

人類基因開關預測及 生醫文件檢索系統製作

指導老師：

資訊管理學系 柯淑津老師

微生物學系 楊鉅文老師

組員：

資四A 96156147

資四B 96156204

資四B 96156209

黃羿綺

林耿誠

陳吳嘉

大綱

- 動機
- 背景
- 基因開關預測
- 系統製作
- Demo

動機

- 鴻海搶攻頂級消費群市場
 - 客製化醫療
 - 從個人DNA著手

「砸2億搶生醫 鴻海5星健檢中心曝光 (2010/09/13 21:43)翁錦煖 / 綜合報導

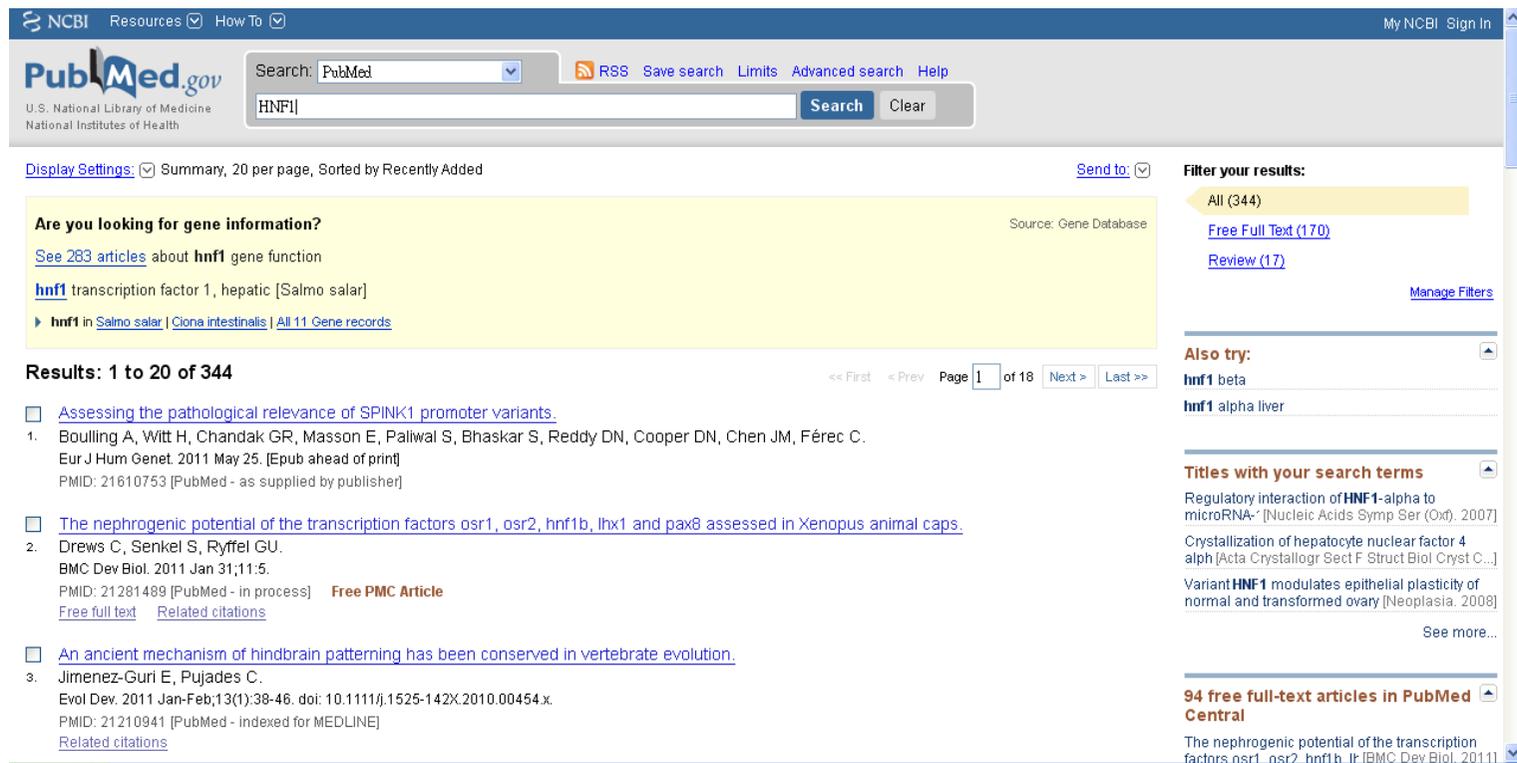
電子一哥鴻海積極跨入健康管理事業，砸下上億元打造的頂級健檢中心，首次在媒體前曝光！裡頭裝潢擺設以橘黃色為基調，走進大廳，彷彿來到五星級飯店一樣高檔。裡頭還有一項檢測項目是全台獨一無二，就是"個人基因檢測"，把原本單提供研究用的DNA排序分析，搬到醫療上，等於能讓人知道自己先天的疾病隱憂。」

背景

- 現有國外生物資訊工具
 - PubMed
 - PubGene
 - JASPAR
 - Weblogo

PubMed

• 生物醫學文獻資料庫



The screenshot displays the PubMed website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' links. The main search area features the 'PubMed.gov' logo and a search box containing 'HNF1'. Below the search box, there are links for 'RSS', 'Save search', 'Limits', 'Advanced search', and 'Help'. The search results are displayed in a list format, with the first three results visible. The first result is 'Assessing the pathological relevance of SPINK1 promoter variants' by Boulling A, Witt H, Chandak GR, Masson E, Paliwal S, Bhaskar S, Reddy DN, Cooper DN, Chen JM, Férec C. The second result is 'The nephrogenic potential of the transcription factors *osr1*, *osr2*, *hnf1b*, *lhx1* and *pax8* assessed in *Xenopus* animal caps' by Drews C, Senkel S, Ryffel GU. The third result is 'An ancient mechanism of hindbrain patterning has been conserved in vertebrate evolution' by Jimenez-Guri E, Pujades C. The right side of the page shows a 'Filter your results' section with 'All (344)' selected, and a 'Titles with your search terms' section listing several articles related to HNF1.

NCBI Resources How To My NCBI Sign In

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed
HNF1

RSS Save search Limits Advanced search Help

Search Clear

Display Settings: Summary, 20 per page, Sorted by Recently Added Send to:

Are you looking for gene information?
See 283 articles about **hnf1** gene function
hnf1 transcription factor 1, hepatic [*Salmo salar*]
▶ **hnf1** in [Salmo salar](#) | [Clona intestinalis](#) | [All 11 Gene records](#)

Source: Gene Database

Results: 1 to 20 of 344

1. [Assessing the pathological relevance of SPINK1 promoter variants.](#)
Boulling A, Witt H, Chandak GR, Masson E, Paliwal S, Bhaskar S, Reddy DN, Cooper DN, Chen JM, Férec C.
Eur J Hum Genet. 2011 May 25. [Epub ahead of print]
PMID: 21610753 [PubMed - as supplied by publisher]

2. [The nephrogenic potential of the transcription factors *osr1*, *osr2*, *hnf1b*, *lhx1* and *pax8* assessed in *Xenopus* animal caps.](#)
Drews C, Senkel S, Ryffel GU.
BMC Dev Biol. 2011 Jan 31;11:5.
PMID: 21281499 [PubMed - in process] **Free PMC Article**
[Free full text](#) [Related citations](#)

3. [An ancient mechanism of hindbrain patterning has been conserved in vertebrate evolution.](#)
Jimenez-Guri E, Pujades C.
Evol Dev. 2011 Jan-Feb;13(1):38-46. doi: 10.1111/j.1525-142X.2010.00454.x.
PMID: 21210941 [PubMed - indexed for MEDLINE]
[Related citations](#)

Filter your results:
All (344)
[Free Full Text \(170\)](#)
[Review \(17\)](#)
[Manage Filters](#)

Also try:
hnf1 beta
hnf1 alpha liver

Titles with your search terms
Regulatory interaction of **HNF1**-alpha to microRNA- [Nucleic Acids Symp Ser (Oxf). 2007]
Crystallization of hepatocyte nuclear factor 4 alpha [Acta Crystallogr Sect F Struct Biol Cryst C...]
Variant **HNF1** modulates epithelial plasticity of normal and transformed ovary [Neoplasia. 2008]
[See more...](#)

94 free full-text articles in PubMed Central
The nephrogenic potential of the transcription factors *osr1*, *osr2*, *hnf1b*, *lhx1*. [BMC Dev Biol. 2011]

PubGene

- 以圖形網絡介面呈現基因關係

The screenshot displays the PubGene interface for the BRCA1 gene. The top navigation bar includes 'Bio Networks', 'Bio Associations', 'Sequence Homology', and 'My Workspace Login'. The main search area has fields for 'Organism' (set to Homo sapiens), 'Gene / Protein' (BRCA1), and 'Biological term'. Below the search area, a network view is shown with the title 'Network based on co-occurrence in article for the "BRCA1" Gene and Proteins in Homo sapiens'. The network view shows a central node connected to several other nodes, representing gene relationships. To the left of the network view is a table with 'Property' and 'Value' columns, listing various attributes of the BRCA1 gene. Below the table are links for 'Full Summary Report', 'Most recent literature', 'Sequence Homology', 'Literature Neighbours', 'GO associations', and 'MeSH associations'.

Property	Value
Term	BRCA1
Name	breast cancer 1, earl...
Organi...	Homo sapiens
Article ...	937
Term T...	PubGene Gene/Prot...
Externa...	EG
Chrom...	
Alt. Sym...	BRCA1
Alt. Sym...	BRCC1
Alt. Sym...	IRIS
Alt. Sym...	PSCP
Alt. Sym...	RNF53

JASPAR

- 哺乳類轉錄因子結合位(TFBS)的PWM資料庫網站工具

The screenshot displays the JASPAR website interface. At the top, it states "You are using the JASPAR server: jaspar.genereg.net". Below this is a large logo for "AGJASPAR" where each letter's height represents its information content in bits. A red starburst graphic next to the logo says "new and improved". The x-axis of the logo is labeled 1 through 10, and the y-axis is labeled "bits" with values 1 and 2. Below the logo, it reads "The high-quality transcription factor binding profile database".

Below the logo, there is a section titled "Browse the JASPAR CORE database directly:" with six navigation buttons:

- JASPAR CORE Vertebrata (with a fish icon)
- JASPAR CORE Nematoda (with a nematode icon)
- JASPAR CORE Insecta (with a fly icon)
- JASPAR CORE Plantae (with a fern icon)
- JASPAR CORE Fungi (with a mushroom icon)
- JASPAR CORE by Structural Class (with a protein structure icon)

Weblogo

- 圖形表示DNA多序列

WEBLOGO

· [about](#) · [create](#) · [examples](#) ·

Version 2.8.2 (2005-09-08)

(= [WebLogo 3](#))

References

Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. *Genome Research*, 14:1188-1190, (2004) [Full Text]

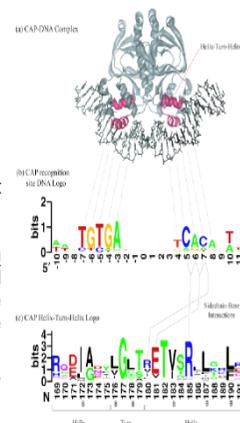
Schneider TD, Stephens RM. 1990. [Sequence Logos: A New Way to Display Consensus Sequences](#). *Nucleic Acids Res.* 18:6097-6100

Introduction

[WebLogo](#) is a web based application designed to make the [generation](#) of sequence logos as easy and painless as possible. Click [here](#) to create your own sequence logos.

[Sequence logos](#) are a graphical representation of an amino acid or nucleic acid multiple sequence alignment developed by [Tom Schneider](#) and [Mike Stephens](#). Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of, for example, a binding site, than would a consensus sequence.

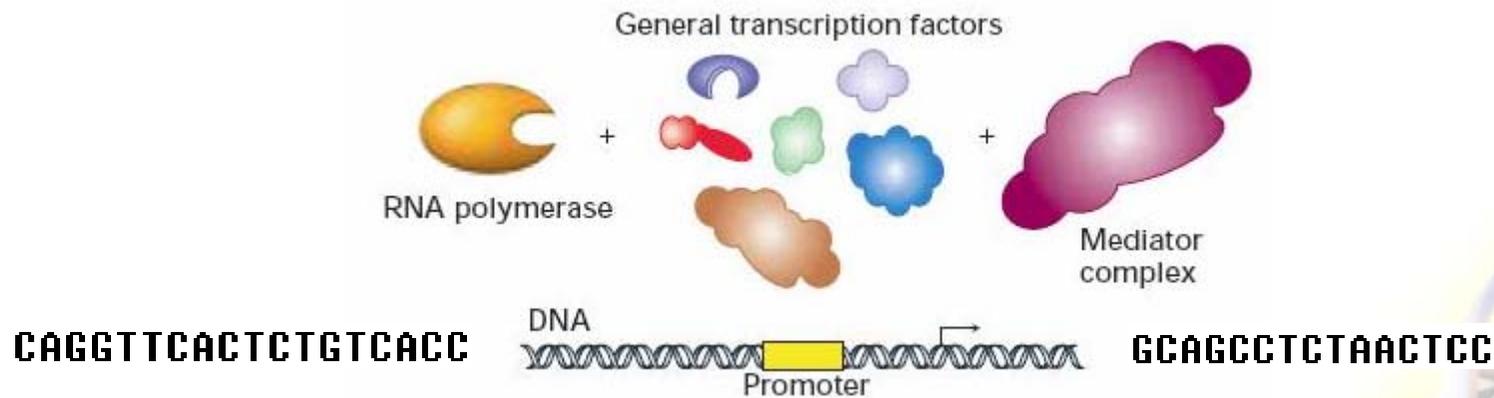
[WebLogo Source Code](#)



```
CREB1_1 ATGCGTCA
CREB1_2 TGACGTCC
CREB1_3 TGACGTCA
CREB1_4 TGACGTCA
CREB1_5 CCGCGTCA
CREB1_6 TGACGTCA
CREB1_7 TGACGTCA
CREB1_8 TGACCTCA
CREB1_9 TTGCGTCA
CREB1_10 TGACGTCT
CREB1_11 TGACGTCA
```

基因開關預測

- 基因轉錄調控(transcription regulation)
 - 嬰兒胚胎發育、成長階段細胞生長、細胞分化到生活中細胞對環境刺激的反應
 - 細胞內最重要核心調控機制
 - 基因轉錄調控主要依賴
 - 轉錄因子(transcription factor, TF)
 - 特別具專一性的DNA片段



基因開關預測

- PWM (Position Weight Matrix)

DNA sequences matrix

Site 1	G	A	C	C	A	A	A	T	A	A	G	G	C	A	
Site 2	G	A	C	C	A	A	A	T	A	A	G	G	C	A	
Site 3	T	G	A	C	T	A	T	A	A	A	A	G	G	A	
Site 4	T	G	A	C	T	A	T	A	A	A	A	G	G	A	
Site 5	T	G	C	C	A	A	A	A	G	T	G	G	T	C	
Site 6	C	A	A	C	T	A	T	C	T	T	G	G	G	C	
Site 7	C	A	A	C	T	A	T	C	T	T	G	G	G	C	
Site 8	C	T	C	C	T	T	A	C	A	T	G	G	G	C	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14

Source binding sites

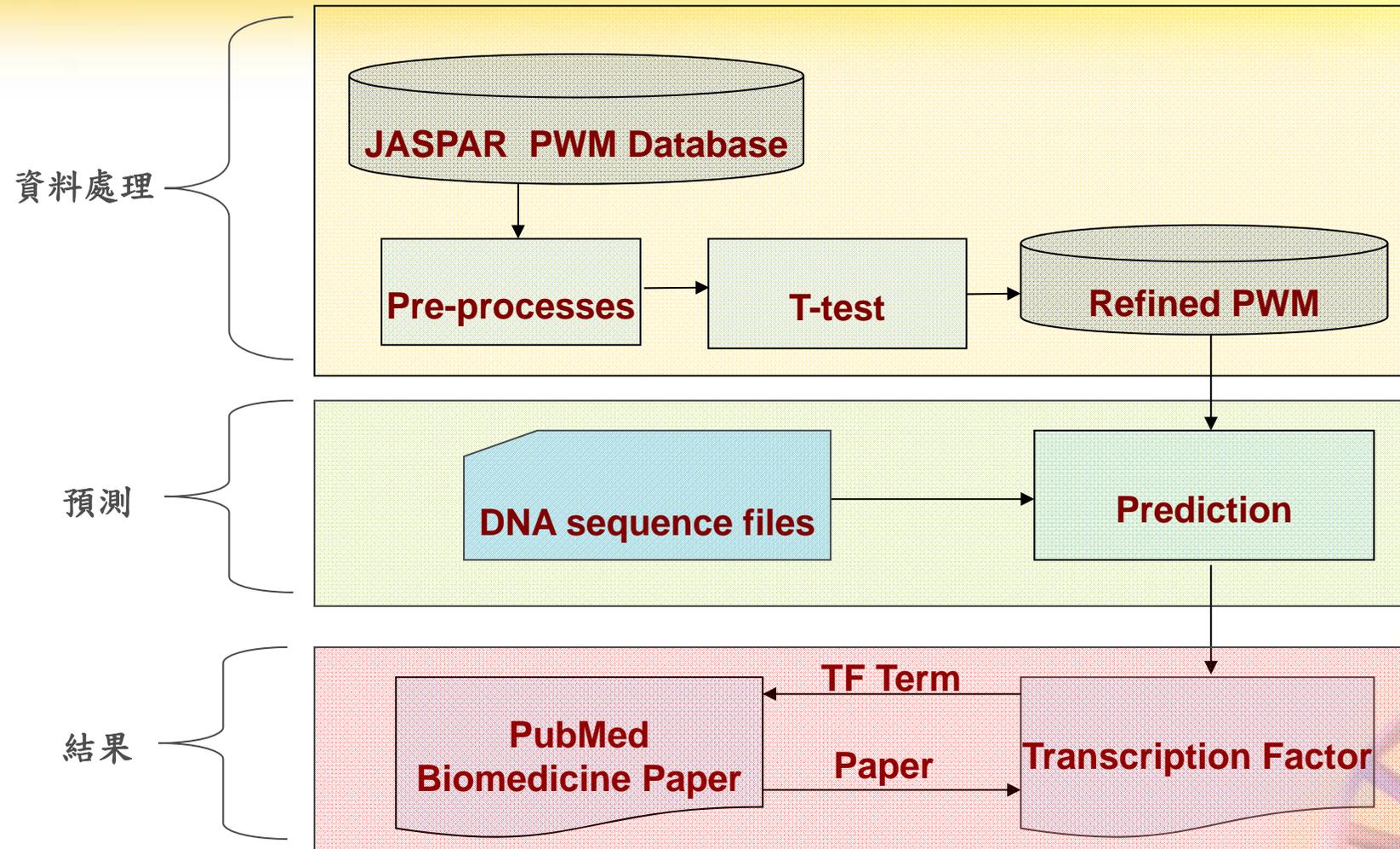
Position frequency matrix (PFM)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	0	4	4	0	3	7	4	3	5	4	2	0	0	4
C	3	0	4	8	0	0	0	3	0	0	0	0	2	4
G	2	3	0	0	0	0	0	0	1	0	6	8	5	0
T	3	1	0	0	5	1	4	2	2	4	0	0	1	0

Position weight matrix (PWM)

A	-1.93	0.79	0.79	-1.93	0.45	1.50	0.79	0.45	1.07	0.79	0.00	-1.93	-1.93	0.79
C	0.45	-1.93	0.79	1.68	-1.93	-1.93	-1.93	0.45	-1.93	-1.93	-1.93	-1.93	0.00	0.79
G	0.00	0.45	-1.93	-1.93	-1.93	-1.93	-1.93	-1.93	0.66	-1.93	1.30	1.68	1.07	-1.93
T	0.15	0.66	-1.93	-1.93	1.07	0.66	0.79	0.00	0.00	0.79	-1.93	-1.93	-0.66	-1.93

系統流程圖



系統製作

- Pre-processes
 - 去除品質低的序列。

```
ARNT_1 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT CACGTG 0
ARNT_2 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT CACGTG 0
ARNT_3 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT CACGTG 0
ARNT_4 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT CACGTG 0
ARNT_5 MA0004.1.sites_ARNT:-193-222 MA0004.1.sites_ARNT CACGTG 0
ARNT_6 MA0004.1.sites_ARNT:-193-222 MA0004.1.sites_ARNT CACGTG 0
ARNT_7 MA0004.1.sites_ARNT:-197-218 MA0004.1.sites_ARNT CACGTG 0
ARNT_8 MA0004.1.sites_ARNT:-195-220 MA0004.1.sites_ARNT CACGTG 0
ARNT_9 MA0004.1.sites_ARNT:-196-219 MA0004.1.sites_ARNT CACGTG 0
ARNT_10 MA0004.1.sites_ARNT:-195-220 MA0004.1.sites_ARNT CACGTG 0
ARNT_11 MA0004.1.sites_ARNT:-194-221 MA0004.1.sites_ARNT CACGTG 0
ARNT_12 MA0004.1.sites_ARNT:-196-219 MA0004.1.sites_ARNT CACGTG 0
ARNT_13 MA0004.1.sites_ARNT:-193-222 MA0004.1.sites_ARNT CACGTG 0
ARNT_14 MA0004.1.sites_ARNT:-197-218 MA0004.1.sites_ARNT CACGTG 0
ARNT_15 MA0004.1.sites_ARNT:-194-221 MA0004.1.sites_ARNT CACGTG 0
ARNT_16 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT AACGTG 2.69675054227128
ARNT_17 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT AACGTG 2.69675054227128
ARNT_18 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT AACGTG 2.69675054227128
ARNT_19 MA0004.1.sites_ARNT:-200-215 MA0004.1.sites_ARNT AACGTG 2.69675054227128
ARNT_20 MA0004.1.sites_ARNT:-194-221 MA0004.1.sites_ARNT CGCGTG 8.16194612644868
```

系統製作

- T-test
 - 預測出每個矩陣分別適合幾個字的預測方法。

mon_New_Sci_MA0004.1.sites			rand_mon_New_Sci_MA0004.1.sites_ARNT_hhh6		
檔案(F)	編輯(E)	格式(O)	檢視(V)	說明(H)	
ARNT_1	CACGTG	0	S0	GGTTA	26.2915967653797
ARNT_2	CACGTG	0	S1	TTACGA	28.8461278467417
ARNT_3	CACGTG	0	S2	TATGCG	17.5698805841205
ARNT_4	CACGTG	0	S3	CCTACG	21.39825770366
ARNT_5	CACGTG	0	S4	TCTAGC	28.8461278467417
ARNT_6	CACGTG	0	S5	TGCGAT	20.5011180562956
ARNT_7	CACGTG	0	S6	TCGAGG	26.2953300739461
ARNT_8	CACGTG	0	S7	TGACTT	23.6890011857745
ARNT_9	CACGTG	0	S8	GTAGCC	21.1442318780754
ARNT_10	CACGTG	0	S9	GCACGT	26.4956564073599
ARNT_11	CACGTG	0	S10	TATCTG	20.3030912534757
ARNT_12	CACGTG	0	S11	GGCGTG	16.3972371307432
ARNT_13	CACGTG	0	S12	TCTTAT	26.4956564073599
ARNT_14	CACGTG	0	S13	TGCATG	17.5275248559453

Ttest_MA0004.1.sites_ARNT_hhh6hhh.txt.txt - 記事本				
檔案(F)	編輯(E)	格式(O)	檢視(V)	說明(H)
mon	50.8450638531741			
di	69.1986140915567			
tri	89.5276945418501			
tet	103.22468148513			
pen	103.24690628183			

result.txt - 記事本				
檔案(F)	編輯(E)	格式(O)	檢視(V)	說明(H)
MA0004.1.sites	ARNT	pen	5	6
MA0006.1.sites	Ahr-ARNT	tet	4	
MA0007.1.sites	Androgen	pen	5	
MA0009.1.sites	T	tet	4	11
MA0014.1.sites	Pax5	pen	5	20
MA0017.1.sites	NR2F1	pen	5	14
MA0018.2.sites	CREB1	mon	5	8
MA0019.1.sites	Ddit3-Cebpa	pen	5	
MA0024.1.sites	E2F1	pen	5	8

Demo

- 人類基因開關預測及生醫文件檢索系統

人類基因開關預測系統製作及生醫文件檢索

Prediction

Mammalian Estrogen Receptor Binding Site Prediction

- [Paste the DNA sequence](#)
- [Upload files](#)

