- FOXA1, FOXA2, FOXA3 (See also [Hepatocyte nuclear factors](#))
- FOXB1, FOXB2
- FOXC1, FOXC2
- FOXD1, FOXD2, FOXD3, FOXD4, FOXD5, FOXD6
- FOXE1, FOXE2, FOXE3
- FOXF1, FOXF2
- FOXG1
- FOXH1
- FOXI1, FOXI2
- FOXJ1, FOXJ2, FOXJ3
- FOXK1, FOXK2
- FOXL1, FOXL2
- FOXM1
- FOXN1, FOXN2, FOXN3, FOXN4
- FOXO1, FOXO2, FOXO3, FOXO4
- FOXP1, FOXP2, FOXP3, FOXP4
- FOXQ1
- FOXR1, FOXR2

Start codon mutagenesis

What we tried in previous effort:

cccgccctccgatgcgctcagcg-3 '
Primer Start1 5' -ccgccctccgCtgcgctcagc-3 '

Note, running the primer design tool returned a slightly longer oligo sequence

What we are trying this time, introducing an XhoI:

cccgccctccgatgcgctcagcg
Primer Start1Mut 5' -cgcgcccgccctcGAGtgcgctcagcggc-3 '

Didn't get in 1st try
last week