

Анализ совместной встречаемости мотивов пионерских и партнерских транскрипционных факторов на основе данных ChIP-Seq по массовому секвенированию

Выполнил: магистрант КИБ Радица Владимир Вячеславович

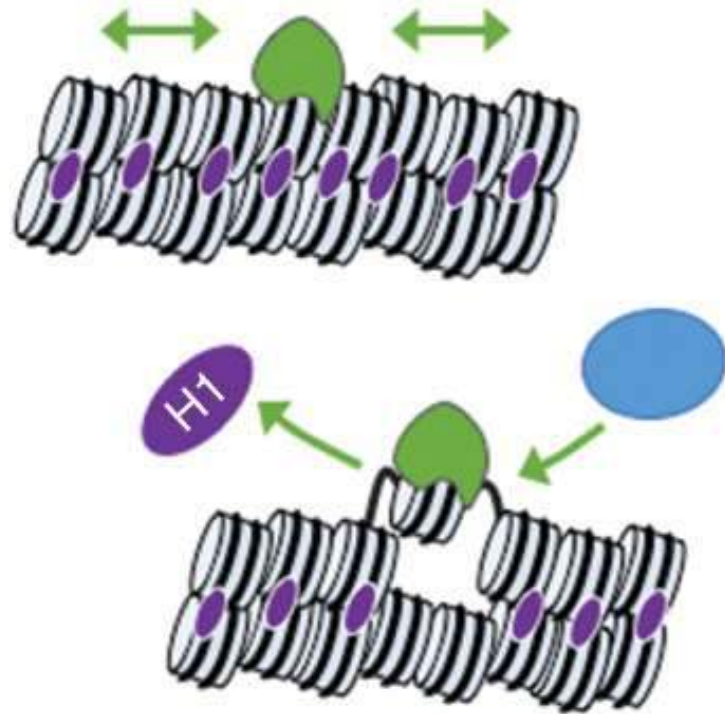
Научный руководитель: к.б.н. Левицкий Виктор Георгиевич
с.н.с лаборатории эволюционной биоинформатики и теоретической генетики ИЦиГ СО РАН

Актуальность

- 1) Транскрипция генов эукариот производится под контролем транскрипционных факторов.
- 2) Взаимодействия транскрипционных факторов с ДНК как правило кооперативные, то есть транскрипция гена происходит за счет взаимодействий не одного, а сразу нескольких транскрипционных факторов.
- 3) Развитие методов полногеномного картирования сайтов связывания транскрипционных факторов дает возможность исследовать специфичность совместного действия транскрипционных факторов.

Пионерские транскрипционные факторы

Пионерский транскрипционный фактор

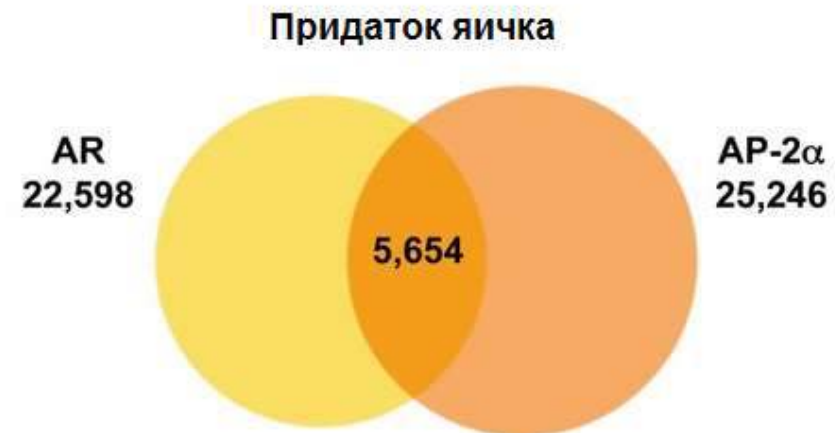
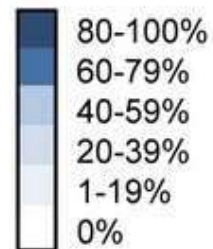


Пионерские транскрипционные факторы являются группой регуляторных белков способных распознавать и связываться с закрытыми участками хроматина.

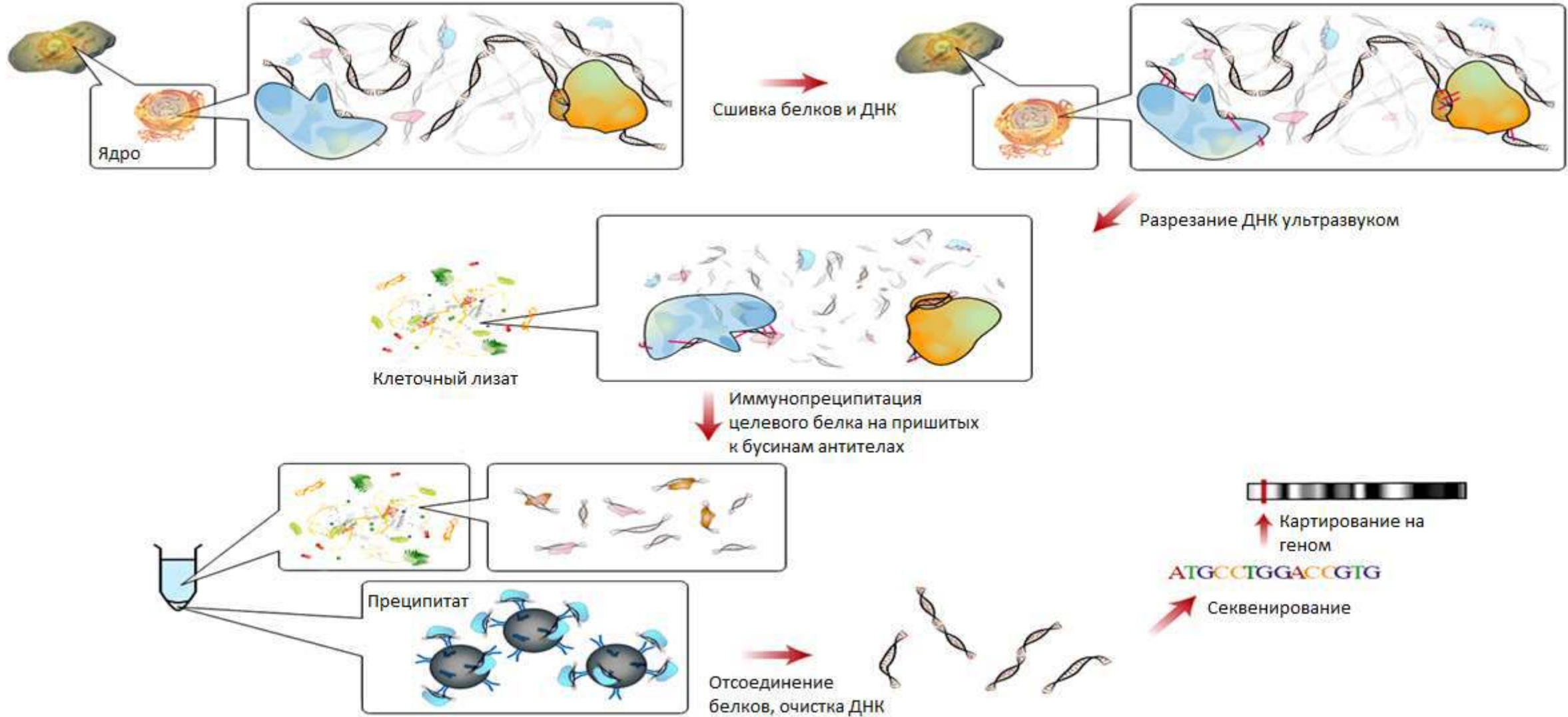
Взаимодействие пионерских транскрипционных факторов с нуклеосомной ДНК приводит к открытию хроматина, тем самым создавая нужные условия для активации транскрипции гена.

Тканеспецифичность транскрипционных факторов

		Pro	Kid	Epi
Nuclear receptor	AR	80-100%	60-79%	80-100%
	ER	60-79%		40-59%
	Hnf4 α		80-100%	
	RXR		80-100%	
	COUP-TF		80-100%	
	NR2F6		80-100%	
	Nur77		80-100%	
	PPAR α	80-100%	80-100%	80-100%
Forkhead	FoxA1	40-59%		
	FoxA2	40-59%		
	NFIC	80-100%	80-100%	
Helix-loop-helix	Tcf12	80-100%	60-79%	
	AP-2 α			80-100%
	AP-2 β			40-59%
	AP-2 γ			40-59%
	E2F	80-100%	80-100%	80-100%
$\beta\beta\alpha$ -zinc finger	SP1		60-79%	
	SP2	80-100%		80-100%
	Mtf1	80-100%		80-100%
	Zfp691	80-100%	40-59%	60-79%
	Znf263	80-100%		80-100%



ChIP-Seq



Cistrome data browser



Не защищено | cistrome.org/db/#/



- Check what factors regulate your gene of interest, what factors bind in your interval or have a significant binding overlap with your peak set. Have a try at [CistromeDB Toolkit](#).
- If you have a **Transcription Factor** ChIP-seq (and TF perturbed expression) data, [Cistrome-GO](#) help you predict the function of this TF.
- Please help us curate the samples which has incorrect meta-data annotation by clicking the button on the inspector page. Thank you!

Containing word(s):

Search

Options ▾

Species

All

Homo sapiens
Mus musculus

Biological Sources

All

1-cell pronuclei
1015c
10326
1064Sk
106A

Factors

All

AATF
ABCC9
ACSS2
ACTB
ADNP

Results

Batch	Species	Biological Source	Factor	Publication	Quality Control
<input type="checkbox"/>	Homo sapiens	HeLa; Epithelium; Cervix	BTAF1	Johannes F, et al. Bioinformatics 2010	●●●●●
<input type="checkbox"/>	Homo sapiens	HeLa; Epithelium; Cervix	GAPDH	Johannes F, et al. Bioinformatics 2010	●●●●●
<input type="checkbox"/>	Homo sapiens	K562; Erythroblast; Bone Marrow	EGR1	Tang C, et al. Electrophoresis 2010	●●●●●
<input type="checkbox"/>	Homo sapiens	LS174T; Epithelium; Colon	TCF4	Mokry M, et al. PLoS ONE 2010	●●●●●
<input type="checkbox"/>	Homo sapiens	LS174T; Epithelium; Colon	TCF4	Mokry M, et al. PLoS ONE 2010	●●●●●
<input type="checkbox"/>	Homo sapiens	LS174T; Epithelium; Colon	TCF4	Mokry M, et al. PLoS ONE 2010	●●●●●
<input type="checkbox"/>	Homo sapiens	LS174T; Epithelium; Colon	TCF4	Mokry M, et al. PLoS ONE 2010	●●●●●

Определения

Композиционный элемент – функциональная единица, в которой взаимодействия ТФ-ДНК и ТФ-ТФ вносят вклад в специфичный паттерн регуляции экспрессии генов (*Olga V. Kel-Margoulis et al. (2000) COMPEL: a database on composite regulatory elements providing combinatorial transcriptional regulation. Nucleic Acids Res.*).

Мотив – предпочитаемая последовательность нуклеотидов с которой связывается транскрипционный фактор.

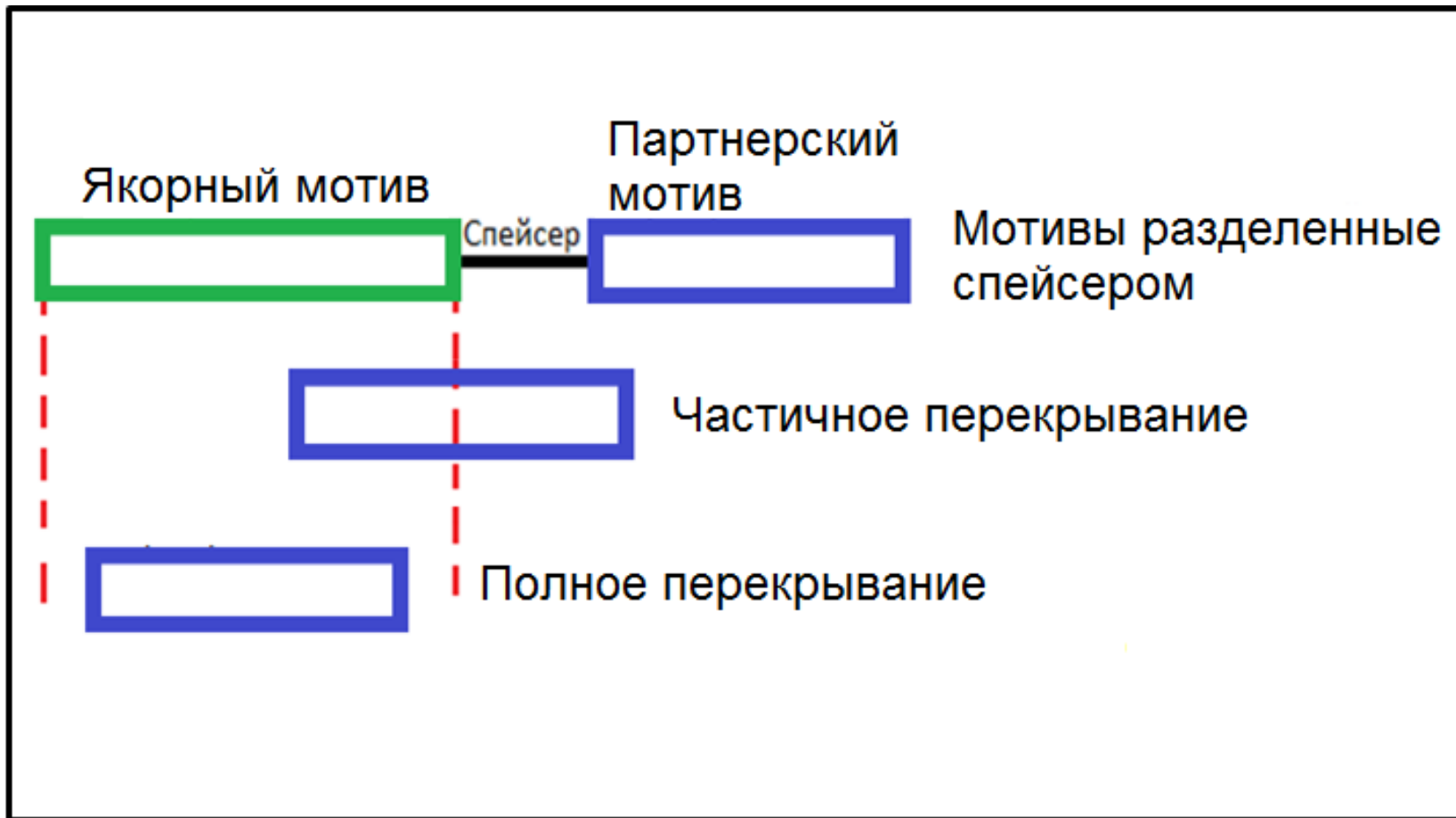
Два мотива могут:

- Перекрываться
- Разделяться спейсером

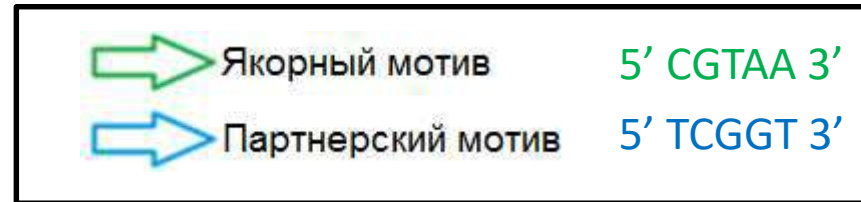
Якорный мотив – мотив для которого делали ChIP-Seq эксперимент.

Партнерский мотив – находящийся вблизи якорного мотива мотив партнерского ТФ.

Типы расположения мотивов относительно друг друга



Типы ориентации в паре мотивов



Мотивы в одной цепи

Мотивы в разных цепях

Якорь - партнер

Партнер - якорь

Инвертированный

Эвертированный

+ цепь



- цепь



Сравнение программ по поиску КЭ

Имя	Анализ одного набора пиков	Предсказание КЭ с перекрытием МОТИВОВ	URL	Ссылка
Spamo	Да	Нет	http://meme-suite.org/tools/spamo	Whittington et al.
iTFs	Да	Нет	http://veda.cs.uiuc.edu/iTFs/	Kazemian et al.
Gem	Нет	Да	http://groups.csail.mit.edu/cgs/gem/	Guo et al.
Taco	Нет	Да	http://bioputer.mimuw.edu.pl/taco/	Jankowski et al.
MCOT	Да	Да	https://gitlab.sysbio.cytogen.ru/academiq/mcot-kernel	Levitsky et al.

Whittington et al. (2011) Inferring transcription factor complexes from ChIP-seq data. Nucleic Acids Res.

Kazemian et al. (2013) Widespread evidence of cooperative DNA binding by transcription factors in Drosophila development. Nucleic Acids Res.

Guo et al. (2012) High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. PLoS Comput Biol.

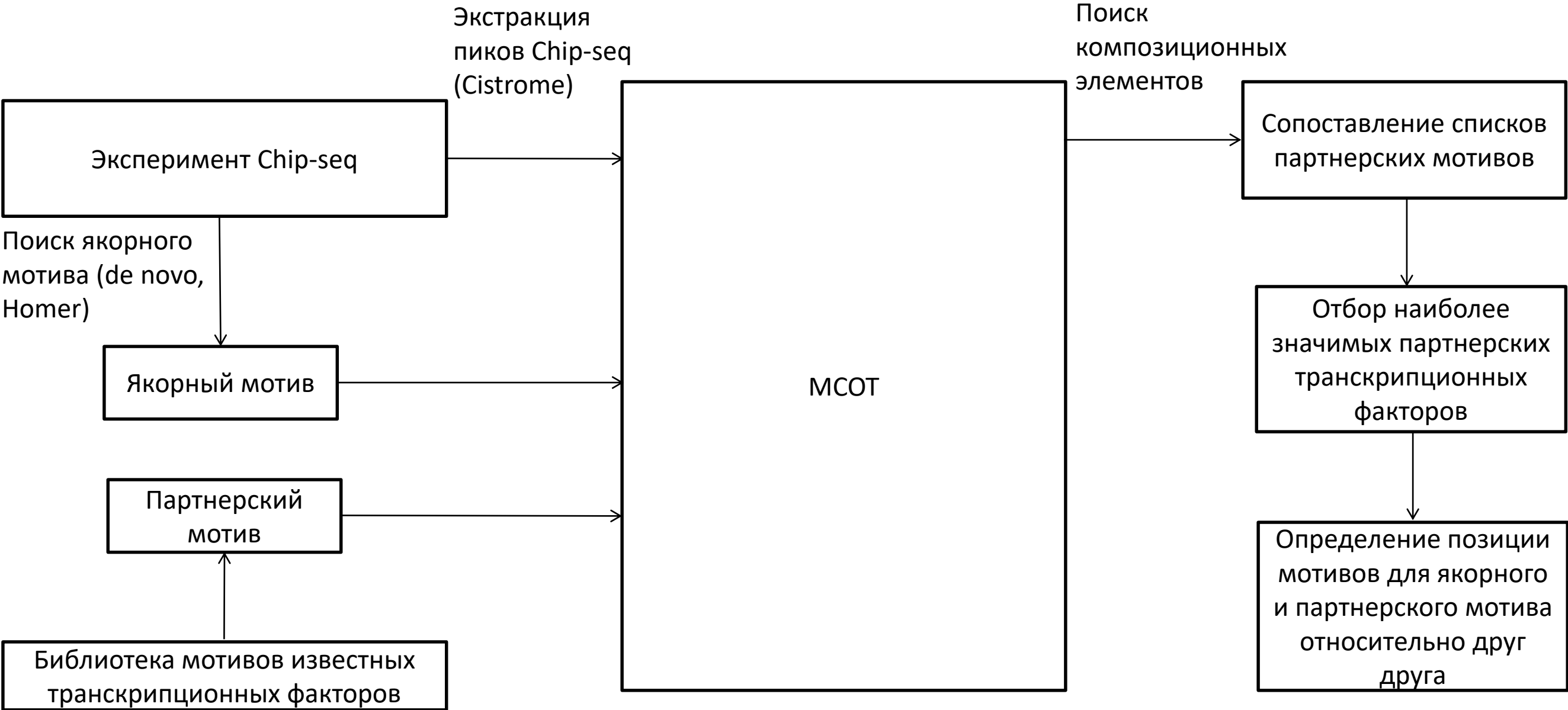
Jankowski et al. (2014) TACO: a general-purpose tool for predicting cell-type-specific transcription factor dimers. BMC Genomics.

Levitsky V et al. (2019) A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. Nucleic Acids Res.

Цель и задачи

- Цель: выявить тканеспецифичные партнерские мотивы для изучаемых якорных транскрипционных факторов.
- Задачи:
 1. Провести анализ литературы с целью выявления известных композиционных элементов.
 2. Подтвердить найденные в литературе известные композиционные элементы.
 3. Предсказать новые тканеспецифичные композиционные элементы для якорных мотивов FOXA1, JUN, ESR1, ZEB1.

Схема работы



Начальные данные

Название клеточной линии	FOXA1	JUN	ESR1	ZEB1
Ткань печени / карцинома печени	Liver		Liver	
Клетки молочной железы / аденокарцинома молочной железы	T47D			Mammary gland
Аденокарцинома эндометрия			ECC1	
Эндотелиальные клетки пупочной вены		HUVEC		
Аденокарцинома легкого		A549		
В-лимфоциты				B cells

Поиск якорного мотива de novo

- Для поиска якорного мотива de novo была использована программа HOMER

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-7835	-1.804e+04	71.70%	18.02%	40.1bp (63.4bp)	Foxa3(Forkhead)/Liver-Foxa3-ChIP-Seq(GSE77670)/Homer(0.977) More Information Similar Motifs Found	motif file (matrix)
2		1e-1515	-3.489e+03	20.33%	5.39%	48.8bp (63.9bp)	HNF4a(NR),DR1/HepG2-HNF4a-ChIP-Seq(GSE25021)/Homer(0.955) More Information Similar Motifs Found	motif file (matrix)
3		1e-1135	-2.614e+03	27.93%	11.51%	50.4bp (62.3bp)	CEBPA/MA0102.4/Jaspar(0.885) More Information Similar Motifs Found	motif file (matrix)
4		1e-652	-1.502e+03	4.15%	0.44%	49.6bp (66.5bp)	CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.910) More Information Similar Motifs Found	motif file (matrix)
5		1e-466	-1.075e+03	7.90%	2.39%	52.0bp (59.3bp)	Arid5a/MA0602.1/Jaspar(0.803) More Information Similar Motifs Found	motif file (matrix)
6		1e-457	-1.053e+03	5.34%	1.17%	51.1bp (61.7bp)	HNF1b(Homeobox)/PDAC-HNF1B-ChIP-Seq(GSE64557)/Homer(0.926) More Information Similar Motifs Found	motif file (matrix)
7		1e-422	-9.721e+02	28.09%	17.13%	53.5bp (63.3bp)	NF1-halfsite(CTF)/LNCaP-NF1-ChIP-Seq(Unpublished)/Homer(0.965) More Information Similar Motifs Found	motif file (matrix)

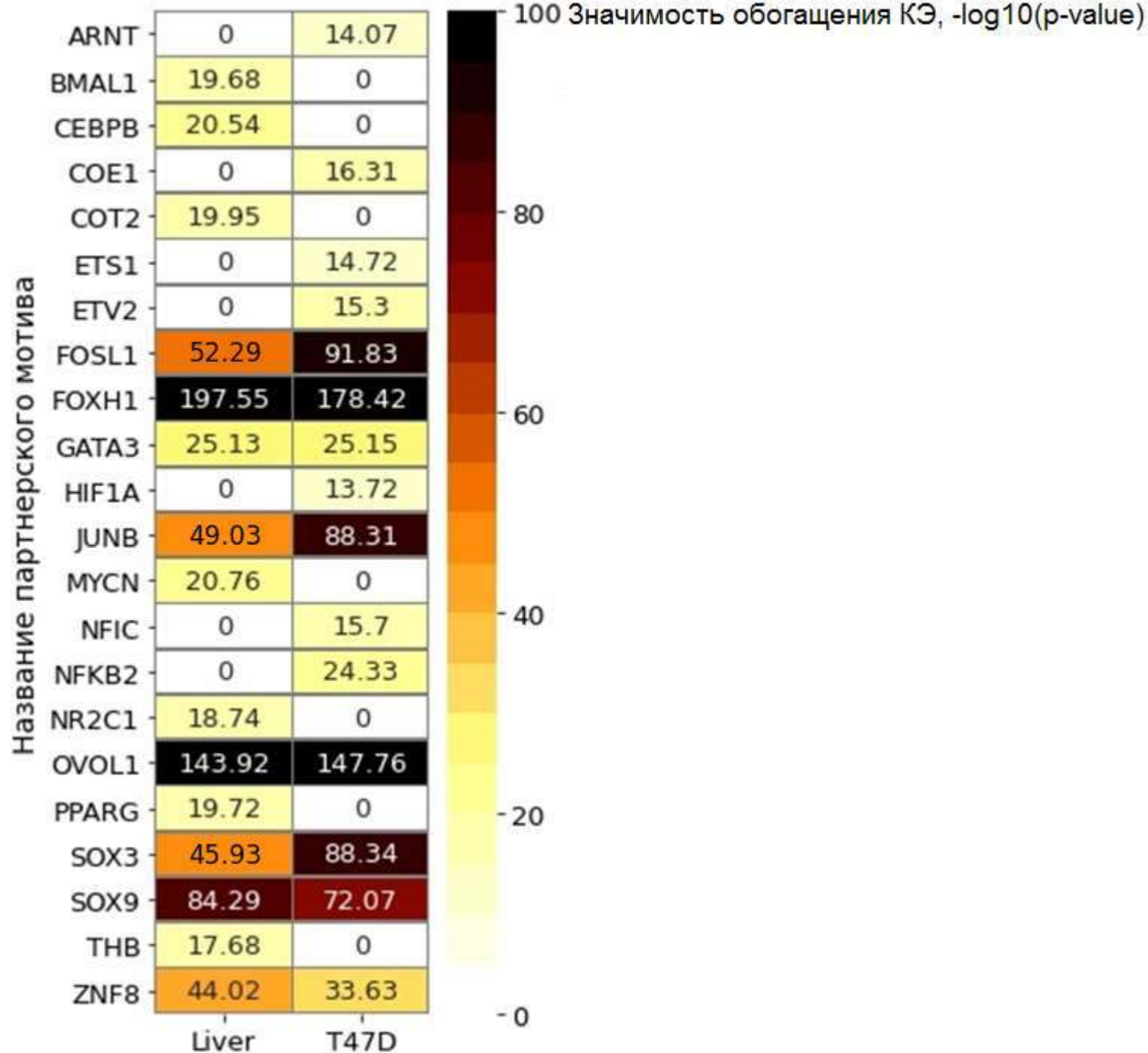
Неполный список найденных мотивов de novo



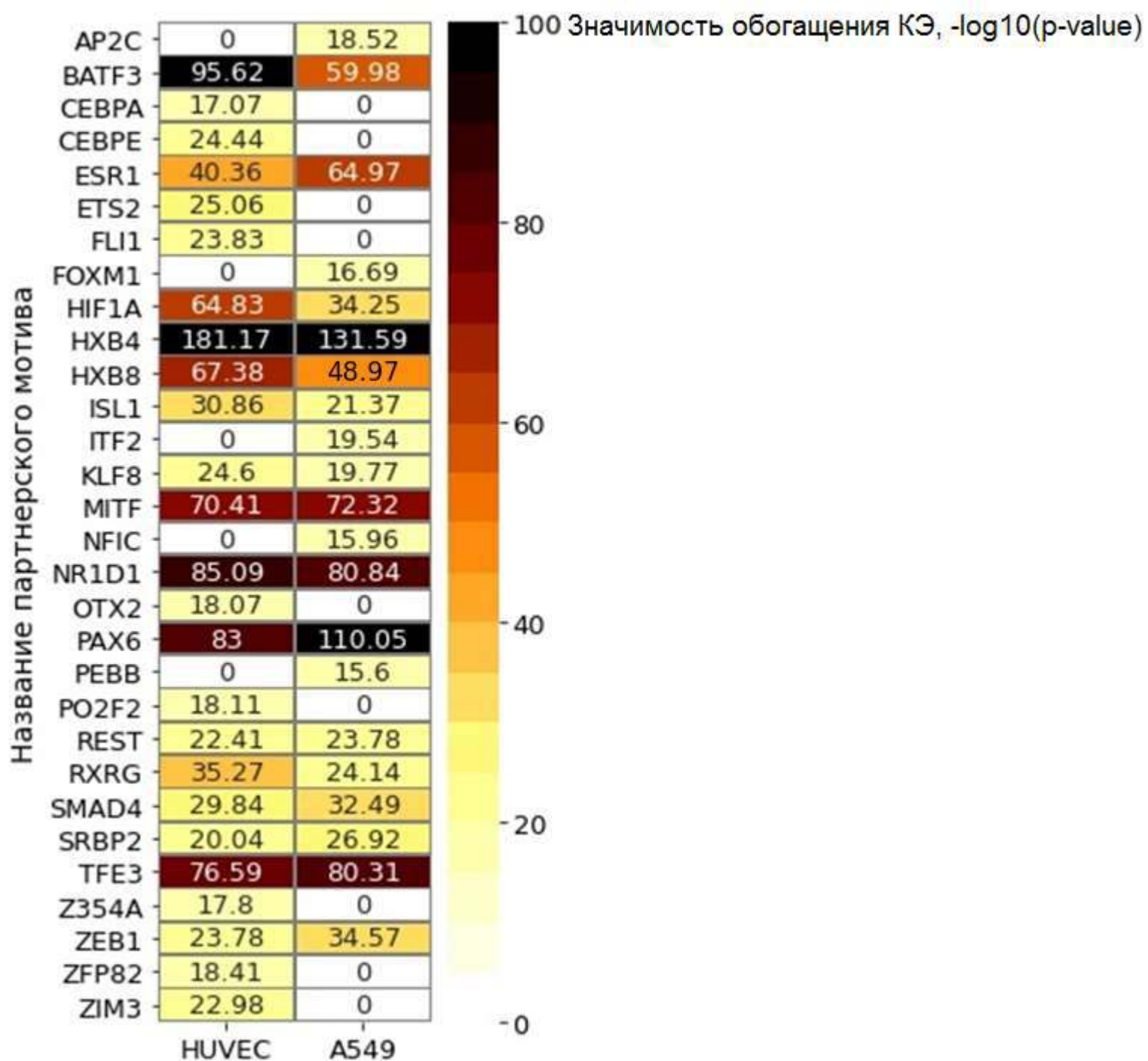
p-value:	1e-7835
log p-value:	-1.804e+04
Information Content per bp:	1.720
Number of Target Sequences with motif	18760.0
Percentage of Target Sequences with motif	71.70%
Number of Background Sequences with motif	4578.4
Percentage of Background Sequences with motif	18.02%
Average Position of motif in Targets	100.5 +/- 40.1bp
Average Position of motif in Background	101.2 +/- 63.4bp
Strand Bias (log2 ratio + to - strand density)	0.0
Multiplicity (# of sites on avg that occur together)	1.35
Motif File:	file (matrix) reverse opposite
SVG Files for Logos:	forward logo reverse opposite

Якорный мотив найденный HOMER для данных из [GSE96114](#), FOXA1

Тепловая карта значимых партнерских мотивов для якорного мотива FOXA1.



Тепловая карта
 значимых партнерских
 мотивов для якорного
 мотива JUN.



Количество различных типов КЭ для JUN/HIF1A в зависимости от взаимной ориентации и расположения ССТФ в клетках HUVEC

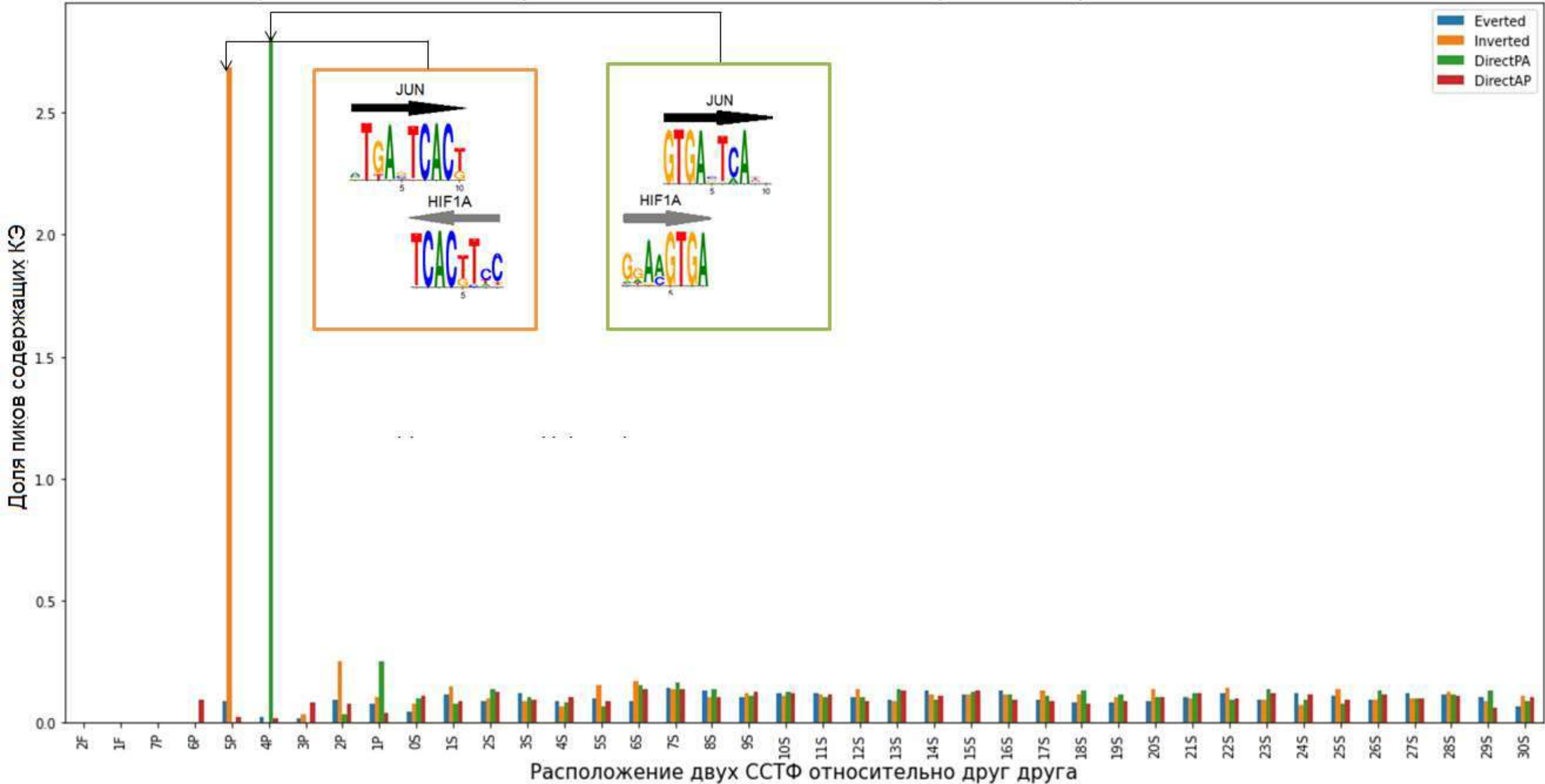
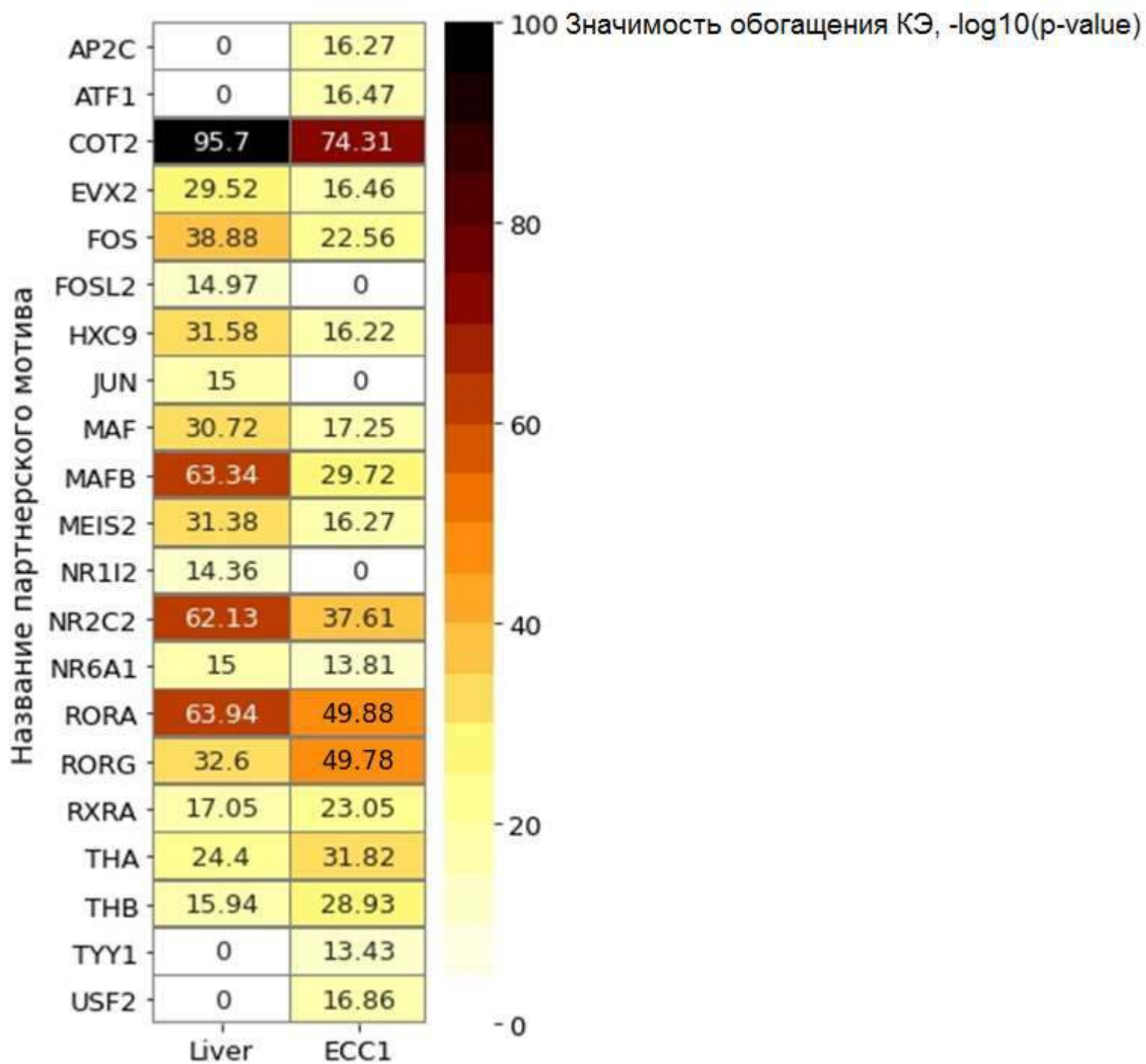


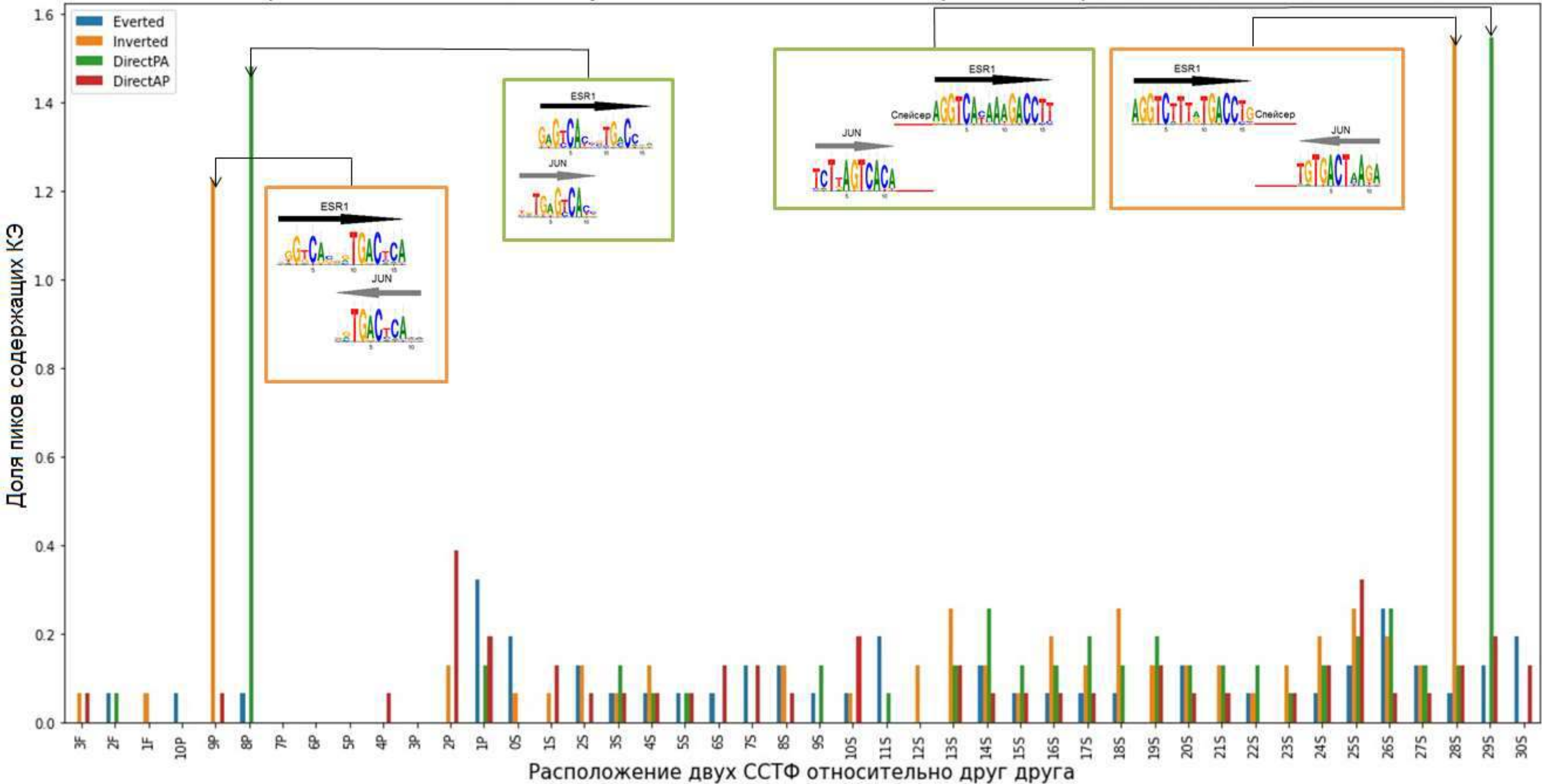
Таблица найденных партнерских мотивов для JUN, подтвержденных в литературе

Партнерский мотив для ТФ JUN	Статья
HIF1A	<i>Arántzazu Alfranca et al., c-Jun and Hypoxia-Inducible Factor 1 Functionally Cooperate in Hypoxia-Induced Gene Transcription // Molecular and Cellular Biology (2001)</i> <i>Robert Carlsson et al., STAT3 precedes HIF1α transcriptional responses to oxygen and oxygen and glucose deprivation in human brain pericytes // PLOS ONE (2018)</i>
BATF3	<i>Rachel C. Lynn et al., c-Jun overexpression in CAR T cells induces exhaustion resistance // Nature (2019)</i>
ESR1	<i>Joseph M. Collins et al., ESR1 ChIP-Seq Identifies Distinct Ligand-Free ESR1 Genomic Binding Sites in Human Hepatocytes and Liver Tissue // International Journal of Molecular Sciences (2021)</i>
ZEB1	<i>Nora Feldker et al. Genome-wide cooperation of EMT transcription factor ZEB1 with YAP and AP-1 in breast cancer // The EMBO Journal (2020)</i>

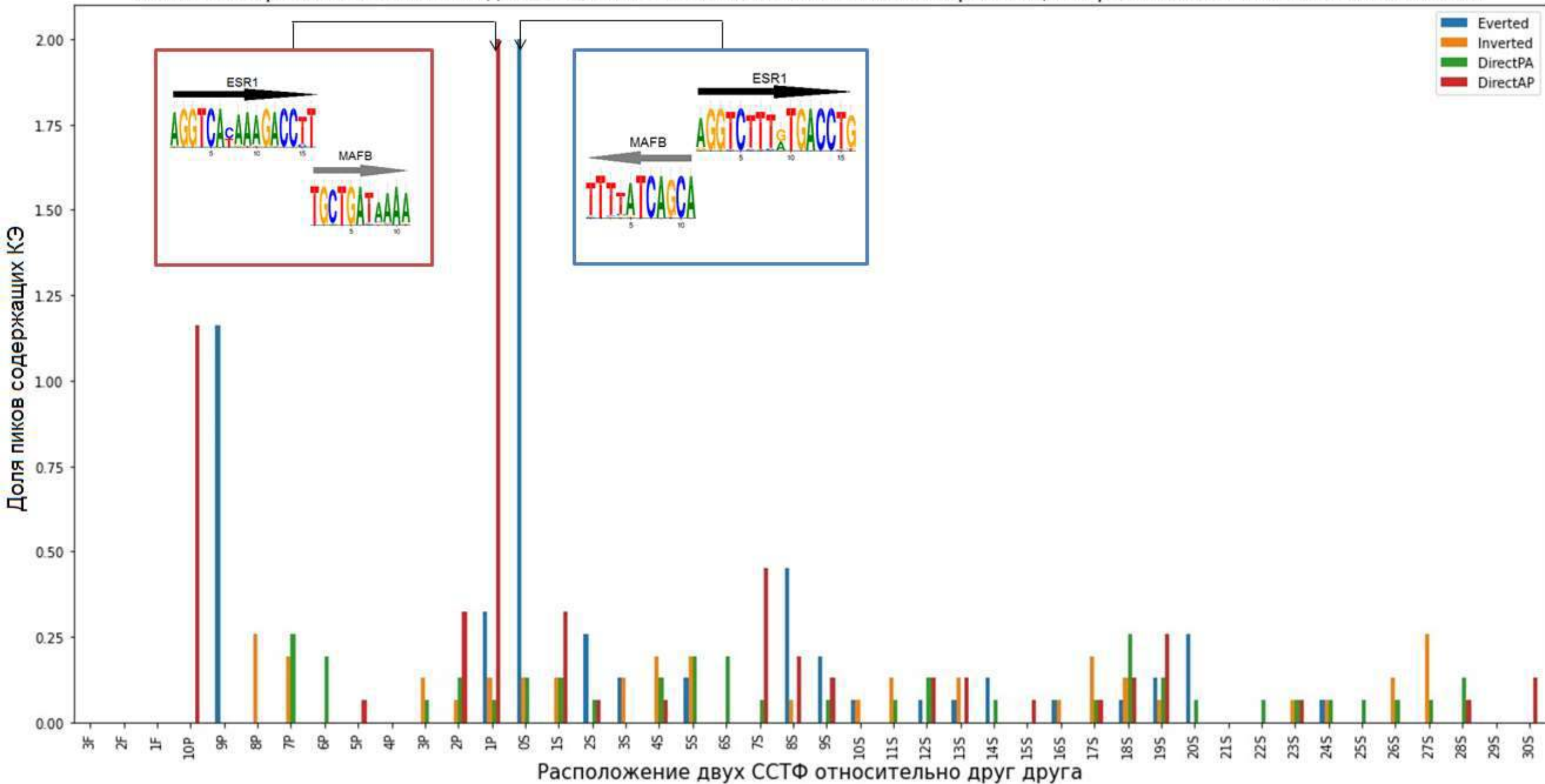
Тепловая карта значимых партнерских мотивов для якорного мотива ESR1.



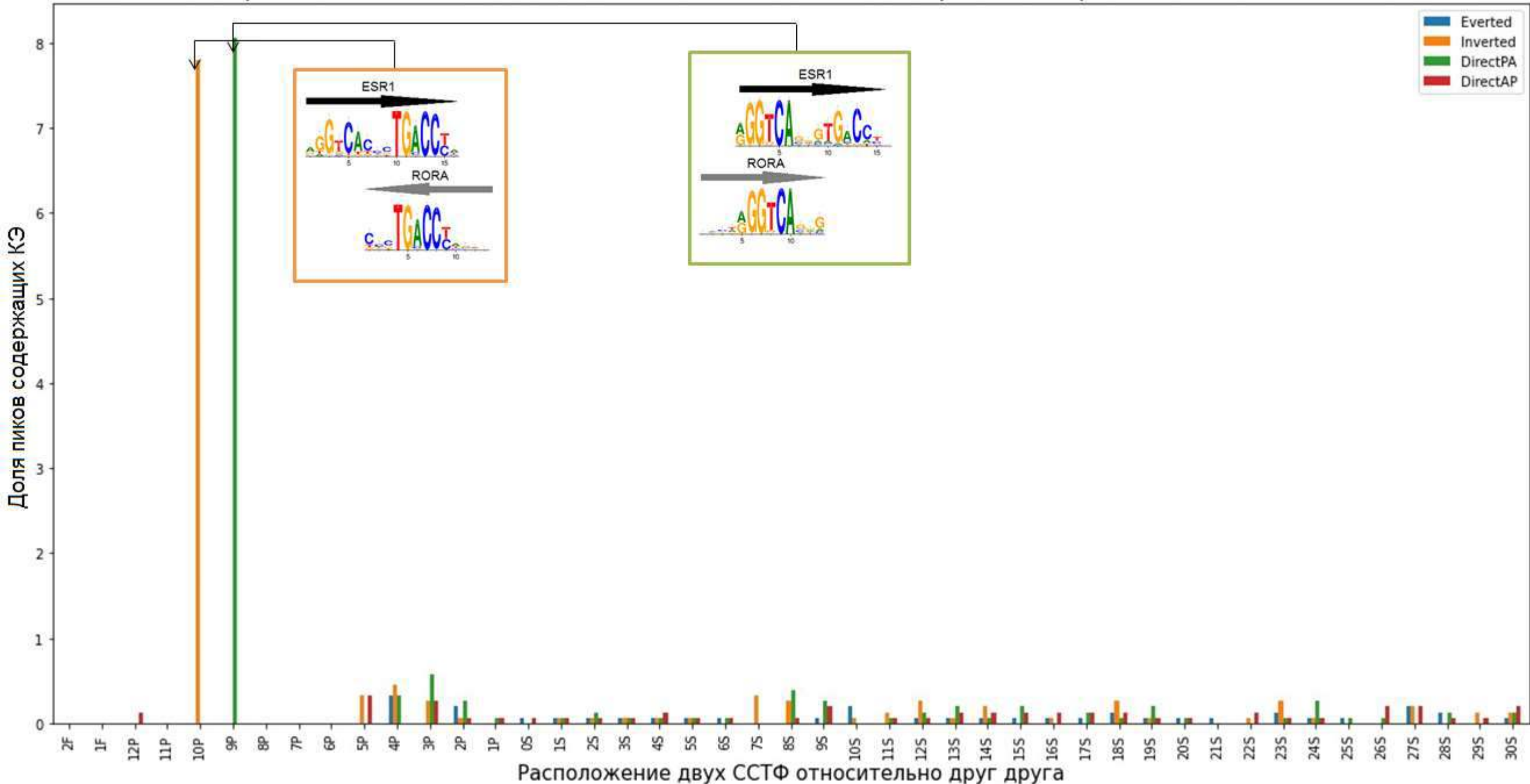
Количество различных типов КЭ для ESR1/JUN в зависимости от взаимной ориентации и расположения ССТФ в ткани печени



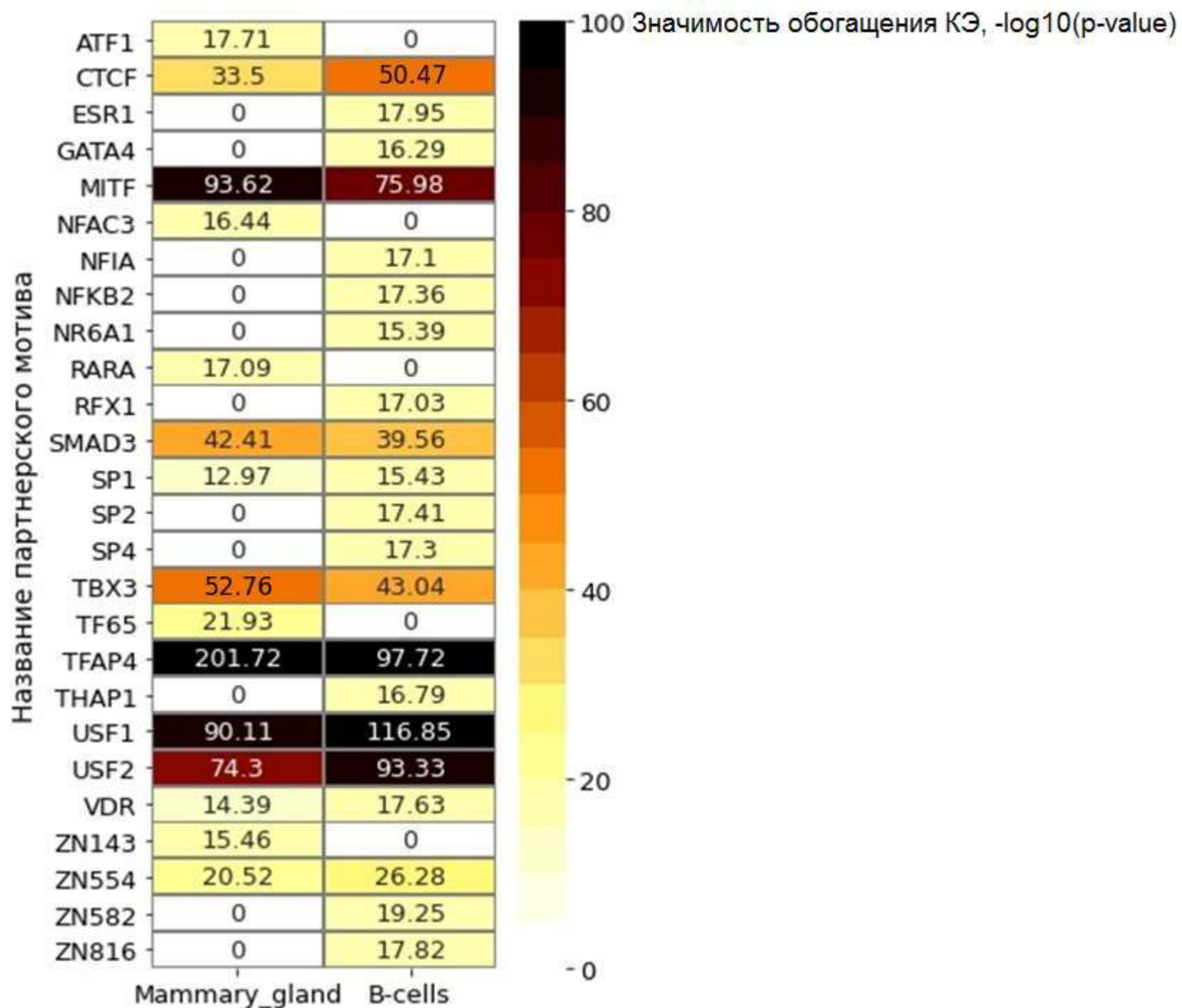
Количество различных типов КЭ для ESR1/MAFB в зависимости от взаимной ориентации и расположения ССТФ в ткани печени



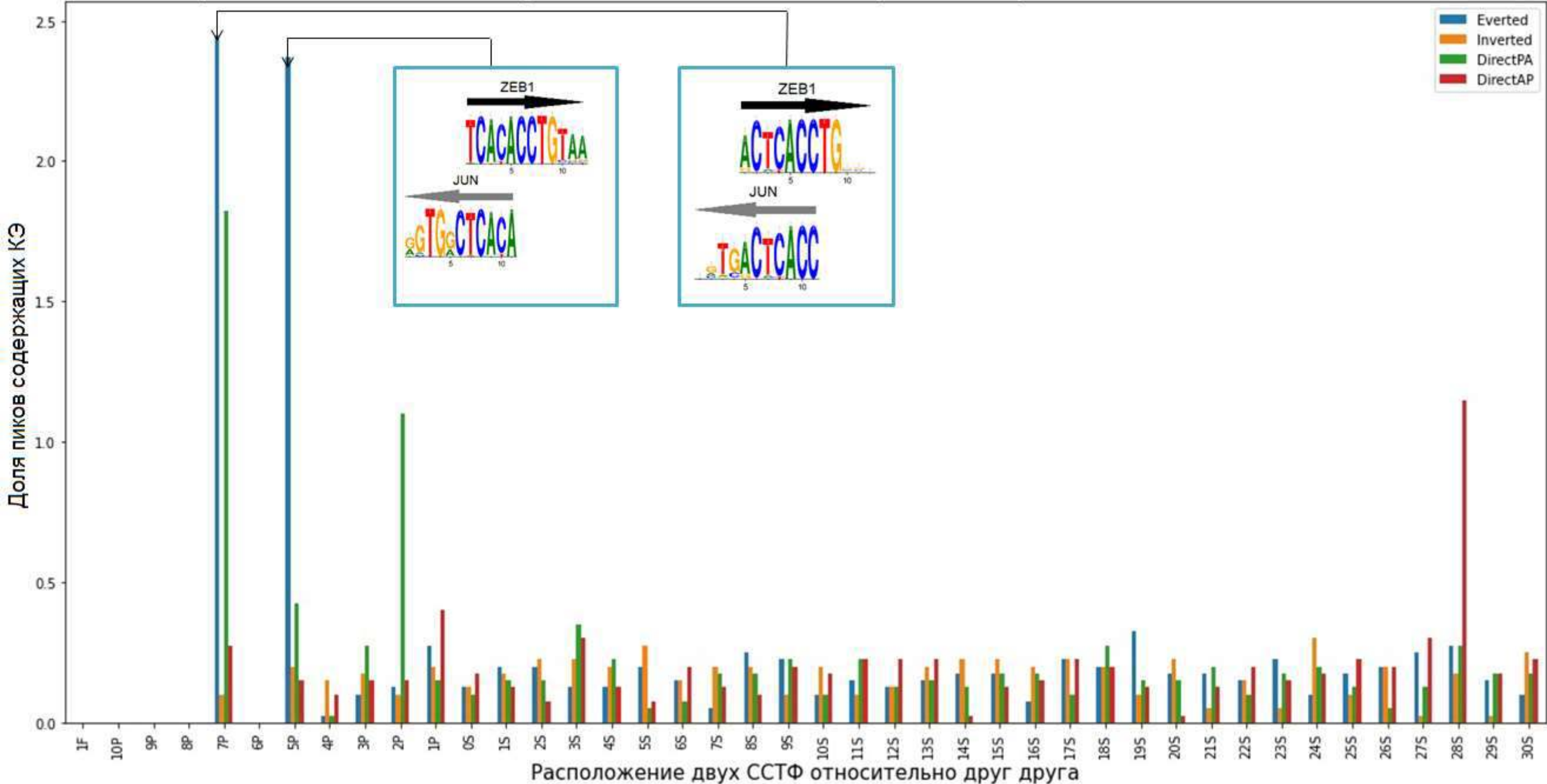
Количество различных типов КЭ для ESR1/RORA в зависимости от взаимной ориентации и расположения ССТФ в ткани печени



Тепловая карта
 значимых партнерских
 мотивов для якорного
 мотива ZEB1.

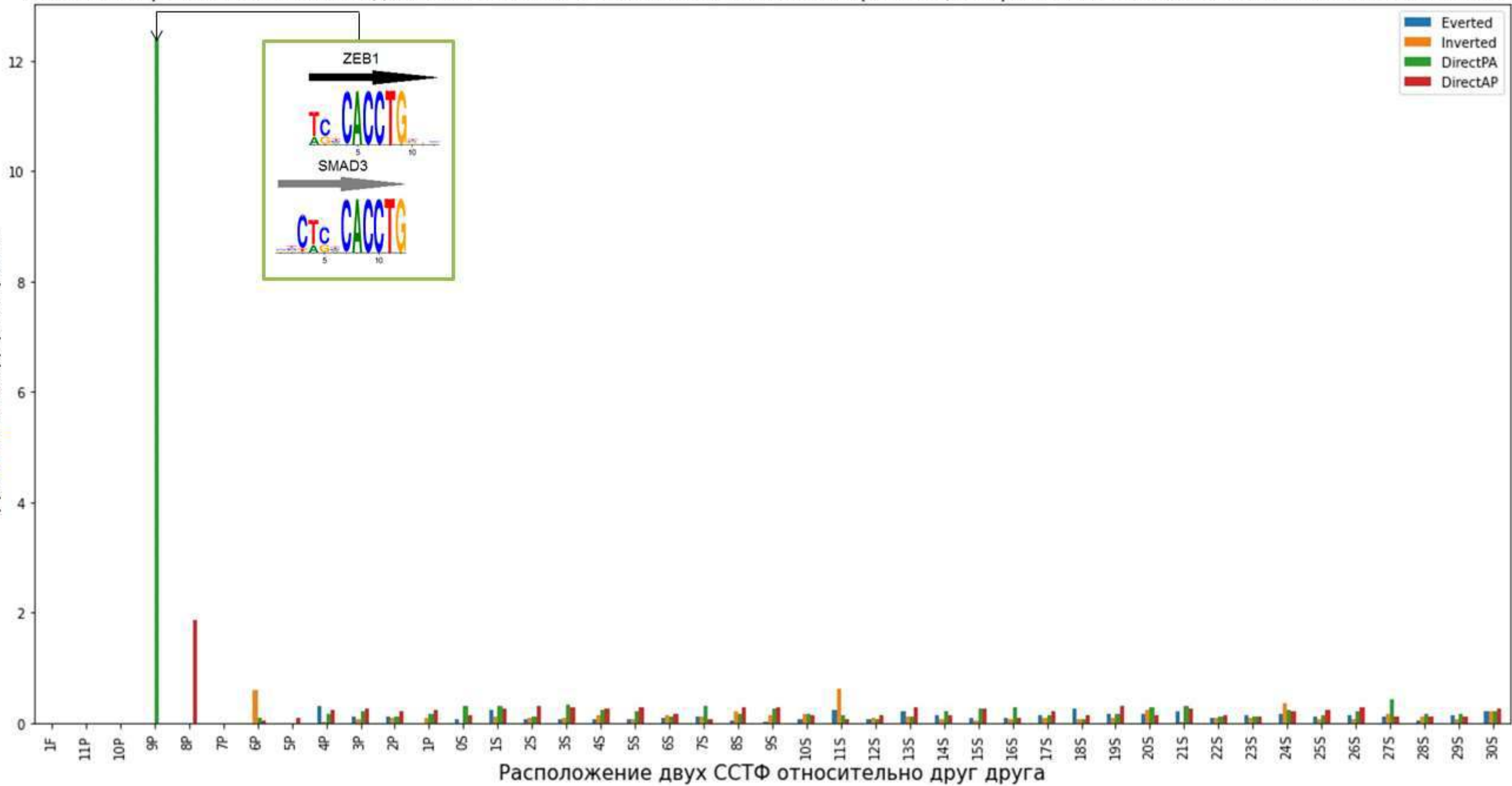


Количество различных типов КЭ для ZEB1/JUN в зависимости от взаимной ориентации и расположения ССТФ в клетках молочной железы



Количество различных типов КЭ для ZEB1/SMAD3 в зависимости от взаимной ориентации и расположения ССТФ в клетках молочной железы

Доля пиков содержащих КЭ



Выводы

1. Был проведен анализ по литературе с целью выявления тех ТФ, которые образуют КЭ с другими ТФ.
2. Подтверждены ранее описанные в литературе композиционные элементы: JUN/HIF1A, ESR1/JUN, ESR1/RXRA, ZEB1/JUN.
3. Были предсказаны новые потенциальные тканеспецифические композиционные элементы для четырех исследуемых якорных мотивов:
 - **FOXA1**: BMAL1, CEBPE, COT2, MYCN, NR2C1, PPARG, THB (ткань печени).
ARNT, COE1, ETS1, ETV2, HIF1A, NFIC, NFKB2, (клетки рака молочной железы).
 - **JUN**: CEBPA, CEBPE, ETS2, FLI1, OTX2, PO2F2, Z354A, ZFPB2, ZIM3 (клетки пупочной вены).
AP2C, FOXM1, ITF2, NFIC, PEBB (аденокарцинома легкого).
 - **ESR1**: FOSL2, JUN, NR1H2 (ткань печени).
ATF1, TYY1, USF2 (Аденокарцинома эндометрия).
 - **ZEB1**: ATF1, CTCF, NFAC3, RARA, TF65, ZN143 (клетки молочной железы).
ESR1, GATA4, NFIA, NFKB2, NR6A1, RFX1, SP1/2/4, THAP1, ZN582, ZN816 (В-лимфоциты).