生物信息资源与实践第二次作业

提高1501班

周克坚

U201513485

# 一、PubMed数据库搜索5篇相应的论文，阅读摘要并总结其功能

Aurora-B基因与人类癌症密切相关，是有丝分裂的关键调控因子，能够调节人体细胞的末端转换。



# 二、序列数据库获取核酸和蛋白质序列

核酸序列见“sequence.fasta.txt”



蛋白质序列见“Q96GD4.fasta.txt”



# 三、基因功能分析，查找基因组的定位、功能结构域和GO注释信息

## 基因组的定位：

Location: 17p13.1

Annotation release Status Assembly Chromosome Location

-------------------------------------------------------------------------------------------------

109 current GRCh38.p12 (GCF\_000001405.38) 17 NC\_000017.11 (8204731..8210767, complement)

105 previous assembly GRCh37.p13 (GCF\_000001405.25) 17 NC\_000017.10 (8108049..8113944, complement)

Exont count: 10

## 功能结构域：

丝氨酸/苏氨酸蛋白激酶，催化结构域



Conserved Domains (2)summary

smart00220:

Location:36 - 286

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:29 - 298

PKc\_like; Protein Kinases, catalytic domain

NM\_001284526.1 → NP\_001271455.1  REVIEWEDaurora kinase B isoform 3

Description: Transcript Variant: This variant (3) uses an al...

Source sequence(s): BC009751, BC013300, BE514591, BP225027

Consensus CDS: CCDS67162.1

UniProtKB/Swiss-Prot: Q96GD4

Related Ensembl: ENSP00000313950.6, OTTHUMP00000259369, ENST00000316199.10, OTTHUMT00000441842

Conserved Domains (2)summary

smart00220:

Location:78 - 328

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:71 - 340

PKc\_like; Protein Kinases, catalytic domain

NM\_001313950.1 → NP\_001300879.1  REVIEWEDaurora kinase B isoform 1

Description: Transcript Variant: This variant (4) uses an al...

Source sequence(s): BC009751, BE514591, BP225027, CX781462

Consensus CDS: CCDS11134.1

UniProtKB/Swiss-Prot: Q96GD4

Conserved Domains (1)summary

cl21453:

Location:70 - 339

PKc\_like; Protein Kinases, catalytic domain

NM\_001313951.1 → NP\_001300880.1  REVIEWEDaurora kinase B isoform 2

Description: Transcript Variant: This variant (5) contains a...

Source sequence(s): BC009751, BE797606, BP225027

Consensus CDS: CCDS58514.1

UniProtKB/Swiss-Prot: Q96GD4

Conserved Domains (2)summary

smart00220:

Location:36 - 286

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:29 - 298

PKc\_like; Protein Kinases, catalytic domain

NM\_001313952.1 → NP\_001300881.1  REVIEWEDaurora kinase B isoform 4

Description: Transcript Variant: This variant (6) lacks an e...

Source sequence(s): BC009751, BE514591, BM905813, BP225027

UniProtKB/Swiss-Prot: Q96GD4

Conserved Domains (2)summary

smart00220:

Location:37 - 287

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:30 - 299

PKc\_like; Protein Kinases, catalytic domain

NM\_001313953.2 → NP\_001300882.1  REVIEWEDaurora kinase B isoform 5

Description: Transcript Variant: This variant (7) uses alter...

Source sequence(s): AC135178, BC009751, BU153790

Consensus CDS: CCDS82065.1

UniProtKB/Swiss-Prot: Q96GD4

Related Ensembl: ENSP00000462207.1, OTTHUMP00000259372, ENST00000578549.5

Conserved Domains (2)summary

smart00220:

Location:102 - 295

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:70 - 307

PKc\_like; Protein Kinases, catalytic domain

NM\_001313954.1 → NP\_001300883.1  REVIEWEDaurora kinase B isoform 6

Description: Transcript Variant: This variant (8) uses an al...

Source sequence(s): AF004022, BC009751, BE514591, BP225027, BU181480

UniProtKB/Swiss-Prot: Q96GD4

Conserved Domains (2)summary

smart00220:

Location:27 - 175

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:21 - 187

PKc\_like; Protein Kinases, catalytic domain

NM\_001313955.1 → NP\_001300884.1  REVIEWEDaurora kinase B isoform 7

Description: Transcript Variant: This variant (9) lacks two ...

Source sequence(s): BC009751, BE514591, BI199088, BP225027

UniProtKB/Swiss-Prot: Q96GD4

Conserved Domains (2)summary

smart00220:

Location:10 - 159

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:10 - 171

PKc\_like; Protein Kinases, catalytic domain

NM\_004217.3 → NP\_004208.2  REVIEWEDaurora kinase B isoform 1

See proteins identical to NP\_004208.2 at http://130.14.18.104/ipg/NP\_004208.2/

Description: Transcript Variant: This variant (1) represents...

Source sequence(s): AC135178, BC000442, BC009751

Consensus CDS: CCDS11134.1

UniProtKB/Swiss-Prot: Q96GD4

Related Ensembl: ENSP00000463999.1, OTTHUMP00000135331, ENST00000585124.5, OTTHUMT00000226995

Conserved Domains (1)summary

cl21453:

Location:70 - 339

PKc\_like; Protein Kinases, catalytic domain

RNA

NR\_132730.1 RNA Sequence

Description: Transcript Variant: This variant (10) uses an a...

Source sequence(s): BC009751, BE514591, BP225027, CN286704

NR\_132731.1 RNA Sequence

Description: Transcript Variant: This variant (11) lacks an ...

Source sequence(s): AY677083, BC009751

RefSeqs of Annotated Genomes: Homo sapiens Annotation Release 109

The following sections contain reference sequences that belong to a specific genome build.

This section includes genomic Reference Sequences (RefSeqs) from all assemblies on which

this gene is annotated, such as RefSeqs for chromosomes and scaffolds (contigs) from both

reference and alternate assemblies. Model RNAs and proteins are also reported here.

Reference GRCh38.p12 Primary Assembly

Genomic

NC\_000017.11

Range 8204730..8210766, complement

Download GenBank FASTA Sequence Viewer (Graphics)

mRNA and Protein(s)

XM\_017025307.2 → XP\_016880796.1 aurora kinase B isoform X1

UniProtKB/Swiss-Prot: Q96GD4

Conserved Domains (2)summary

smart00220:

Location:36 - 286

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:29 - 298

PKc\_like; Protein Kinases, catalytic domain

XM\_017025308.2 → XP\_016880797.1 aurora kinase B isoform X2

XM\_011524072.3 → XP\_011522374.1 aurora kinase B isoform X1

See proteins identical to XP\_011522374.1 at http://130.14.18.104/ipg/XP\_011522374.1/

UniProtKB/Swiss-Prot: Q96GD4

Conserved Domains (2)summary

smart00220:

Location:36 - 286

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:29 - 298

PKc\_like; Protein Kinases, catalytic domain

XM\_017025310.1 → XP\_016880799.1 aurora kinase B isoform X3

Conserved Domains (2)summary

smart00220:

Location:27 - 175

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:21 - 187

PKc\_like; Protein Kinases, catalytic domain

XM\_017025309.1 → XP\_016880798.1 aurora kinase B isoform X3

Conserved Domains (2)summary

smart00220:

Location:27 - 175

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:21 - 187

PKc\_like; Protein Kinases, catalytic domain

XM\_017025311.1 → XP\_016880800.1 aurora kinase B isoform X3

Conserved Domains (2)summary

smart00220:

Location:27 - 175

S\_TKc; Serine/Threonine protein kinases, catalytic domain

cl21453:

Location:21 - 187

PKc\_like; Protein Kinases, catalytic domain

## GO注释信息：

Gene Ontology Provided by GOA (www.ebi.ac.uk/GOA)

Function GO ID Evidence

-------------------------------------------------------------------------------------------------

ATP binding GO:0000000 -

histone serine kinase activity GO:0000000 PMID: 21873635

histone serine kinase activity GO:0000000 -

kinase binding GO:0000000 PMID: 23036704

metal ion binding GO:0000000 -

protein binding GO:0000000 PMID: 12689593,15064709,16179162,16239925,16291752,17099693,17681274,17726514,17956729,18243099,18477699,19357306,20562864,20739936,20959462,21051298,21346195,21988832,22939624,23036704,24571573,24857377,25036637,26829474,27332895,29340707,29568061

protein serine/threonine kinase activity GO:0000000 PMID: 21397845

protein serine/threonine kinase activity GO:0000000 PMID: 22724069

protein serine/threonine kinase activity GO:0000000 -

protein serine/threonine/tyrosine kinase... GO:0000000 PMID: 19774610

Processs GO ID Evidence

-------------------------------------------------------------------------------------------------

abscission GO:0000000 -

aging GO:0000000 -

anaphase-promoting complex-dependent cat... GO:0000000 -

attachment of spindle microtubules to ki... GO:0000000 PMID: 19774610

cell proliferation GO:0000000 -

cellular response to UV GO:0000000 PMID: 20959462

cleavage furrow formation GO:0000000 PMID: 16103226

histone H3-S28 phosphorylation GO:0000000 -

histone modification GO:0000000 PMID: 19774610

mitotic cytokinesis checkpoint GO:0000000 -

mitotic spindle assembly checkpoint GO:0000000 -

mitotic spindle midzone assembly GO:0000000 PMID: 16103226

mitotic spindle midzone assembly GO:0000000 PMID: 19774610

mitotic spindle organization GO:0000000 PMID: 21873635

negative regulation of B cell apoptotic ... GO:0000000 PMID: 20959462

negative regulation of cytokinesis GO:0000000 -

negative regulation of protein binding GO:0000000 PMID: 21820309

negative regulation of transcription by ... GO:0000000 PMID: 20959462

positive regulation of cytokinesis GO:0000000 PMID: 16103226

positive regulation of cytokinesis GO:0000000 PMID: 19774610

positive regulation of telomerase activi... GO:0000000 PMID: 21531765

positive regulation of telomere capping GO:0000000 PMID: 21531765

positive regulation of telomere maintena... GO:0000000 PMID: 21531765

protein autophosphorylation GO:0000000 PMID: 19774610

protein localization to kinetochore GO:0000000 PMID: 19468067

protein phosphorylation GO:0000000 PMID: 21820309,22724069

regulation of chromosome segregation GO:0000000 PMID: 19774610

regulation of cytokinesis GO:0000000 PMID: 21873635

regulation of signal transduction by p53... GO:0000000 -

spindle organization GO:0000000 PMID: 21820309

Component GO ID Evidence

-------------------------------------------------------------------------------------------------

chromocenter GO:0000000 -

chromosome passenger complex GO:0000000 PMID: 21873635

chromosome passenger complex GO:0000000 PMID: 15260989,18591255

chromosome passenger complex GO:0000000 PMID: 19774610

condensed chromosome, centromeric region GO:0000000 PMID: 19465021

condensed nuclear chromosome, centromeri... GO:0000000 PMID: 21873635

condensed nuclear chromosome, centromeri... GO:0000000 PMID: 18195732

cytosol GO:0000000 -

kinetochore GO:0000000 PMID: 22024163

midbody GO:0000000 PMID: 17726514

midbody GO:0000000 PMID: 19774610

mitotic spindle midzone GO:0000000 -

nucleoplasm GO:0000000 -

nucleoplasm GO:0000000 -

nucleus GO:0000000 PMID: 20959462

spindle GO:0000000 PMID: 19774610

spindle microtubule GO:0000000 PMID: 21873635

spindle midzone GO:0000000 PMID: 21873635

spindle pole centrosome GO:0000000 PMID: 21873635

# 四、三级结构信息获取

下图为4AF3的三级结构图

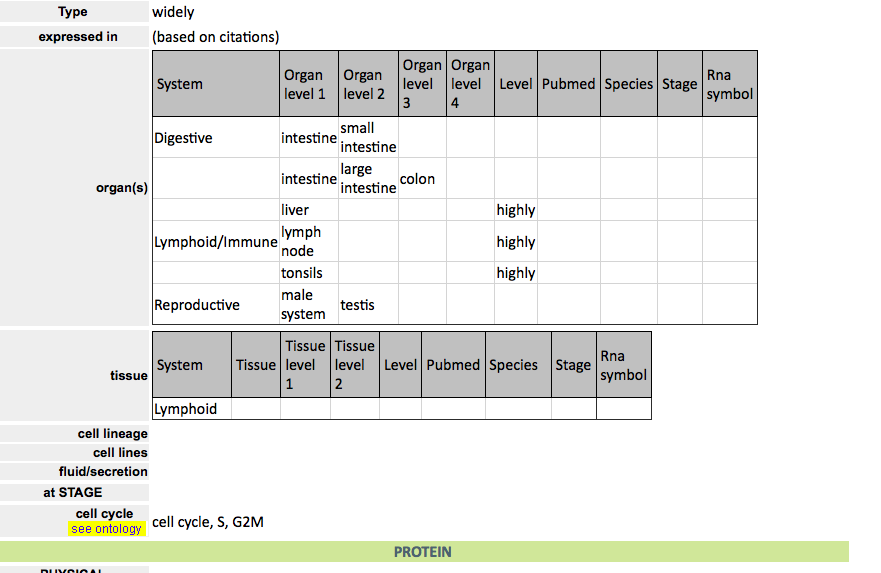


# 五、基因表达、变异和疾病相关等分析

## GeneCards对AURKB基因的总结：

AURKB（Aurora Kinase B）是蛋白质编码基因。与AURKB相关的疾病包括巨细胞胶质母细胞瘤和脊髓小脑共济失调10。其相关途径是基因表达和细胞周期，有丝分裂。与该基因相关的GO注释包括转移酶活性，转移含磷基团和蛋白酪氨酸激酶活性。

## GENATLAS对AURKB基因与疾病有关的信息如下：



# 六、序列处理根据人类Aurora-B基因，寻找符合[AILVFPMW]K.[DE]模体的位置记相应序列并输出到文本文件，结果格式为：>PIk1Position 1 – position 2XXX



输出结果保存在：

