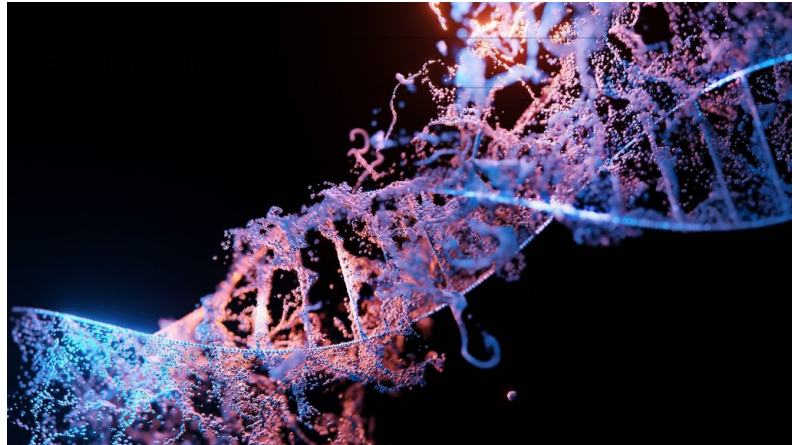


Групповой проект H3K4me. Результаты



Methylation of lysine 4 of histone H3

Зарипов Данис	PHF8	Histone modification erase
Васильева Дарья	KMT2A	Histone modification write
Сильвестрова Анна	MLL2	Histone modification write
Лапшина Ксения	AIRE	Histone modification read, TF
Степанянц Максим	RAG2	Histone modification read
Ярцев Кирилл	ING4	Histone modification read
Епифанов Евгений	ZCWPW1	Histone modification read
Митурич Василий	TAF3	Histone modification read
Сушкова Дарья	WDR5	Histone modification read
Белова Наталья	CCDC101	Histone modification read

Информация о белке

Ген	Кол-во изоформ RefSeq	ID белка	Длина белка	Где экспрессируется
PHF8	4	NP_001171825.1	1060	testis, ovary, uterus
KMT2A	3	NP_001399526.1	4002	brain, ovary, artery
MLL2 (KMT2B)	7	NP_055542.1	2715	testis, thyroid, ovary
AIRE	1	NP_000374.1	545	Brain, Lymph node, Thymus
RAG2	3	NP_001230715.1	527	thyroid, testis, bone marrow
ING4	9	NP_001121054.1	249	Brain cerebellar hemisphere, Brain cerebellum, Thyroid
ZCWPW1	3	NP_001372939.1	649	Testis(144.0)
TAF3	1	NP_114129.1	929	Artery- Brain - cerebellar hemisphere
WDR5	13	NP_001371343.1	334	testis, brain, cells, skin
CCDC101	1	NP_612423.1	293	testis, ovary

Как проводился BLASTp и отбор далёких гомологов

- Поиск BLASTp по протеомам был проведен через сервер
- Порог как в задании (if evalue < 1e-300, evalue = 1e-300, хоть это и не совсем порог)
- Уменьшалось ли количество хитов при удалении от человека по эволюционной лестнице? В целом да, но не всегда
- Дальний гомолог имел e-value < 1e-10
- Его длина была сопоставима с длиной исходного человеческого белка (разница не более 100 или 200 а.о.).

Информация о находках: КОЛ-ВО ХИТОВ

	hum	mous	zebraf	droso	c.eleg	ciliate	yeast	metha	therm	e.coli	tuberc
phf8	319	266	142	74	69	11	19	3	6	3	8
kmt2a	799	679	365	130	82	53	19	8	10	3	20
mlI2	255	549	301	122	138	55	34	7	7	5	2
aire	100	100	100	63	13	0	100	0	0	0	0
rag2	10	10	34	1	2	11	10	5	8	5	1
ing4	378	10	10	36	95	140	3	41	7	399	9
zcwpw1	196	215	67	26	8	9	8	3	7	4	1
taf3	127	104	77	156	76	56	27	8	9	5	4
wdr5	1677	1621	1559	921	495	901	278	8	12	10	8
ccdc101	11	3	2	39	2	0	1	0	0	0	0

Далёкие гомологи: общая информация

Ген	ID белка гомолога и название гена	Организм гомолога	Разность длин данного белка и его гомолога	E-value
PHF8	XP_001023333.2 (TTHERM_00444700) NP_010971.1 (JHD1) NP_500611.1 (LSD7)	ciliate yeast c.elegans	1060-1836 =776 1060-492 = 568 1060-897 = 163	4.10e-09 8.96e-49 1.75e-74
KMT2A	XP_001029767.2 (TTHERM_01309130) NP_011987.1 (SET1)	yeast ciliate	4002-2445 =1557 4002-1080 =2922	6,06E-37 6,15E-30
MLL2	NP_011987.1 XP_009290151.1 XP_001029767.2	yeast zebrafish ciliate	2715-1080= 1635 3771-2715=1056 2715-2445=270	3.87e-35 1.49e-166 6.34e-31
AIRE	CAI7203189.1 CQS_1a_G0010160.mRNA.1.CD S.1 AGJ94565.1 RE59080p1 NP_001103484.1 AIRE	yeast drosophila zebrafish	545-684 =138 545-1561 =1016 545-511 =34	8E-12 2E-13 5E-83
RAG2	NP_571460.2 RAG2	zebrafish	527-530 =3	0.0

Далёкие гомологи: общая информация

Ген	ID белка гомолога и название гена	Организм гомолога	Разность длин данного белка и его гомолога	E-value
ING4	NP_496909.1(IGP) XP_012652826.1(PHD) NP_014302.3(Pho23p)	c.elegans ciliate yeast	490-249 = 241 361-249 = 112 330-249=81	1.12e-16 1.55e-29 1.42e-21
ZCWPW1	NP_001076020.1 (NSD2) NP_001346952.1 (zinc finger CW-type PWWP domain protein 1)	zebrafish MOUSE	649-1461 = 812 649-630 = 19	2.52e-11 0.0
TAF3	NP_001257272.1 (PHD-type domain-containing protein) NP_001257272.1(TBP-associated factor 3) NP_001036209.1 (transcription initiation factor TFIID subunit 3)	c. elegans drosophila zebrafish	929-1007 =78 929-1406 =477 929-898 =31	1.38e-17 7.97e-24 1.31e-75
WDR5	NP_009734.1 (SWD3)	yeast	334-315 =19	1.22e-46
CCDC101	NP_500635.3 XP_017960484.1	c. elegans drosophila	293 - 236 = 57 293 - 289 = 4	3e-48 4e-93

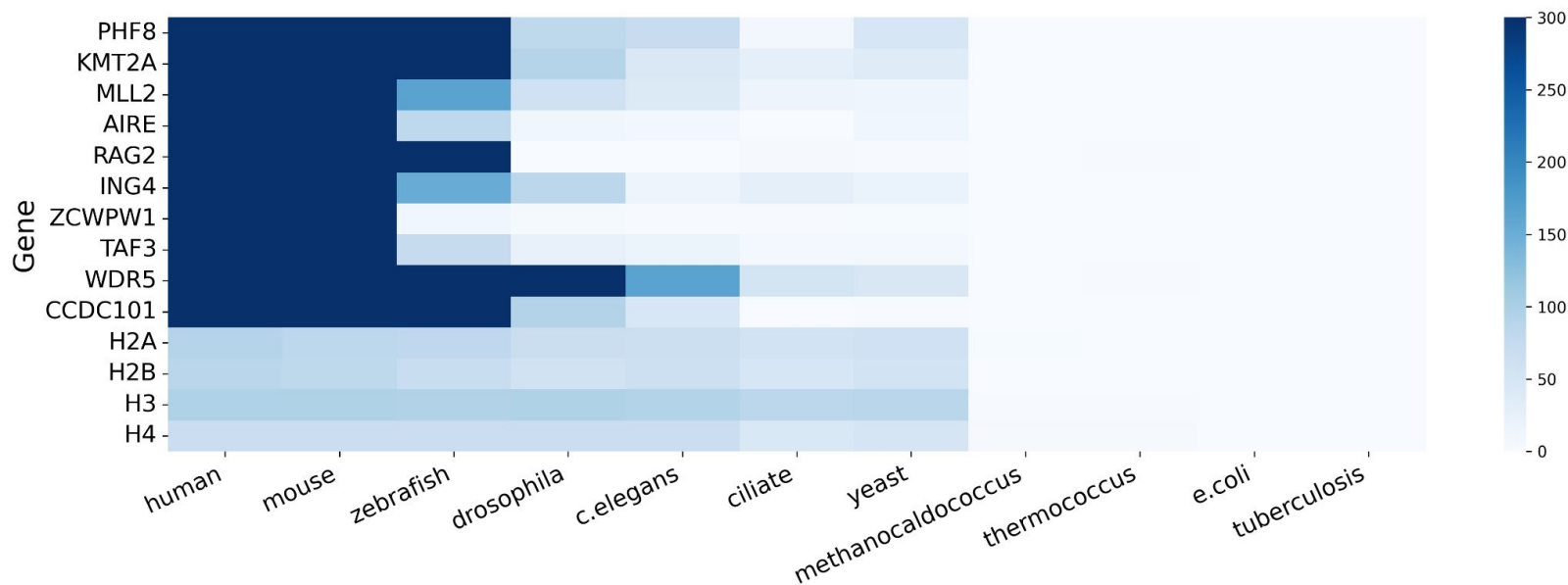
Далёкие гомологи: домены и функция 1

Ген	Домены гена	Домены гомолога	Функция гомолога
PHF8	JHD, PHD_PHF8, cupin_RmlC, JmjC	PHD_TCF19_like, cupin_RmlC-like super family, JHD super family	The function of Lysine-specific demethylase 7 homolog in <i>Caenorhabditis elegans</i> (<i>C. elegans</i>) is to remove specific methyl groups from lysine residues in histone proteins.
kmt2a	SET_KMT2A_2B, ePHD_KMT2A, Bromo_ALL-1, PHD3_KMT2A, PHD1_KMT2A, FYRC, PHD2_KMT2A, zf-CXXC, FYRN	SET_SETD1-like, PHD_SF super family, Bromodomain, PHD, PHD_SF super family, LIM	Molecular function: dna binding, chromatin binding, dna-binding transcription factor activity, rna binding, mrna binding, structural constituent of ribosome, gtpase activity Biological process: ribosomal subunit export from nucleus, mitotic cell cycle, conjugation with cellular fusion, cell morphogenesis, cytokinesis, invasive growth in response to glucose limitation
mlI2	Extended PHD finger, SET_KMT2A_2B, ePHD_KMT2B, Bromo_ALL-1, PHD3(2,1)_KMT2B, FYRC, FYRN, zf-HC5HC2H	SET domain protein histone methyltransferase SET1 Bromodomain BROMO PHD ePHD2_KMT2C_like	Enables RNA binding activity and protein-lysine N-methyltransferase activity. Contributes to histone methyltransferase activity (H3-K4 specific). Involved in several processes, including nucleic acid metabolic process; peptidyl-lysine methylation; and regulation of gene expression.
aire	HSR super family, PHD1_AIRE, PHD2_AIRE, SAND super family	HSR, PHD1_AIRE, PHD2_AIRE, SAND super family	Molecular Function: chromatin binding, histone binding, metal ion binding, RNA polymerase II transcription regulatory region sequence-specific DNA binding, translation regulator activity. Biological Process: central tolerance induction to self antigen, humoral immune response, negative thymic T cell selection, peripheral T cell tolerance induction, regulation of thymocyte migration, thymus epithelium morphogenesis. [UniProt]
rag2	RAG2 super family, PHD_SF super family	RAG2 super family, PHD_SF super family	This gene encodes a protein that is involved in the initiation of V(D)J recombination during B and T cell development. This protein forms a complex with the product of the adjacent recombination activating gene 1, and this complex can form double-strand breaks by cleaving DNA at conserved recombination signal sequences.

Далёкие гомологи: домены и функция 2

Ген	Домены гена	Домены гомолога	Функция гомолога
ing4	ING_ING4, TNG2 super family	TNG2	Pho23p is a component of the PHO signaling pathway, which regulates the expression of genes involved in phosphate metabolism and uptake. When cells are starved for phosphate, Pho23p is translocated from the cytoplasm to the nucleus where it interacts with other proteins to activate expression of PHO-regulated genes
zcwpw1	PWWP super family zf-CW super family	PWWP_ZCWPW1 zf-CW SCP-1 super family	The function of the zinc finger CW-type PWWP domain protein 1 (ZCWPW1) is not extensively studied, and its precise role in cellular processes is still being investigated.
taf3	PHD_TAF3 BTP	PHD_TAF3 BTP	Ортолог человеческого гена. Скорее всего, та же функция a general initiation factor (GTF) that plays a key role in preinitiation complex (PIC) assembly through core promoter recognition. The interaction of H3K4me3 with TAF3 directs global TFIID recruitment to active genes, which regulates gene-selective functions of p53 in response to genotoxic stress.
wdr5	WD40 super family, CaiC super family, PHA03247 super family	WD40 super family	Contributes to histone methyltransferase activity (H3-K4 specific). Involved in histone H3-K4 methylation; regulation of meiotic DNA double-strand break formation; and telomere organization. Part of Set1C/COMPASS complex. Orthologous to several human genes including WDR5 (WD repeat domain 5).
ccdc101	DUF1325 super family	DUF1325 super family	NP_500635.3 enables methylated histone binding.

Визуализация результатов: белки



Хитмэп по результатам BLASTp: в ячейках $-\log_{10}(\text{evalue})$

Обобщение об эволюции

- В большинстве белков первый гомолог появляется в дрожжах (одноклеточный эукариот)
- Первый “уверенный” гомолог для всех белков – рыба данио-рерио (многоклеточный позвоночный)
- Белки, новейшие гомологи которых связаны с H3K4me, появились не вполне одновременно
- У дрожжей есть как минимум по 1 белку, родственному с белками, отвечающими за erase, write и read H3K4me
- Можно считать, что H3K4me уверенно находится в *c.elegans*

Спасибо за внимание!

