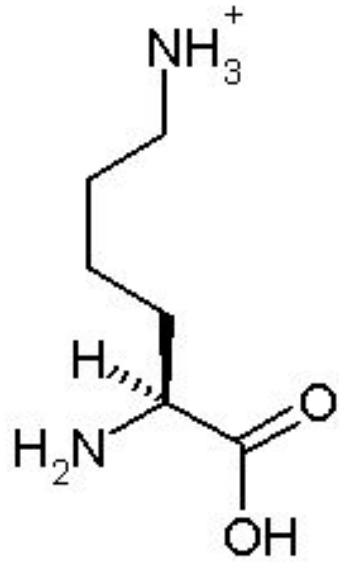
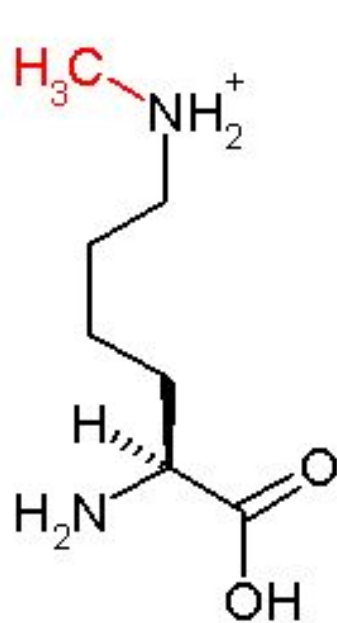


# Групповой проект H3K4me

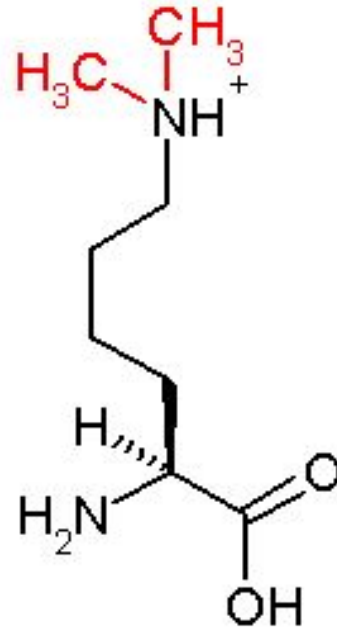
## Methylation of lysine 4 of histone H3



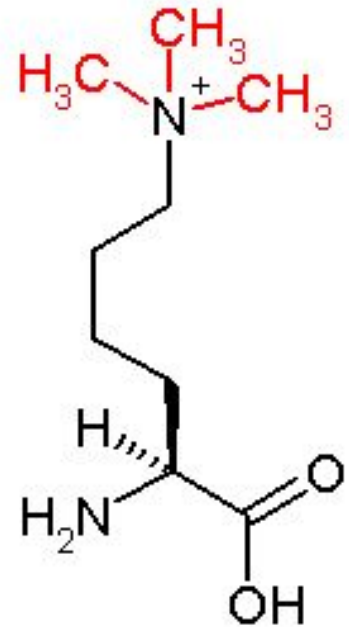
Lysine



Mono-methyl  
lysine



Dimethyl lysine



Trimethyl lysine

# Эпигенетическая модификация - Histone methylation

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

AIRE	<p>"... AIRE selectively interacts with histone H3 through its first plant homeodomain (PHD) finger (AIRE-PHD1) and preferentially binds to non-methylated H3K4 (H3K4me0) ... AIRE-PHD1 is an important member of a newly identified class of PHD fingers that specifically recognize H3K4me0 ..." Org et al., 2008</p> <p>"Certain posttranslational modifications of H3 tails, notably dimethylation or trimethylation at H3K4, abrogated binding by Aire, whereas others were tolerated." Koh et al., 2008</p>
KMT2A	<p>"Histone lysine methyltransferase 2 (KMT2) proteins form multimeric enzymatic complexes that methylate lysine 4 on histone H3 (H3K4) at transcription regulatory elements in the genome" - doi: 10.1016/j.bbagr.2020.194545</p> <p>"KMT2A (MLL1) is responsible for generating mono-, di-, and tri-methylated H3K4 through its SET domain and by interaction with cofactors"-doi:10.2217/epi.15.1</p> <p>"Of the many PHD fingers only PHD3 of KMT2A in KMT2 family exhibit H3K4 binding properties" - doi: 10.3390/cells7030017</p>
RAG2	<p>RAG2 (Recombination activating gene) contains a plant homeodomain (PHD) near its C terminus (RAG2-PHD) that <b>recognizes histone H3 methylated at lysine 4 (H3K4me)</b> and influences V(D)J recombination. Two aspects of RAG2-PHD are unique. First, in the absence of the modified peptide, a peptide N-terminal to RAG2-PHD occupies the substrate-binding site, which may reflect an autoregulatory mechanism. Second, in contrast to other H3K4me3-binding PHD domains, RAG2-PHD substitutes a carboxylate that interacts with arginine 2 (R2) with a Tyr, resulting in binding to H3K4me3 that is enhanced rather than inhibited by dimethylation of R2. <a href="#">source</a></p>
PHF8 (1, 2)	<p>1) PHD finger protein 8 (<b>PHF8</b>) belongs to the JmJc domain-containing family of proteins that can <b>remove methyl groups</b> from arginine or lysine residues (...) the identified <b>H3K4me3</b> readers Sgf29, TRRAP, <b>PHF8</b> (...)</p> <p>2) PHF8 is associated with <b>hypomethylated</b> rRNA genes</p>
ING4	<p>Full-length ING4 recognizes methylated H3K4 peptides, but an ING4 derivative lacking the PHD finger, ING4<sub>ΔPHD</sub> (aa 1–194), does not, demonstrating that the PHD finger of ING4 is necessary and sufficient for the H3K4me peptide binding activity of ING4.</p>
ZCWPW1	<p><a href="#">Here</a>, we demonstrate that Zcwpw1 is an H3K4me3 reader that is required for DSB repair and synapsis in mouse testes.</p>
MLL2	<p>Here, we <a href="#">show</a> that Mll2, one of the six Set1/Trithorax-type H3K4 methyltransferases in mammals, is required for trimethylation of bivalent promoters in mouse embryonic stem cells.</p>
TAF3	<p>TAF3 contributes to promoter recognition and selectivity and acts as antiapoptotic factor.</p> <p><i>Further analyses show that H3K4me3 enhances p53-dependent transcription by stimulating preinitiation complex (PIC) formation; (...) H3K4me3-TAF3/TFIID interactions regulate geneselective functions of p53 in response to genotoxic stress.</i></p>
WDR5	<p>"Here we show that ... the WD40-repeat protein WDR5 directly associates with histone H3 di- and trimethylated at K4 and with H3-K4-dimethylated nucleosomes. <b>WDR5 is required for binding of the methyltransferase complex to the K4-dimethylated H3 tail as well as for global H3 K4 trimethylation</b> ..." - doi: 10.1016/j.cell.2005.03.036</p>
CCDC101	<p>"SGF29 (CCDC101) specifically recognizes and binds methylated 'Lys-4' of histone H3 (H3K4me), with a preference for trimethylated form (H3K4me3)... In the SAGA-type complexes, SGF29 is required to recruit complexes to H3K4me..."</p>

## В какие комплексы входят выбранные белки?

AIRE	Not found
KMT2A	MLL-HCF, CHD8, COMPASS-like MLL1,2
RAG2	not found
PHF8	PHF8-TOPBP1
ING4	HBO1
ZCWPW1	Not found
MLL2(KMT2D)	Set1/COMPASS,SWI/SNF (BAF),NuA4 (Tip60-p400) ,MLL2
TAF3	TFIID
WDR5	ATAC, NSL, RING2-L3MBTL2, MLL1/2, MLL3/4
CCDC101	ATAC



## Экспрессия (ткани)

Ген	Название ткани + медианный TPM
<a href="#">AIRE</a>	<b>GTEx</b> Brain: Hypothalamus (4.017), Anterior cingulate cortex (1.691), Cortex (1.383), Frontal Cortex (1.305), Nucleus accumbens (basal ganglia) (1.152) <b>NCBI</b> Lymph node (0.592); Thymus (0.443); Adrenal (0.069), Lymph node (0.056), Brain (0.044)
KMT2A	<b>GTEx:</b> Brain - Cerebellum 34.84, Brain - Cerebellar Hemisphere 32.50, Ovary 26.88, Artery - Tibial 25.50, Uterus 24.74, Nerve - Tibial 21.77, Artery - Aorta 21.16,
RAG2	В организме человека почти только в щитовидной железе (RPKM 16.6), но также понемногу в семенниках и костном мозге [ncbi]. По GTEx, есть на минимальном уровне везде, но в ощутимых размерах в щитовидной железе и немного в семенниках.
PHF8	Testis (59.4), Ovary (30.1), Uterus (24.3)
ING4	Brain cerebellar hemisphere 79.6, Brain cerebellum 76.35, Thyroid 69.97, Endocervix 67.59, Ectocervix 60.35, Uterus 60.29

## Экспрессия (ткани)

ZCWPW1	Testis(144.0)
MLL2	adrenal gland (2.783),appendix (4.121 ),bone marrow (9.459),cerebral cortex (3.186 ),colon (4.104),duodenum (2.446 )endometrium (4.128 ),esophagus (2.945 ) ,adipose tissue(3.081) ,gall bladder (3.217),heart (2.105) ,kidney (3.749),liver (1.068) ,lung (3.531) ,lymph node (3.521),ovary (5.62) ,pancreas (1.316) ,placenta (4.067) ,prostate (3.391) ,salivary gland (1.87) ,skin (5.759) ,small intestine (2.624),spleen (5.17) ,tomach (3.076),testis (5.013),thyroid (4.12)
TAF3	Артерия (16.29) - Мозг (14.95) - полушарие мозжечка (13.74)
WDR5	Testis (95.45), Brain - Cerebellar Hemisphere (62.63), Brain - Cerebellum (59.44), Cells - Cultured fibroblasts (58.67), Cells - EBV-transformed lymphocytes (51.34) (Данные взяты из БД GTEx)
CCDC101	Ubiquitous expression in testis (13.2) and ovary (12.9)



## Доменная структура

AIRE

KMT2A

RAG2

PHF8

ING4

ZCWPW1

MLL2

TAF3

WDR5

CCDC101



# Доменная структура AIRE

## Graphical summary

☐ Zoom to residue level

[show extra options »](#)

RF +1

Superfamilies

PHA03247

RF +2

Superfamilies

PHA03

histone H3 binding site

histone H3 binding site

RF +3

Superfamilies

S

PRK1232



## List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	PHA03247 super family	c133720	large tegument protein UL36; Provisional	5917-7209	8.26e-03
[+]	PHD_SF super family	c122851	PHD finger superfamily; The PHD finger superfamily includes a canonical plant homeodomain (PHD) ...	5270-5374	3.13e-13
[+]	HSR super family	c103917	HSR domain; The Sp100 protein is a constituent of nuclear domains, also known as nuclear dots ...	215-301	6.02e-06
[+]	PHA03247 super family	c133720	large tegument protein UL36; Provisional	50-556	2.16e-04
[+]	PHD_SF super family	c122851	PHD finger superfamily; The PHD finger superfamily includes a canonical plant homeodomain (PHD) ...	7970-8065	5.37e-04
[+]	HSR super family	c103917	HSR domain; The Sp100 protein is a constituent of nuclear domains, also known as nuclear dots ...	735-896	2.05e-12
[+]	SAND super family	c102536	SAND domain; The DNA binding activity of two proteins has been mapped to the SAND domain. The ...	3819-3953	2.40e-04
[+]	PRK12323 super family	c136108	DNA polymerase III subunit gamma/tau;	5286-5930	2.64e-04

# Доменная структура KMT2A

Conserved domains on [ref|NP\_001184033.1]

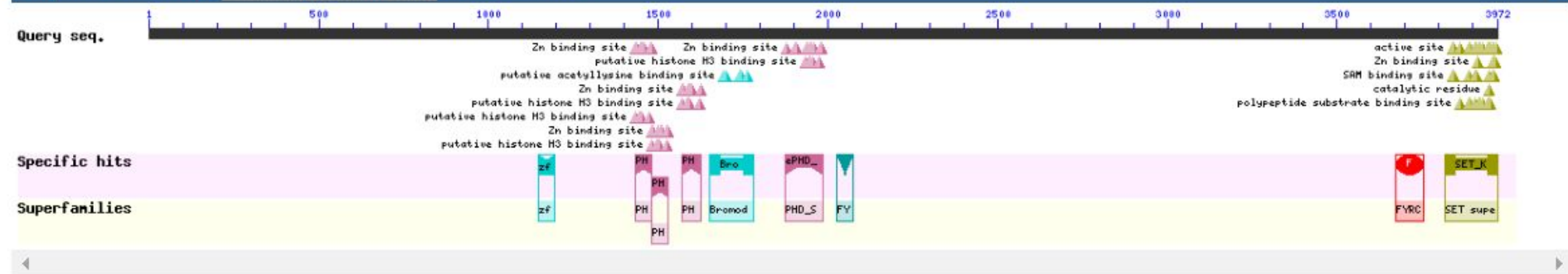
View **Concise Results** ?

:1-3972 histone-lysine N-methyltransferase 2A isoform 1 [Homo sapiens]

## Graphical summary

☐ Zoom to residue level

[show extra options >](#)



[Search for similar domain architectures](#) ?

[Refine search](#) ?


## List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	SET_KMT2A_2B	cd19170	SET domain (including post-SET domain) found in histone-lysine N-methyltransferase 2A (KMT2A), ...	3819-3972	2.03e-115
[+]	ePHD_KMT2A	cd15693	Extended PHD finger found in histone-lysine N-methyltransferase 2A (KMT2A); The extended plant ...	1874-1986	3.34e-83
[+]	Bromo_ALL-1	cd05493	Bromodomain, ALL-1 like proteins. ALL-1 is a vertebrate homologue of Drosophila trithorax and ...	1650-1780	6.84e-52
[+]	PHD3_KMT2A	cd15592	PHD finger 3 found in histone-lysine N-methyltransferase 2A (KMT2A); KMT2A (also termed ALL-1, ...	1568-1627	5.59e-37
[+]	PHD1_KMT2A	cd15588	PHD finger 1 found in histone-lysine N-methyltransferase 2A (KMT2A); KMT2A (also termed ALL-1, ...	1433-1479	4.36e-33
[+]	FYRC	smart00542	FY-rich domain, C-terminal region; is sometimes closely juxtaposed with the N-terminal region ...	3672-3755	2.00e-32
[+]	PHD2_KMT2A	cd15590	PHD finger 2 found in histone-lysine N-methyltransferase 2A (KMT2A); KMT2A (also termed ALL-1, ...	1481-1530	3.46e-32
[+]	zf-CXXC	pfam02008	CXXC zinc finger domain; This domain contains eight conserved cysteine residues that bind to ...	1147-1194	2.17e-21
[+]	FYRN	pfam05964	F/Y-rich N-terminus; This region is normally found in the trithorax/ALL1 family proteins. It ...	2027-2074	4.65e-16



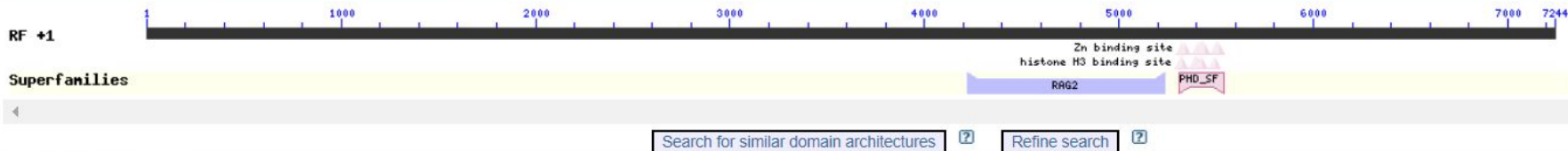
# Доменная структура RAG2

Conserved domains on [lcl|seqsig\_ACTCT\_e2fa10616ef857d71a3ad2f2b252e927]




View Concise Results 

Local query sequence

**Graphical summary** ☐ Zoom to residue level [show extra options »](#) 



## List of domain hits

	Name	Accession	Description	Interval	E-value
	RAG2 super family	cl28748	Recombination activating protein 2; V-D-J recombination is the combinatorial process by which ...	4219-5235	0e+00
	PHD_SF super family	cl22851	PHD finger superfamily; The PHD finger superfamily includes a canonical plant homeodomain (PHD) ...	5308-5541	1.74e-49

PF13341

**RAG2** PHD domain

PFAM

This domain is found at the C-terminus of the **RAG2** protein. The structure of this domain has been shown bound to histone H3 trimethylated at lysine 4 (H3K4me3) [\[\[cite:PUB00057307\]\]](#).

Note: существует RAG1, но он выполняет другую эпигенетическую функцию, экспрессируется в других тканях. Генетическая общая в том, что Recombination activation (The protein encoded by this gene is involved in activation of immunoglobulin V-D-J recombination. The encoded protein is involved in recognition of the DNA substrate, but stable binding and cleavage activity also requires RAG2.)

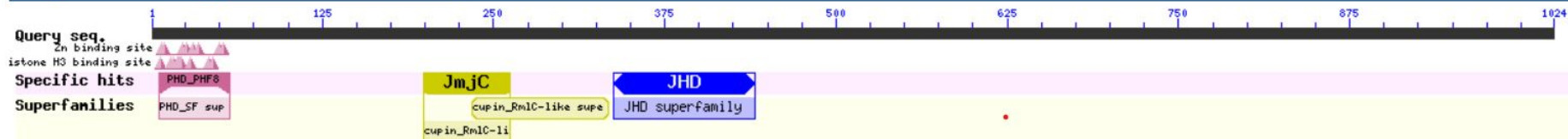
# Доменная структура PHF8

Conserved domains on [ref|NP\_055922.1]

View Concise Results ?

:1-1024 histone lysine demethylase PHF8 isoform 2 [Homo sapiens]

**Graphical summary** ☐ Zoom to residue level show extra options »



?

?

## List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	JHD	pfam17811	Jumonji helical domain; This 4-helix bundle domain is associated with the Jumonji domain ...	338-441	3.77e-62
[+]	PHD_PHF8	cd15642	PHD finger found in histone lysine demethylase PHF8; PHF8, also termed PHD finger protein 8, ...	6-57	1.21e-33
[+]	cupin_RmlC-like super family	cl40423	RmlC-like cupin superfamily; This superfamily contains proteins similar to the RmlC (dTDP ...	234-334	1.92e-15
[+]	JmjC	smart00558	A domain family that is part of the cupin metalloenzyme superfamily; Probable enzymes, but of ...	199-262	2.54e-12

## References:

- Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", **Nucleic Acids Res.**45(D)200-3.
- Marchler-Bauer A et al. (2015), "CDD: NCBI's conserved domain database.", **Nucleic Acids Res.**43(D)222-6.
- Marchler-Bauer A et al. (2011), "CDD: a Conserved Domain Database for the functional annotation of proteins.", **Nucleic Acids Res.**39(D)225-9.
- Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", **Nucleic Acids Res.**32(W)327-331.



# Доменная структура ING4

Conserved domains on [lcl|seqsig\_INGMA\_d2f22314399346f2b214e9ef4e2fe96f]

View Concise Results ?

Local query sequence

## Protein Classification

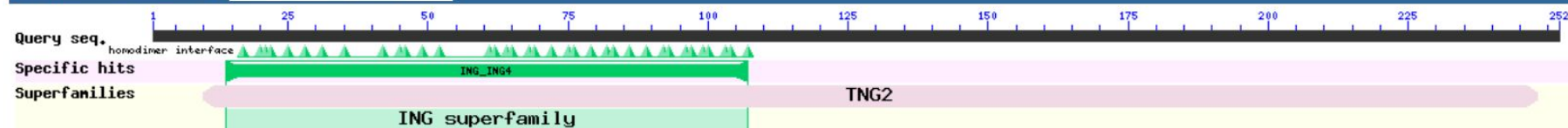
**ING\_ING4 and PHD\_ING4 domain-containing protein**( domain architecture ID 13415405)

ING\_ING4 and PHD\_ING4 domain-containing protein

## Graphical summary

☐ Zoom to residue level

[show extra options >](#)



[Search for similar domain architectures](#) ?

[Refine search](#) ?

## List of domain hits

	Name	Accession	Description	Interval	E-value
<a href="#">+</a>	ING_ING4	cd16862	Inhibitor of growth (ING) domain of inhibitor of growth protein 4 (ING4); ING4, also termed ...	14-107	1.75e-63
<a href="#">+</a>	TNG2 super family	cl34876	Chromatin remodeling protein, contains PhD zinc finger [Chromatin structure and dynamics];	10-248	5.69e-36

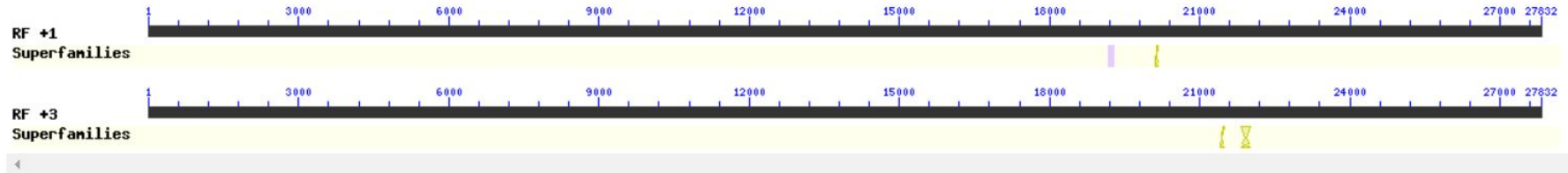


# Доменная структура ZCWPW1

## Graphical summary

☐ Zoom to residue level

[show extra options »](#)



[Search for similar domain architectures](#) [?](#)

[Refine search](#) [?](#)

## List of domain hits

+	Name	Accession	Description	Interval	E-value
[+]	PWWP super family	cl02554	PWWP (Pro-Trp-Trp-Pro) domain; The PWWP domain, named for a conserved Pro-Trp-Trp-Pro motif, ...	20104-20166	2.05e-04
[+]	zf-CW super family	cl06504	CW-type Zinc Finger; This domain appears to be a zinc finger. The alignment shows four ...	19171-19296	3.60e-04
[+]	PWWP super family	cl02554	PWWP (Pro-Trp-Trp-Pro) domain; The PWWP domain, named for a conserved Pro-Trp-Trp-Pro motif, ...	21825-22016	7.03e-10
[+]	PWWP super family	cl02554	PWWP (Pro-Trp-Trp-Pro) domain; The PWWP domain, named for a conserved Pro-Trp-Trp-Pro motif, ...	21402-21491	2.24e-03

# Доменная структура MLL2

Conserved domains on [lcl|seqsig\_MDSQK\_a354bfe1de21c999e8b503585ab8bee0]

View Concise Results

NP\_003473.3 KMT2D [organism=Homo sapiens] [GeneID=8085]

Graphical summary

☐ Zoom to residue level

[show extra options »](#)





[Search for similar domain architectures](#) <sup>?</sup>

[Refine search](#) <sup>?</sup>

## List of domain hits

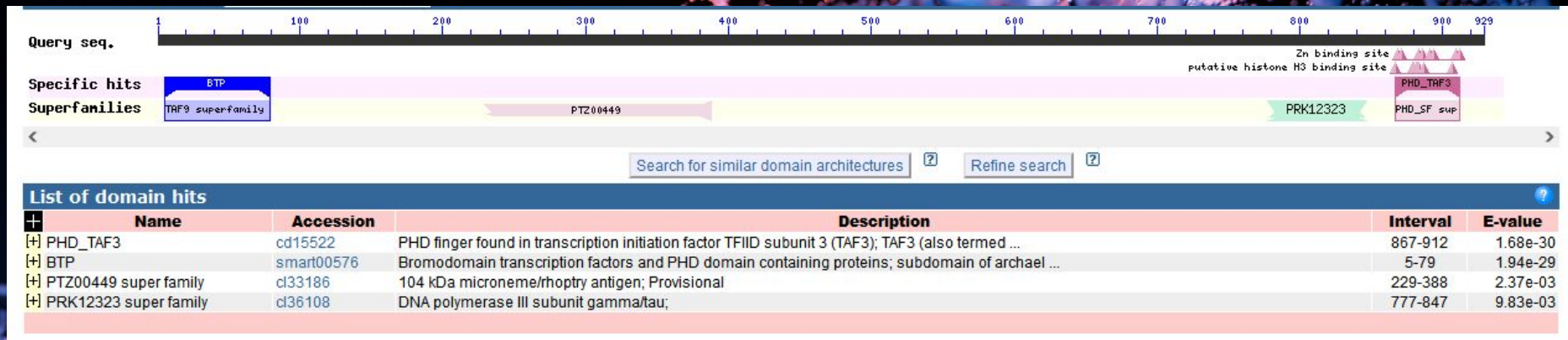
Name	Accession	Description	Interval	E-value
SET_KMT2D	cd19209	SET domain (including post-SET domain) found in histone-lysine N-methyltransferase 2D (KMT2D) ...	5382-5536	1.44e-...
ePHD2_KMT2D	cd15698	Extended PHD finger 2 found in histone-lysine N-methyltransferase 2D (KMT2D); The extended ...	5032-5138	9.33e-...
HMG-box_KMT2D	cd22027	high mobility group (HMG)-box found in histone-lysine N-methyltransferase 2D (KMT2D) and ...	1995-2078	9.47e-...
ePHD1_KMT2D	cd15695	Extended PHD finger 1 found in histone-lysine N-methyltransferase 2D (KMT2D); The extended ...	134-217	4.43e-...
PHD5_KMT2D	cd15601	PHD finger 5 found in Histone-lysine N-methyltransferase 2D (KMT2D); KMT2D, also termed ...	1506-1556	6.72e-...
PHD3_KMT2D	cd15597	PHD finger 3 found in Histone-lysine N-methyltransferase 2D (KMT2D); KMT2D, also termed ...	1378-1428	7.92e-...
FYRC	smart00542	FY-rich domain, C-terminal region; is sometimes closely juxtaposed with the N-terminal region ...	5240-5327	1.88e-...
PHD5_KMT2C_like	cd15513	PHD finger 5 found in Histone-lysine N-methyltransferase 2C (KMT2C) and PHD finger 4 found in ...	1429-1475	6.34e-...
PHD_SF super family	cl22851	PHD finger superfamily; The PHD finger superfamily includes a canonical plant homeodomain (PHD) ...	275-320	1.45e-...
PHD1_KMT2C_like	cd15509	PHD finger 1 found in Histone-lysine N-methyltransferase 2C (KMT2C) and 2D (KMT2D); KMT2C, ...	228-273	1.26e-...
FYRN	pfam05964	F/Y-rich N-terminus; This region is normally found in the trithorax/ALL1 family proteins. It ...	5181-5232	2.08e-...
PHA03247 super family	cl33720	large tegument protein UL36; Provisional	2108-2600	2.17e-...
PHA03247 super family	cl33720	large tegument protein UL36; Provisional	690-1272	1.02e-...
PHA03247 super family	cl33720	large tegument protein UL36; Provisional	469-860	8.84e-...
PHA03247 super family	cl33720	large tegument protein UL36; Provisional	4162-4426	4.96e-...

## References:

-  Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", **Nucleic Acids Res.**45(D)200-3.
-  Marchler-Bauer A et al. (2015), "CDD: NCBI's conserved domain database.", **Nucleic Acids Res.**43(D)222-6.
-  Marchler-Bauer A et al. (2011), "CDD: a Conserved Domain Database for the functional annotation of proteins.", **Nucleic Acids Res.**39(D)225-9.
-  Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", **Nucleic Acids Res.**32(W)327-331.

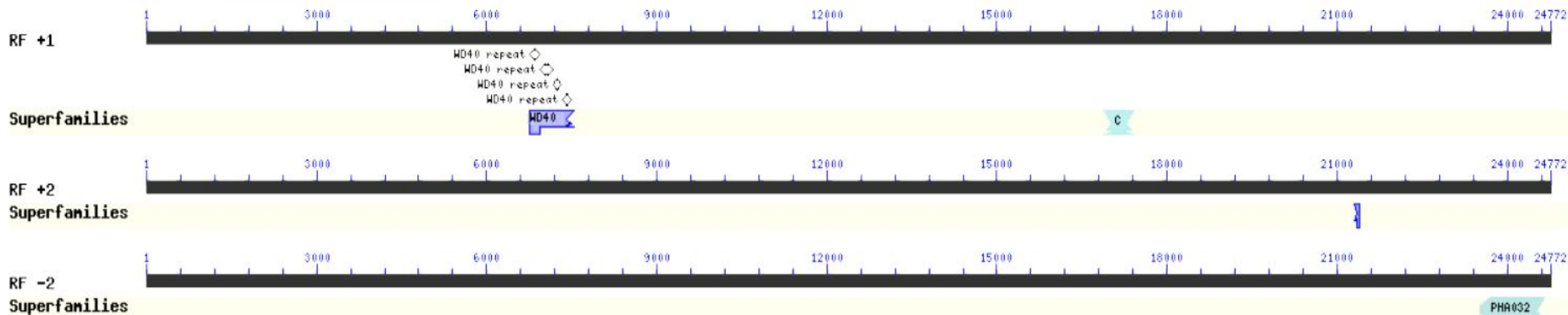


# Доменная структура TAF3



# Доменная структура WDR5

Graphical summary ☐ Zoom to residue level [show extra options »](#)



[Search for similar domain architectures](#)

[Refine search](#)

## List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	WD40 super family	cl29593	WD40 domain, found in a number of eukaryotic proteins that cover a wide variety of functions ...	6745-7539	1.73e-13
[+]	CaiC super family	cl43103	Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II [Lipid transport and metabolism, ...	16879-17427	2.31e-03
[+]	WD40 super family	cl29593	WD40 domain, found in a number of eukaryotic proteins that cover a wide variety of functions ...	21308-21400	8.32e-03
[+]	PHA03247 super family	cl33720	large tegument protein UL36; Provisional	23519-24667	3.16e-03



# Доменная структура CCDC101

Conserved domains on [gi|215820637|ref|NM\_138414.2|]

View Concise Results 

Homo sapiens SAGA complex associated factor 29 (SGF29), mRNA

**Graphical summary** ☐ Zoom to residue level [show extra options »](#) 



[Search for similar domain architectures](#) 

[Refine search](#) 

## List of domain hits

	Name	Accession	Description	Interval	E-value
	DUF1325 super family	cl18505	SGF29 tudor-like domain; This domain is found in the yeast protein SAGA-associated factor 29. ...	657-1049	6.67e-61