



WIE-biotech

立效生物

DNA 甲基化

基础版

方案组

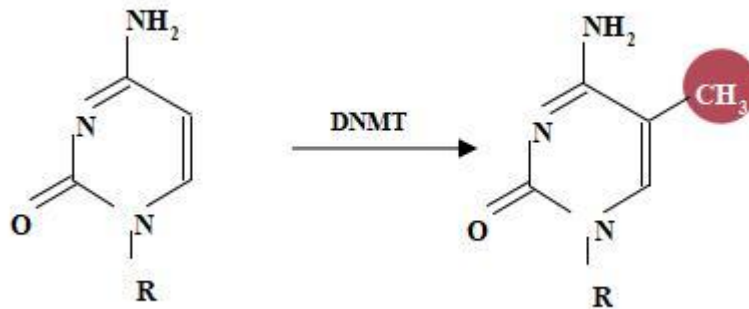
2023/10/26

目录

一、什么是 DNA 甲基化？	2
二、DNA 甲基化发生的机制是什么？	2
三、DNA 甲基化发生的意义是什么？	2
四、甲基化检测方法有哪些？	3
五、DNA 甲基化的临床应用如何？	3
六、怎么才能知道基因启动子区有没有 cpg 岛呢？以及 cpg 岛具体位置？	4
七、DNA 甲基化数据库	10
八、其它甲基化数据库（RNA，组蛋白）	12

一、什么是 DNA 甲基化？

DNA 甲基化是指基因组胞嘧啶-鸟嘌呤二核苷酸 (CpG) 上的 5 位碳原子，在 DNA 甲基转移酶家族 (DNMTs, DNMT1、DNMT3a 和 DNMT3b) 的介导下，共价结合一个游离的甲基形成 5-甲基胞嘧啶的过程。



二、DNA 甲基化发生的机制是什么？

CPG 二核苷酸不均匀地分布于基因组中，人们把基因组中 CPG 二核苷酸出现频率较高的一段区域称为 CPG 岛 (CGI)。

在人类中，CGI 一般存在于基因组的第 1 外显子区和启动子区，具有调控基因表达的作用。

正常人体组织中，大部分基因组重复区域的 CPG 二核苷酸是甲基化的，而 CGI 内的 CPG 二核苷酸则通常保持非甲基化状态；当发生肿瘤（包括肺癌）时，则可出现特征性的全局 DNA 低甲基化（通常以非编码区为靶点）和肿瘤抑制基因启动子中 CPG 的位点特异性高甲基化。

三、DNA 甲基化发生的意义是什么？

当抑癌基因启动子区高甲基化时，可引起抑癌基因转录抑制、失活或沉默；基因组整体甲基化水平降低，可引起原基因组不稳定、癌基因活化等，导致肿瘤的形成或进展；其中，抑癌基因启动子区高甲基化或原癌基因低甲基化都可导致

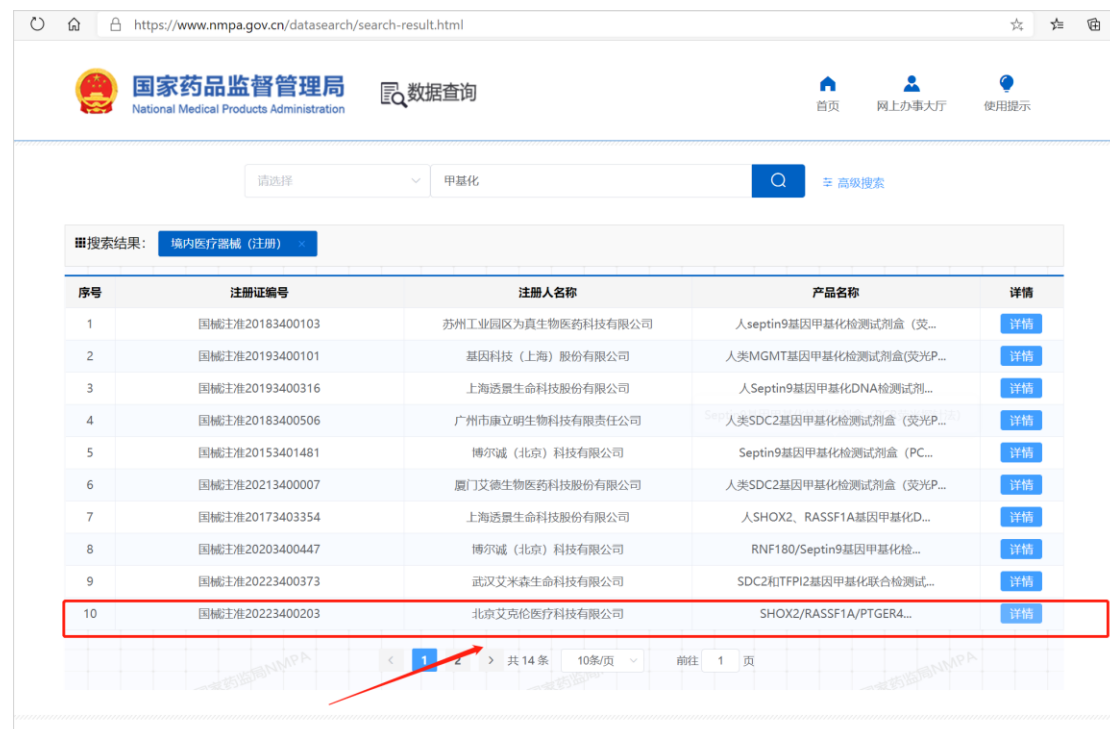
肿瘤的发生[5]。

四、甲基化检测方法有哪些？

随着医学研究的进步，各种各样的甲基化检测方法层出不穷。

目前 DNA 甲基化预处理的方法主要有 3 类，即甲基化敏感的限制性内切酶消化、亚硫酸氢盐转换以及亲和富集法。

其中基于亚硫酸氢盐转换的甲基化特异性 PCR 是一种被广泛使用的经典的甲基化检测方法。荧光 PCR 法操作简便，且具有灵敏度高、特异性强、重复性好等特点。不断受到市场青睐。多个公司的荧光 PCR 法的甲基化检测试剂盒受到 NMPA 的批准。艾克伦的肺癌甲基化产品也于 2022 年 02 月 15 日获批。



国家药品监督管理局
National Medical Products Administration

数据查询

甲基化

高级搜索

搜索结果: 境内医疗器械 (注册)

序号	注册证编号	注册人名称	产品名称	详情
1	国械主准20183400103	苏州工业园区为真生物医药科技有限公司	人septin9基因甲基化检测试剂盒 (荧...	详情
2	国械主准20193400101	基因科技 (上海) 股份有限公司	人类MGMT基因甲基化检测试剂盒 (荧光P...	详情
3	国械主准20193400316	上海透景生命科技股份有限公司	人Septin9基因甲基化DNA检测试剂...	详情
4	国械主准20183400506	广州市康立明生物科技有限责任公司	人类SDC2基因甲基化检测试剂盒 (荧光P...	详情
5	国械主准20153401481	博尔诚 (北京) 科技有限公司	Septin9基因甲基化检测试剂盒 (PC...	详情
6	国械主准20213400007	厦门艾德生物医药科技股份有限公司	人类SDC2基因甲基化检测试剂盒 (荧光P...	详情
7	国械主准20173403354	上海透景生命科技股份有限公司	人SHOX2、RASSF1A基因甲基化D...	详情
8	国械主准20203400447	博尔诚 (北京) 科技有限公司	RNF180/Septin9基因甲基化检...	详情
9	国械主准20223400373	武汉艾米森生命科技有限公司	SDC2和TFPI2基因甲基化联合检测试...	详情
10	国械主准20223400203	北京艾克伦医疗科技有限公司	SHOX2/RASSF1A/PTGER4...	详情

共 14 条 10 条/页 前往 1 页

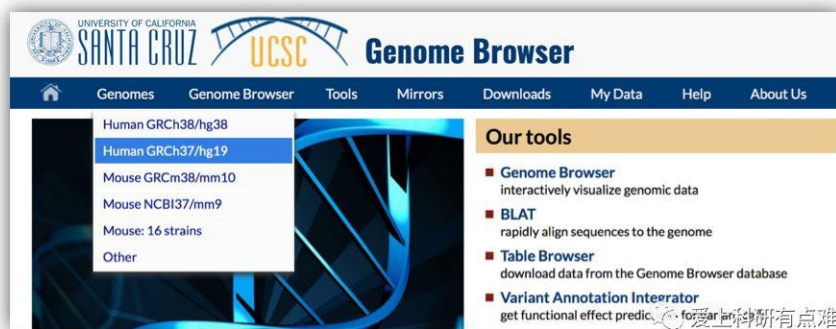
五、DNA 甲基化的临床应用如何？

肺癌的早期筛查和诊断能延长患者生存期和降低病死率。有实验表明，吸烟患者低剂量 CT(LDCT)早期检测能降低肺癌患者的死亡率，然而，由于 LDCT 存在费用过高和辐射暴露危险等缺陷，限制了其临床应用。游离 DNA 甲基化可作

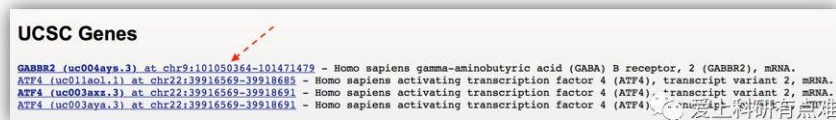
为强有力的生物标志物，准确且定量检测游离 DNA 甲基化水平对于肿瘤的早期辅助诊断和疗效监测具有十分重要的意义。目前，DNA 甲基化作为生物标志物已广泛用于肺癌、胃癌、乳腺癌、结直肠癌、肝癌、胰腺癌等癌症的诊疗。

六、怎么才能知道基因启动子区有没有 cpG 岛呢？以及 cpG 岛具体位置？（快速通道：直接跳 3）

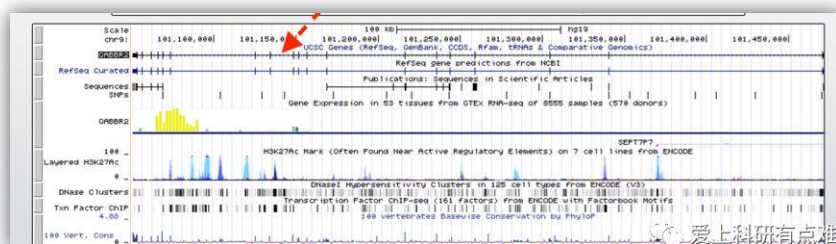
1、首先下载基因的启动子序列 → 然后把序列导入预测软件预测 cpG 岛的位置 → 先下载基因启动子序列 → 首先走 UCSC genome browser(<http://genome.ucsc.edu>) → 选取 Genomes 里的 hg19 版本（为了后面适配大多数的软件/数据库）



➤ 以 GABBR2 基因为例，输入后搜索



➤ 点击最上面一条 GABBR2...再次进入



➤ 再次点击 UCSC gene(也就是图示红色箭头指向的位置，进入...)

Page Index	Sequence and Links	UniProtKB Comments	Genetic Associations	MalaCards	CTD
Gene Alleles	RNA-Seq Expression	Microarray Expression	RNA Structure	Protein Structure	Other Species
GO Annotations	mRNA Descriptions	Pathways	Other Names	Model Information	Methods

Data last updated: 2013-06-14

Sequence and Links to Tools and Databases

Genomic Sequence (chr9:101,050,364-101,471,479)			mRNA (may differ from genome)	Protein (941 aa)	
Gene Sorter	Genome Browser	Other Species FASTA	VisiGene	Gene interactions	Table Schema
BioGPS	CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards
GeneNetwork	Gepis Tissue	H-INV	HGNC	HPRD	Human Cortex Gene Expression
Lynx	MGI	MOPED	neXtProt	OMIM	PubMed
Reactome	Stanford SOURCE	Treefam	UniProtKB	Wikipedia	

- 点击 Genomic Sequence(进入如下的序列设置界面)

Get Genomic Sequence Near Gene

Note: if you would prefer to get DNA for more than one feature of this track at a time, try the [Table Browser](#) using

Sequence Retrieval Region Options:

☒ Promoter/Upstream by bases

☒ 5' UTR Exons

☒ CDS Exons

☒ 3' UTR Exons

☒ Introns

☐ Downstream by bases

☒ One FASTA record per gene.

☐ One FASTA record per region (exon, intron, etc.) with extra bases upstream (5') and extra downstream (3')

☐ Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, the bases will be padded with N's.

Sequence Formatting Options:

☐ Exons in upper case, everything else in lower case.

☒ CDS in upper case, UTR in lower case.

☐ All upper case.

☐ All lower case.

☐ Mask repeats: ☒ to lower case ☐ to N

- Promoter 区/Upstream 可以设置宽些,以防有漏网的 CpG 岛 → 此处, 设置在 2000, 其他按图里的设置(尤其需要注意的是,要点选“CDS in upper case, UTR in lower case”)这样可以根据大小写快速定位到转录起始位点 → 点击 submit 后,得到整条基因以及上游 2000bp 的序列 → 此处从第一行开始下滑,第一次出现的大写字母就是转录起始位点 TSS

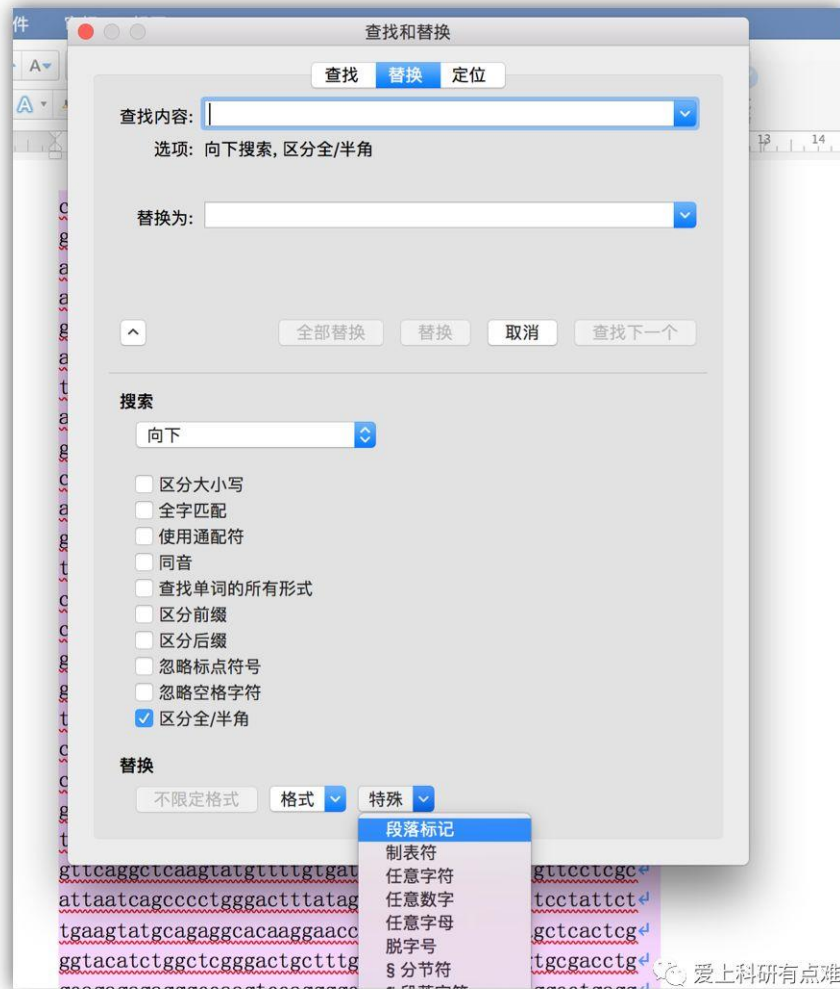
```

>hg19_knownGene_uc004ays.3 range=chr9:101050364-101473479 5'pad=0 3'pad=0 strand=- repeatMasking=none
cagagattggcctggagccatgctgatggagagtggatgggtgccccct
gggtcagggcagatcccaggtcatgctcttggggcaggtcacccacc
ataaaagaaagcagaggggctgtagggcaggatgccaggtaggccttca
acccttagttaaattgcaggcaaatgacaaagttaccagcttcaacttattcca
gatgtagggtcagtgagacactctatattggcacagcttgaaggaggac
acaaatgactaaagaaagaaattagaaaaagaaaaaaccaagtgaaga
tggtcattcagaacctgagctaaaggtttgaggaatggcaggtgccaga
agactatgactggaggatatttggatgtcaagtggagacagcagttgagc
gagttatctggggattacactgtccttggaccatttcttccctccctcc
cttccctccctccctccctggaggagatgctcaggtgagttctgtctatgc
aaacctccagcttgggaaagaaatgccaagtggagccgggagaggt
ggagttctatagaagcctggaaagagaattagggcagagagaatttggcc
tggttccctggcgtgtccaaagttactctctagtagttctgtcagcct
ccactcacatcttctgtcttctaacaaactcagcctggattcagagctgc
ccttgcctcactccagtagagctggtgtgtggccatcatcatcatcct
gggtcaaatccctcaccatagtcggagggaatcgatccagagcaact
gagtcagggcagtgctcctcctggagagaaacttgagctccacatccc
tggtccctcactgactgtgcagaggttctgtggctctgggttccgtt
ccttcagtagctgtgtgtcttccactagctacagaacctctctgagcac
cagtgatctgcctgtaaagttaggtgagtaacacagtagtctacagg
gcagttgtggggattacatcataagcacacagtaagtagctcagcgaagg
tcaaaacttccattcttgaaggtcgtcgatcttccacagtcctgt
gttcaggtcgaagtgttctgtgatttgcagtcctactgtgttctcgc
atatacagccctgggactttatagattaatgacagtttctattct
tgaagtatgcagagccacaaaggaacccaggtcttagtcagtcactcg
ggtagatctggctgggactgcttggctaaagtggccagtgagcctg
gcagagagagggccaaagtcaggggcgcggcccccgcctggcctgagg
atagtgctcccacccgggagacttctaggggggggccatcaagaatgg
ggagcggggagggtcaaatgcctggagggggcggtagcgcacgtaaaggt
gggacccagcttctgtgggttggcaaggcggtcgctccgtctgtgt
ggggaggaaggggttctcagcagcccaagccctaaaggccgggagcctg
gcaccaacggcggtcagcagccgtggcgcttccagcagcccgga
ctcgttcaggggtcgtgggcagcaaaagacgggagagcgtaggctg
cagatgccttgggagccttctgggaacattctcctcccagacctcc
cggccagacctccgagaagcgttccaaacccagctgaggtgtacagg
cagcgtcggcgagcggcgcgccgactgctgtgagcgcgggtcc
cgggtcccagatcgtgtccccggcgctggtgctgaggtcagcgg
tcggcgggcagcggcagctggcgagcggagggcgggcggtgtgtg
cgtgtgtgggaaggtggcctccccggcgccgagcgcagcgcct
cccccccccgcgcgagcgtctgtcccgctccgtggagggcgccga
ggcggcctgagcagcctcgccttgcctcccgcttctcgtccgc
cctccccggcgagctccaggggtgcccctagcagctccggcgga
gagcgttcagagcgcgcagggggcggggagggcgccggtgcgggg
cgcgcgctggagagagcgcgcgagagcggcccccctctccgc
gttgcctgctgtcctcccgctcccgactcgcctcgtcccccctt
cccgctgattgattgtcacggcgcccgctgcccgcgcgcgcgc
ggcgcttctgagcagcggcaacctagcccagagcggagcggggc
cgggcgcgcttgcggcgcgccggggaagacctggcgcggggc
ggggcgcggttaggcatgcccgcagtgagcggcgcccgagccc
ggggcgcggtatggcttccccgggagctccggcagccggcgccgc
CGCCGCGCCACCGCGCCGCGCGCTGCTACTGCTACTGCTGCG
CTGCTGCTGCTGCGCGCCGCGCGCTGGGGCTGGGCGCGGGCGCC
CGGCGCGCCAGCAGCGCGCGCTGCTCATCATGGGCTCATGCC
TCACCAAGGAGGTGCCAAGGGCAGCATCGGGCGGCTGCTCCCGC
GTGGAACGTGGCATCGAGCAGATCCGACAGGTCATCTGCGCCCTA
CTTCTCGACCTGCGGCTCTATGACACGGAGGTGAGTGCTCCGGTCC

```

爱上科研有点难

从 TSS 往下再数 10 行（500bp 左右）→连同 TSS 到页面最上面第一行的所有小写字母部分就可以定义为此基因启动子区域序列→拷贝到 word 中→替换掉段落标记（换行符）→就得到了 clean DNA sequence



➤ (小 tips) → 这整段序列其实 >2500bp → 因为里面还包含了 5' 端 UTR exon 可以在 Word 中统计出此序列的长

➤ 回归正题，得到序列后 → 开始用软件预测 cpG 岛啦 → 此处推荐两个软件

① 第一个: CpG Finder

(<http://www.softberry.com/berry.phtml?topic=cpfinder&group=programs&subgroup=pro-moter>)

界面如下

CpG Finder

Paste sequence here:

GGTGAGTGTC CGGCTGCCCCTGCCGTCGGGGCTTCGGGGGGCATTGGGGTGGGTGG
TTTGGACACAGGCAACCTCCGGCCTTTGGTCAGATCCCCGATTGACCGGCCCGCGCC
TGCCGAGGGCAGAGAGGGCTGGTGCCCA GTGTTACCCCCGCGGTCTCTGGGGCT
TGGCTGGCACAGCTGGTAACATTGCCAAGCACCCCTCCTCTGGTCTTAT

Or load sequence from file:
Local file name:

选取文件

未选择文件

Min length of island to find

200

Min percent G and C

50

Min CpG number

0

Min gc_ratio = $P(\text{CpG})/(\text{expected})P(\text{CpG})$

0.600

Extend island if its lengths less then required

☐

Process

Reset

[\[Help\]](#)

[\[Example\]](#)

从 word 里贴进序列后一般按默认参数，点击 Process 后静待结果。

② EMBOSS Cpplot (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpplot/)

界面如下

[illegible]

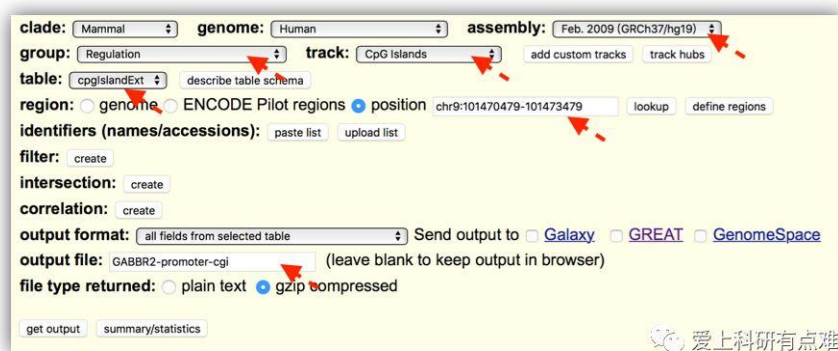
从 word 里贴进序列后一般按默认参数，点击 Submit 后静待结果→这个软件还可以生成整条序列的 cpg 分布情况图试试吧

- 预测软件比较多，原理都差不多了解 CpG 岛的定义可以参见维基百科（https://en.wikipedia.org/wiki/CpG_site#CpG_islands）

二、用基因启动子区染色体位置 → 在 UCSC tables 中已记录的 cpG 岛数据中直接检索 → 还记得第一层级的序列 word 中统计的序列长度吗 → 记为 *prolength* (比如上面例子中 *prolength*=3000) → 然后再回到第一层级中得到的序列界面说起 → 要靠此页面获取此基因启动子区的染色体位置



如图, 如果此页面显示的是 *strand=-* 那么染色体位置就是 chr9: 101473479-*prolength* ~ 101473479 → 比如还是以上面 *GABBR2* 为例, *prolength*=3000 染色体位置是 chr9: 101470479 ~ 101473479 → 而假如此页面显示的是 *strand=+* 那么染色体位置就是 chr9: 101050364 ~ 101050364+*prolength* 小本本记下此染色体位置后 → 开始探索 UCSC tables, 网址如下 <http://genome.ucsc.edu/cgi-bin/hgTables>, 按如下图的设置



➤ 填入上面准备过的启动子染色体位置 → 定义 output file 名称 → 点击 get output 得到结果文件

3、<http://www.bio-bio-bio.com/Home/MiniTools>

如下图...

基因转录起始位点附近 CpG 岛查询

基因转录起始位点附近 CpG 岛 (CGI) 的甲基化情况常常影响基因的表达水平。这个小插件可以帮助科研人员迅速检索到某个基因的 TSS 附近有无 CGI 以及 CGI 的具体信息。(Tips: 这里采用的 TSS 附近是指 TSS 上游 5000bp 到 TSS 下游 1000bp 的区域)

请输入基因的名称

GABBR2

提交

查询结果

检索基因	GENE-chrom	TSSregion-chromStart	TSSregion-chromEnd	CGI-chrom	CGI-chromStart	CGI-chromEnd	CGI-name	CGI-length	CGI-cpgNum
GABBR2	chr9	101470019	101475019	chr9	101469579	101472154	CpG: 229	2575	229

爱上科研有点难

得到 cpG 岛的染色体位置后 → 可以直接调用 UCSC 获得其序列 → 比如上图里的

<http://genome.ucsc.edu/cgi-bin/das/hg19/dna?segment=chr9:101469579,101472154>

(上述链接输入浏览器, 即可瞬间得到 DNA 序列)

七、DNA 甲基化数据库

1、EWAS Data Hub (<https://ngdc.cncb.ac.cn/ewas/datahub/index>)

数据内容:

标准化的 DNA 甲基化芯片数据以及相关的元数据

75,344 个样本, 主要来自 GEO、TCGA、ArrayExpress 和 Encode 等数据库。

EWAS 数据中心能提供不同背景下的参考 DNA 甲基化图谱, 涉及 81 种组织/细胞类型 (包含 25 个脑部和 25 种血细胞类型), 6 个祖先类别和 67 种疾病 (包括 39 种癌症)。

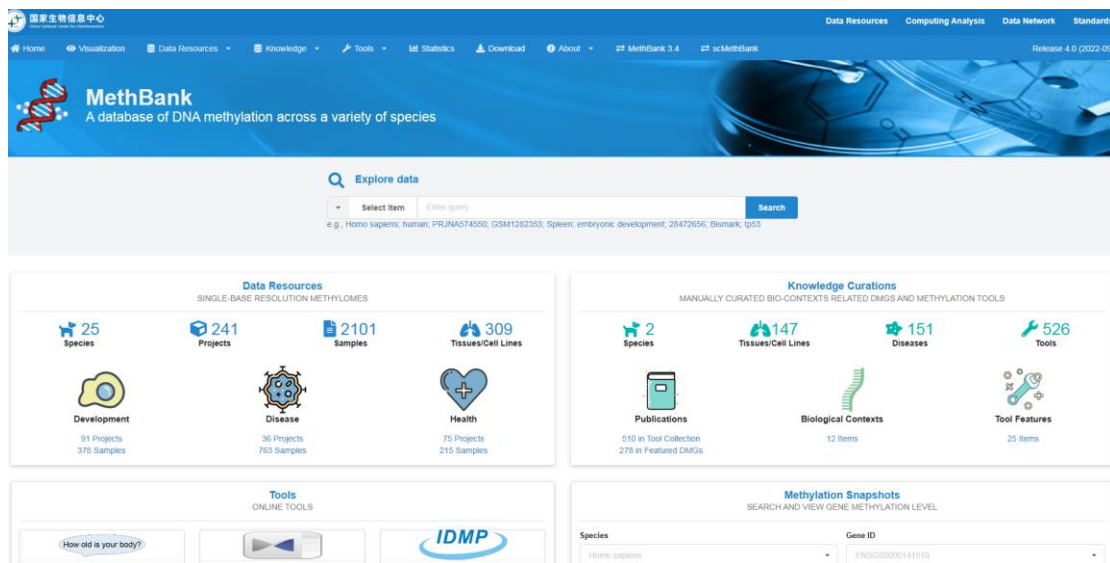
由中国科学院北京基因组研究所开发

功能 & 优点:

有效减弱或去除数据间的批次效应; 提供了高效的查询方式, 可协助检索和发现基于甲基化的生物标记物。

网站操作指南: https://mp.weixin.qq.com/s/aKu27_Cv4w4qYRZh5GvppA

2、MethBank (<http://bigd.big.ac.cn/methbank>)



数据内容:

各种物种的共有参考甲基化组 (CRM), 全基因组单碱基分辨率甲基化组 (SRM), DNA 和 RNA 甲基化工具 (MeTools) 和表观基因组范围关联研究 (EWAS) 相关信息。

34 个来自大量人类样品的共识参考甲基化组织, 336 个来自不同发育阶段 5 个植物组织的单碱基分辨率甲基化组织, 18 个单碱基在人类和小鼠的多个阶段从配子和早期胚胎中分辨甲基化组织。

由中国科学院北京基因组研究所开发。

功能 & 优点:

提供了甲基化数据可视化的交互式浏览器; 高质量甲基化酶大规模整合; 系统地鉴定与年龄密切相关的甲基化位点、在不同年龄段具有恒定甲基化水平的位点、年龄特异性差异甲基化胞嘧啶/区域和甲基化 CpG 岛; 在线估计人甲基化年

龄，鉴定差异甲基化的启动子；对于解读 DNA 甲基化调控机制有很大的帮助。

网站操作指南：<https://mp.weixin.qq.com/s/xootx8owCEdSi-cTUOa3eA>

八、其它甲基化数据库（RNA，组蛋白）

1、RNA 甲基化数据库

RMBase (<https://rna.sysu.edu.cn/rmbase/index.php>)

Welcome to RMBase v2.0 !

Latest version of RMBase (RMBase v3.0) is now available at
<https://rna.sysu.edu.cn/rmbase3/> or <http://bioinformaticsscience.cn/rmbase/>.

RMBase v2.0 is a comprehensive database to integrate epitranscriptome sequencing data for exploring post-transcriptionally modifications of RNAs, as well as their relationships with microRNA binding events, disease-related SNPs and RNA-binding proteins (RBP). RMBase v2.0 was expanded with 566 datasets and 1 397 244 modification sites from 47 studies among 13 species, which represented about 10 times expansion when compared to the previous release. It contains ~1 373 000 N6-Methyladenosines (m⁶A), ~5 400 N1-Methyladenosines (m¹A), ~9 600 pseudouridine (Ψ) modifications, ~1 000 5-methylcytosine (m⁵C) modifications, ~5 100 2'-O-methylations (2'-O-Me), and ~2 800 modifications of 100 other types.

We built two new modules called "Motif" and "modRBP", respectively. "Motif" provides both *de novo* identified position weight matrices (PWMs) and visualized logos of modification motifs. And "modRBP" analyzes the relationships of RNA modifications and RNA-binding Protein (RBP). We also constructed a new web interface, "modSNV", to present the disease link between RNA modification sites and SNVs sites. In addition, we developed a novel web-based tool named "modMetagene" for plotting metagenes of RNA modifications along a transcript model.

How to cite:
 RMBase v2.0: Deciphering the Map of RNA Modifications from Epitranscriptome Sequencing Data. Jia-Jia Xuan, Wen-Ju Sun, Ke-Ren Zhou, Shun Liu, Peng-Hui Lin, Ling-Ling Zheng, Liang-Hu Qu*, Jian-Hua Yang*.
 RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. Wen-Ju Sun, Jun-Hao Li, Shun Liu, Jie Wu, Liang-Hu Qu*, Jian-Hua Yang*. Nucleic Acids Res. 2016 Jan 4;44(D1):D259-65. Epub 2015 Oct 12.

Release notes
 Release version 2.0.0 2017-06-01

- Species: 13
- Sample: 566
- modification sites: 1,397,244
- Technologies: Pseudo-seq, Ψ-seq, CeU-seq, Aza-IP, MeRIP-seq, m⁶A-seq, m¹A-seq, miCLIP, m⁶A-CLIP, RiboMeth-seq
- Modification Types: over 100
- located ~5,400 m¹A (1-methyladenosine) modification sites by *de novo* analyzing from 4 species.
- de novo* identified ~1,373,000 m⁶A (N6-Methyladenosine) modifications across 13 species.

Modification Site Statistics in 13 Species (Genome Versions)

■ Human Mouse Rhesus Chimpanzee
 Rattus Pig Zebrafish A.thaliana
 S.cerevisiae Fly Paeruginosa E.coli
 S.pombe

Modification sites Statistics in 13 Species
 Total modification sites: 1,402,764

Pie chart data (approximate values):

Species	Modification Sites
Human	490704
Mouse	3320
Rhesus	59
Chimpanzee	5814
Rattus	6798
Pig	67671
Zebrafish	20331
A.thaliana	43027
S.cerevisiae	121409
Fly	60769
Paeruginosa	38369
E.coli	1052
S.pombe	97

数据内容：

13 个物种的 47 个研究中的 566 个数据集和 1 397 244 个修饰位点，m⁶A、m⁵C、2'-O-Me 和 100 种其他类型的修饰物。
 由中山大学开发。

功能 & 优点：

目前最全面的 RNA 修饰数据库；整合转录组测序数据，以探索 RNA 的转录后修饰，以基因为关键词，查询每个基因上鉴定到的化学修饰；查询与 miRNA 结合事件，疾病相关 SNP 和 RNA 结合蛋白（RBP）的关系。

2、RMVar (<https://rmvar.renlab.org/index.html>)

数据内容:

影响 **m6A** 修饰的相关变异, 与变异相关的 RBP 结合区、miRNA-靶点和剪接位点。

修饰物: m⁶A, m¹A, ψ , m⁵C, 2'-O-Me 和 m⁷G, 5-甲基尿苷 m⁵U, A 至 I。

来自全基因组关联研究(GWAS)和 ClinVar 的数据。

功能 & 优点:

帮助用户研究 m⁶A 相关变异对转录后调控的影响; 为研究 m⁶A 相关变异和疾病间的关系提供支持。

2、组蛋白甲基化数据库

ChIPBase (<https://rnasysu.com/chipbase3/index.php>)

数据内容:

ncRNAs 和 PCGs 的转录调控网络:

数千个结合 motif 矩阵及其结合位点; 约为 10000 个肿瘤样本和约为 9100 个正常组织和细胞系的基因表达谱

由中山大学开发

功能 & 优点:

预测转录因子(TFs)与基因之间的转录调控关系。

预测涉及或影响 ncRNAs 和 PCGs 转录的 TFs 和组蛋白修饰。

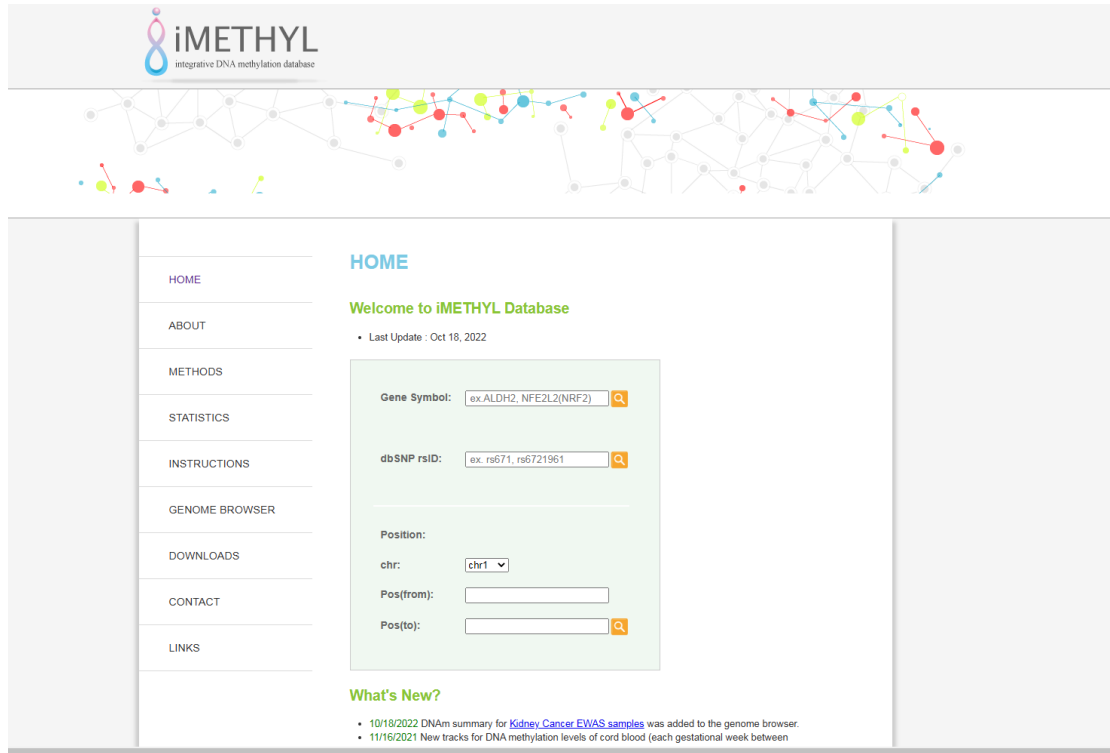
探索 DNA 结合蛋白和各种类型基因之间的共表达模式。

预测不同基因的功能和可视化各种芯片 seq 数据。

网站操作指南: <https://mp.weixin.qq.com/s/7cwpSZwIVrClIGxkIIYlhA>

3、多组学联合数据库

iMETHYL (<http://imethyl.iwate-megabank.org/index.html>)



数据内容:

DNA 甲基化、 SNP 和 RNA_seq 的多组学联合数据库。

CD4 + T 淋巴细胞, 单核细胞和嗜中性粒细胞的全 DNA 甲基化 (~2400 万个常染色体 CpG 位点), 全基因组 (~900 万个单核苷酸变体) 和全转录组 (> 14000 个基因) 数据。

功能 & 优点:

两两之间的关联分析, 推断 DNA 甲基化, 基因组变异和基因表达之间的调控机制。