

Gene-Based Repetitive Element Database (GBRED)

國立成功大學生物資訊研究所：梁凱強

國立成功大學分子醫學研究所：孫孝芳

What is the HGP?



What We've Learned Thus Far

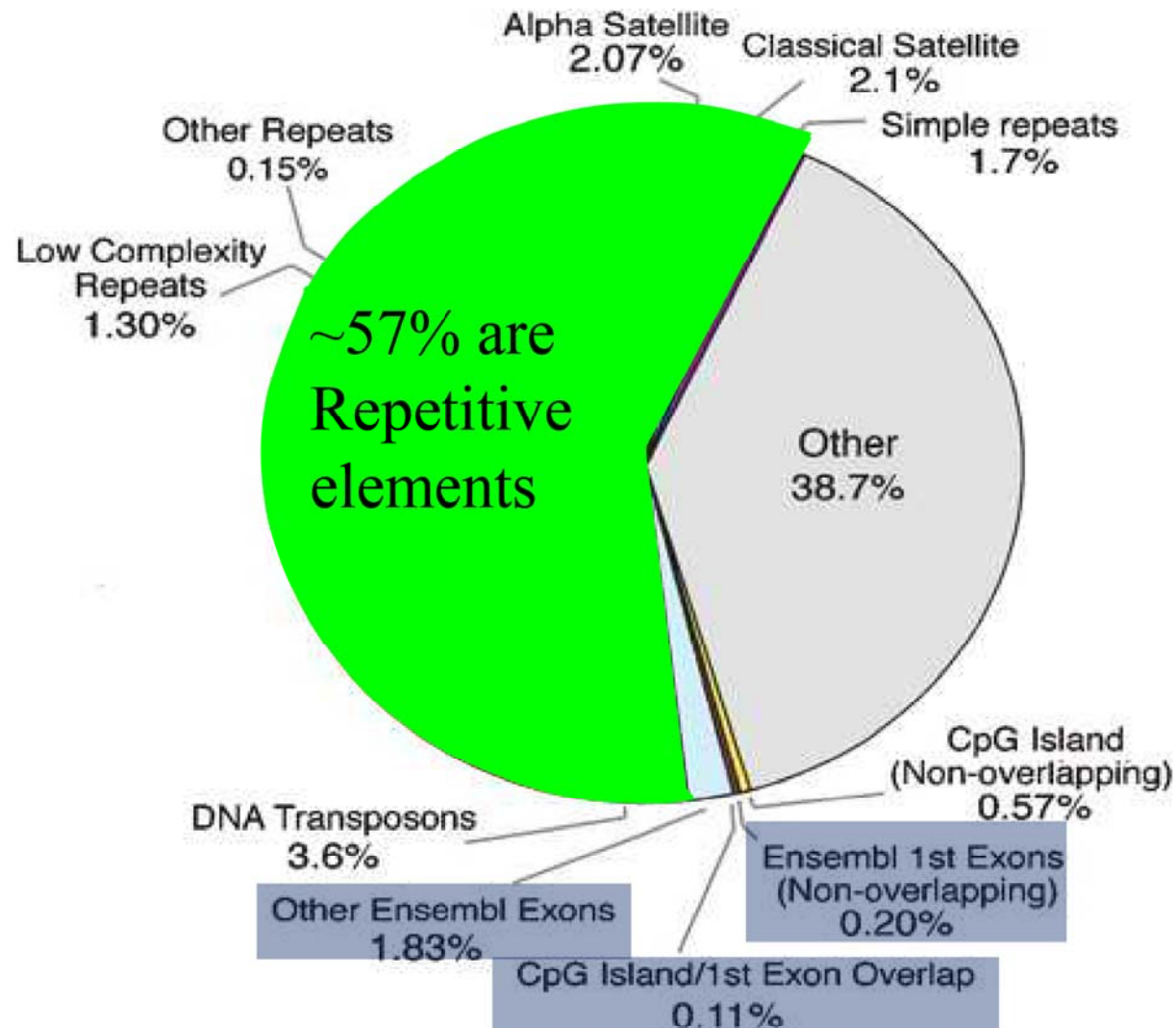
“the more we learn about the human genome, the more there is to explore.”

Repetitive elements

Class	Length	Copy Number in Human Genome
Repetitious DNA		
Simple-sequence DNA	1-500 bp	Variable
Interspersed repeats		
DNA transposons	2-3 kp	300,000
LTR retrotransposons	6-11 kp	440,000
Non-LTR retrotransposons		
LINEs	6-8 kp	860,000
SINEs	100-300 bp	1,600,000

E. S. Lander *et al.*, *Nature*. 409: 860-921 (2001)

Human genome composition



Functional roles of repetitive elements

- Repetitive elements are involved in genome rearrangement and illegitimate recombination

Kreahling et al., Trends Genet., 20, 1-4 (2004)

- Serve as regulatory elements controlling the transcription of nearby genes

- By acting as enhancers or promoters

- Provide alternative splicing sites and polyadenylation tail

Kondo, Y. et al., J. Biol. Chem. 278: 27658-27662 (2003)

- Via epigenetic mechanism

Makałowski, Gene. 259: 61-67 (2000)

- By carry the transcription factor binding site

Eriko A. et al., Nucleic Acids Research. 35: 3383-3390 (2007)

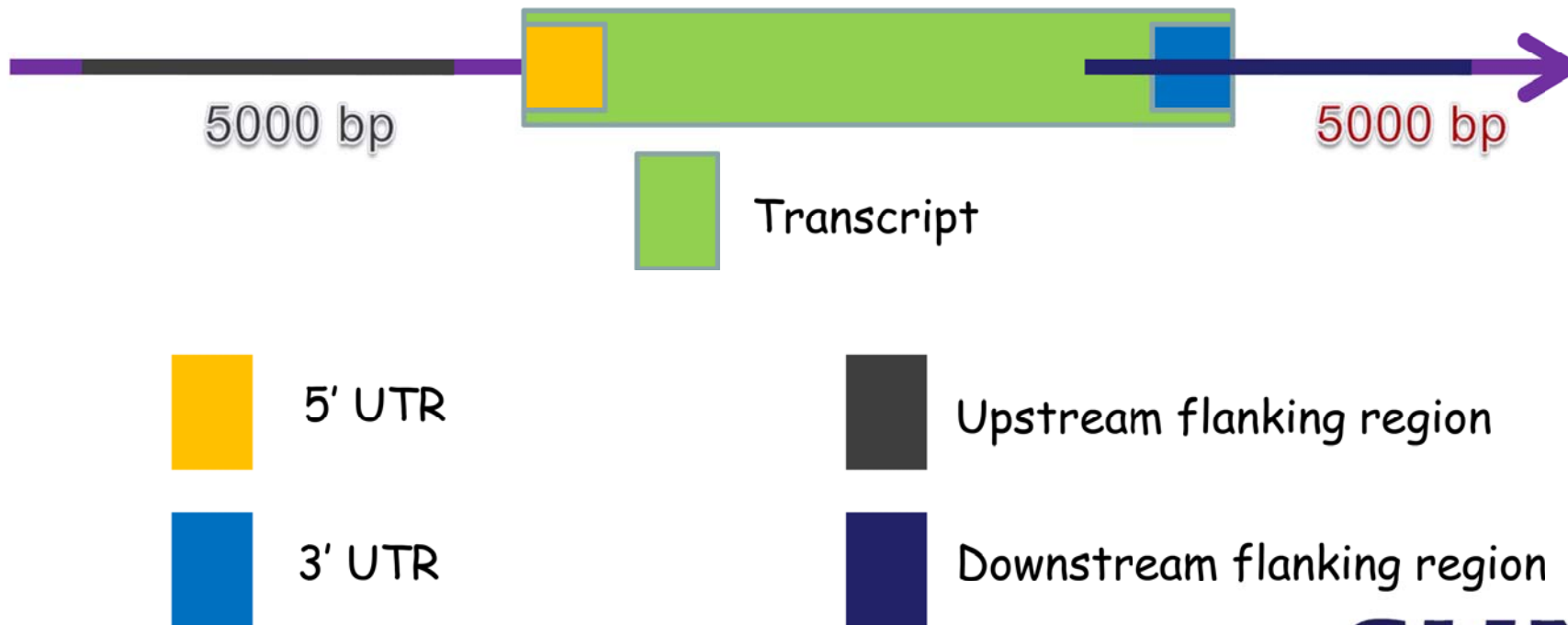
Ting W. et al., Proc Natl Acad Sci USA . 104: 18613-18618 (2007)

Objective

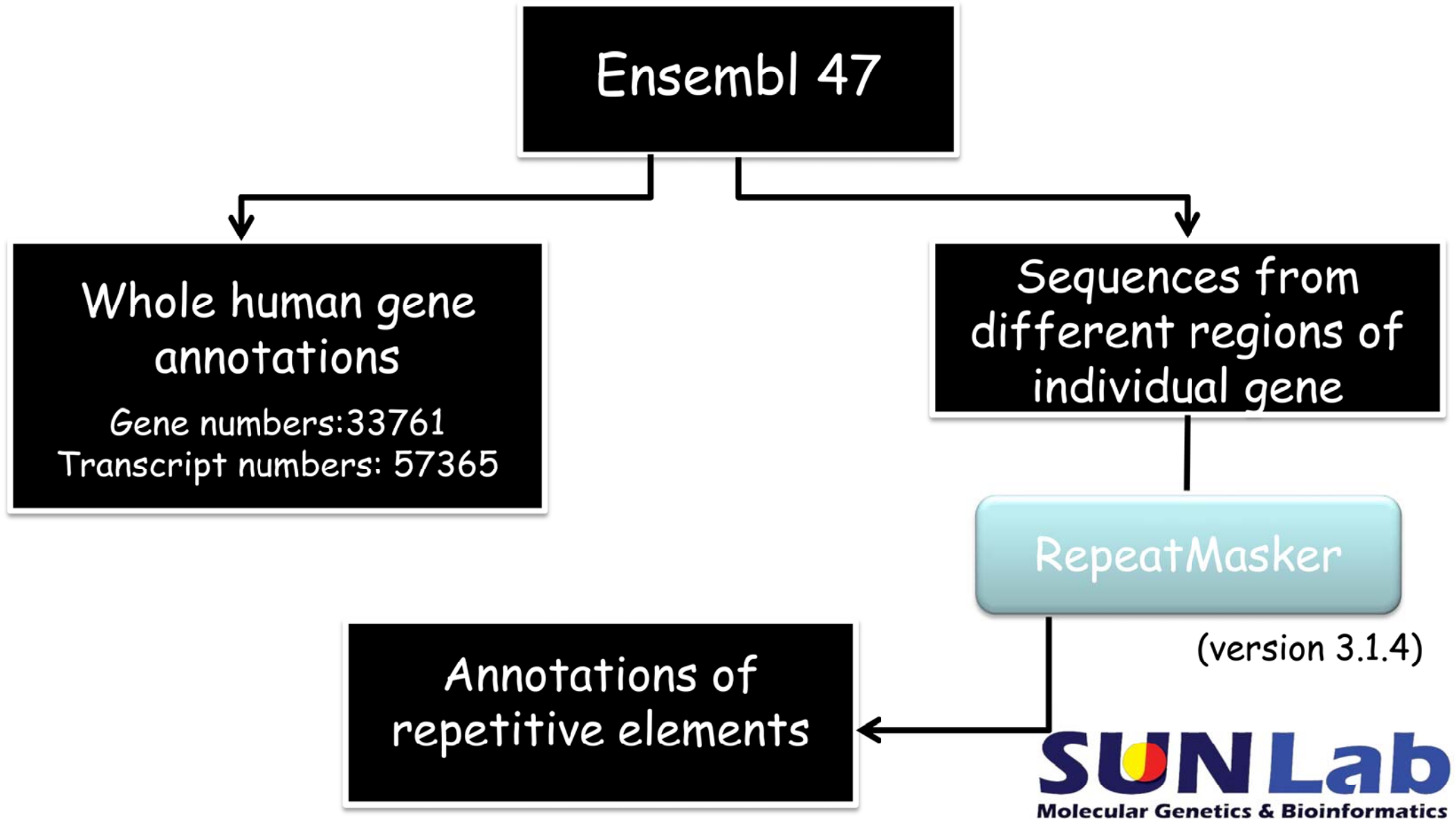
To construct a gene-based repetitive element database to provide all annotated repetitive element information that are associated with particular gene(s)

Data sources

- All sequences and annotated information of human genes were downloaded from Ensembl build version 47



Database construction



NCKU GBRED 首頁

<http://www.binfo.ncku.edu.tw/GBRED/>

The screenshot shows the GBRED website homepage. At the top, there is a header with the 'Binfo' logo and 'NCKU Bioinformatics Center' text. Below this is the main title 'GBRED' in large, stylized letters, followed by the subtitle 'Gene-Based Repetitive Element Database'. The page is divided into two main columns. The left column contains a navigation menu with links for 'Homepage', 'Query', 'Manual', and 'RepeatMasker'. Below this is a 'CONTACT US' section with an 'E-mail' link, and a 'VERSION' section listing 'RepeatMasker version: 3.1.4', 'Rebase version: 20070801', and 'Last updated: 20080928'. The right column features an 'INTRODUCTION' section dated 'September 22, 2008', which describes GBRED as a database of repetitive elements flanking human genes. Callout boxes are overlaid on the image, pointing to specific elements: '進入搜尋頁面' (Enter search page) points to the 'Query' link; '進入說明頁面' (Enter introduction page) points to the 'Manual' link; '進入成大RepeatMasker伺服器' (Enter NCKU RepeatMasker server) points to the 'RepeatMasker' link; '寄信給成大生物資訊中心' (Email NCKU Bioinformatics Center) points to the 'E-mail' link; 'GBRED 簡介' (GBRED Introduction) points to the 'INTRODUCTION' text; and '版本及更新日期' (Version and update date) points to the 'VERSION' section.

進入搜尋頁面

GBRED 簡介

進入說明頁面

進入成大RepeatMasker
伺服器

版本及更新日期

寄信給成大生物資訊中心

GBRED 搜尋介面

選擇搜尋特定區域的
重複性序列

GBRED

Homepage
Query
■ Manual
■ RepeatMasker

CONTACT US

■ E-mail

VERSION

RepeatMasker version: 3.1.4
Rebase version : 20070801
Last updated : 20080928

QUERY

Region :
5UTR
3UTR
Flank_u5000
Flank_d5000
All

RepeatClass :

Subclass :

Gene Symbol :

Chromosome (Default : All)

1 2 3 4 5 6 7 8
 9 10 11 12 13 14 15 16
 17 18 19 20 21 22 X Y All

UTR means untranslated region;
Flank_u5000 means 5kb upstream
from transcription start site;
Flank_d5000 means 5kb downstream
from transcription end site.

GBRED 搜尋介面

GBRED QUERY September 30, 2008

Homepage
Query
■ Manual
■ RepeatMasker

CONTACT US
■ E-mail

VERSION
RepeatMasker version: 3.1.4
Rebase version : 20070801
Last updated : 20080928

Region :

RepeatClass : Select main class of repetitive element

Subclass :

Gene Symbol :

Chromosome (Default):

1 2 3 4 5 6 7 8
 9 10 11 12 13 14 15 16
 17 18 19 20 21 22 X Y All

選擇特定種類的重複性序列

GBRED 搜尋介面

GBRED QUERY September 30, 2008

Homepage
Query
■ Manual
■ RepeatMasker

CONTACT US
■ E-mail

VERSION
RepeatMasker version: 3.1.4
Rebase version : 20070801
Last updated : 20080928

Region :

RepeatClass :

Subclass : Select subclass of repetitive element which

- All
- DNA
- DNA/AcHobo
- DNA/hAT
- DNA/Mariner
- DNA/MER1_type
- DNA/MER1_type?
- DNA/MER2_type
- DNA/Merlin
- DNA/MuDR
- DNA/PiggyBac
- DNA/Tc2
- DNA/Tip100

Gene Symbol

Chromosome

1 2 3 4 5 6 7 8

9 10 11 12 13 14 15 16

17 18 19 20 21 22 X Y All

選擇細部種類的重複性序列

GBRED 搜尋介面

GBRED QUERY September 30, 2008

Homepage
Query
■ Manual
■ RepeatMasker

CONTACT US
■ E-mail

VERSION
RepeatMasker version: 3.1.4
Rebase version : 20070801
Last updated : 20080928

Region :

RepeatClass :

Subclass :

Gene Symbol : Enter the gene symbol to search relative repetitive element information. You can use the first few letters followed by * for a wildcard search. e.g.fgf*.

Chromosome (Default : All)

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1	2	3	4	5	6	7	8		
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	10	11	12	13	14	15	16		
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17	18	19	20	21	22	X	Y	All	

輸入基因的名稱，並
可利用*當作萬用字
元來進行檢索

GBRED 搜尋介面

GBRED QUERY September 30, 2008

Homepage
Query
■ Manual
■ RepeatMasker

CONTACT US
■ E-mail

VERSION
RepeatMasker version: 3.1.4
Rebase version : 20070801
Last updated : 20080928

Region :

RepeatClass :

Subclass :

Gene Symbol :

Chromosome (Default : All)

1 2 3 4 5 6 7 8
 9 10 11 12 13 14 15 16
 17 18 19 20 21 22 X Y All

選擇要搜尋的
染色體

GBRED 搜尋介面

(範例)

GBRED

Homepage
Query
■ Manual
■ RepeatMasker

CONTACT US

■ E-mail

VERSION

RepeatMasker version: 3.1.4
Rebase version : 20070801
Last updated : 20080928

QUERY September 30, 2008

Region :

RepeatClass :

Subclass :

Gene Symbol :

Chromosome (Default : All)

1 2 3 4 5 6 7 8

9 10 11 12 13 14 15 16

17 18 19 20 21 22 X Y All

選擇3UTR

選擇LINEs

選擇LINE/CR1

以染色體9為例

GBRED 搜尋結果

前一頁所設定的搜尋條件

You have selected

- chromosome 09
- gene name is *
- region is 3UTR
- selected repeat is LINE/CR1

[Manual page](#)

基因正式縮寫

Search Result

Gene_ID	Gene_Symbol	Description
ENSG0000023318	TXNDC4	Thioredoxin domain-containing protein 4 precursor (Endoplasmic reticulum resident protein ERp44). [Source:Uniprot/SWISSPROT;Acc:Q9BS26]
ENSG00000095203	EPB41L4B	Band 4.1-like protein 4B (Protein EHM2) (FERM-containing protein CG1). [Source:Uniprot/SWISSPROT;Acc:Q9H329]
ENSG00000137038	C9orf123	Uncharacterized protein C9orf123. [Source:Uniprot/SWISSPROT;Acc:Q96GE9]
ENSG00000137073	UBAP2	Ubiquitin-associated protein 2. [Source:Uniprot/SWISSPROT;Acc:Q5T6F2]
ENSG00000137135	C9orf100	VAV-like protein C9orf100. [Source:Uniprot/SWISSPROT;Acc:Q8N4T4]
ENSG00000148053	NTRK2	BDNF/NT-3 growth factors receptor precursor (EC 2.7.10.1) (Neurotrophic tyrosine kinase receptor type 2) (TrkB tyrosine kinase) (GP145-TrkB) (Trk-B). [Source:Uniprot/SWISSPROT;Acc:Q16620]
ENSG00000148153	C9orf80	Uncharacterized protein C9orf80. [Source:Uniprot/SWISSPROT;Acc:Q9NRY2]
ENSG00000180471	C9orf31	Putative uncharacterized protein C9orf31 (Protein MOST-2). [Source:Uniprot/SWISSPROT;Acc:Q9NRJ2]
ENSG00000203987		CDNA FLJ46557 fis, clone THYMU3039846. [Source:Uniprot/SPTREMBL;Acc:Q6ZR86]

Page 1 Pages 1 (Total Data 9)

[New Query](#)

Total: 9
Page: 1 / 1

Ensembl所提供的基因編號，點擊以取得詳細資料

基因的全名及其來源

GBRED 搜尋結果

若搜尋條件中沒有選擇特定染色體，則結果會列出各基因所在的染色體

[Manual page](#)

列出所在的染色體

連結到說明網頁

[Next](#)

Total: 144
Page: 1 / 5

You have selected

- gene name is *
- region is 3UTR
- selected repeat is LINE/CRT

Search Result

Chromosome	Gene_ID	Gene_Symbol	Description
1	ENSG00000006555	TTC22	Tetratricopeptide repeat protein 22 (TPR repeat protein 22). [Source:Uniprot/SWISSPROT;Acc:Q5TAA0]
1	ENSG00000009790	TRAF3IP3	TRAF3-interacting JNK-activating modulator (TRAF3-interacting protein 3). [Source:Uniprot/SWISSPROT;Acc:Q9Y228]
1	ENSG00000057252	SOAT1	Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1). [Source:Uniprot/SWISSPROT;Acc:P35610]
1	ENSG00000077157	PPP1R12B	Protein phosphatase 1 regulatory subunit 12B (Myosin phosphatase-targeting subunit 2) (Myosin phosphatase target subunit 2). [Source:Uniprot/SWISSPROT;Acc:O60237]
1	ENSG00000097033	SH3GLB1	SH3 domain GRB2-like protein B1 (EC 2.3.1.-) (Endophilin-B1) (Bax-interacting factor 1) (Bif-1). [Source:Uniprot/SWISSPROT;Acc:Q9Y371]
1	ENSG00000117597	C1orf107	Digestive organ expansion factor homolog. [Source:Uniprot/SWISSPROT;Acc:Q68CQ4]
1	ENSG00000118473	SGIP1	SH3-containing GRB2-like protein 3-interacting protein 1 (Endophilin-3-interacting protein). [Source:Uniprot/SWISSPROT;Acc:Q9BQI5]
1	ENSG00000133065	SLC41A1	solute carrier family 41 member 1 [Source:RefSeq_peptide;Acc:NP_776253]

Ensembl所提供的基因編號，點擊以取得詳細資料

在此以TTC22基因為例子，點擊基因編號以取得基因的詳細資料

基因詳細資訊

Ensembl基因編號，可連結至Ensembl網站

基因正式縮寫

基因的全名及其來源

Gene Content

Gene ID [ENSG00000006555](#)

Gene Symbol TTC22

Description Tetratricopeptide repeat protein 22 (TPR repeat protein 22). [Source:Uniprot/SWISSPROT;Acc:Q5TAA0]

Strand -1

Biotype protein_coding

Gene Location Chromosome 01 location 55019504-55039459

Transcript

Transcript ID	Transcript start	Transcript end
ENST00000371274	55022813	55039459

此基因位於染色體的正股或是負股

Please select the transcript ID for further information.

[New Query](#)

基因功能

此基因所具有的transcripts，
點擊可顯示重複性序列的詳細
資料

基因在染色體上的位置

重複性序列詳細資訊

重複性序列所在的區域以及此區域長度

所選擇的transcript ID

Transcript	Transcript ID	Transcript start	Transcript end
	ENST00000371274	55022813	55039459

Graph Table - [ENST00000371274](#)

不同的色塊代表不同種類的重複性序列

Legend - Description of repetitive elements

SINEs	LTR elements	Small RNA	Simple repeats	Unclassified: Other	Transcription Start site
LINEs	DNA elements	Satellites	Low complexity	Unclassified: Unknow	Transcription End site

重複性序列的種類

3'UTR Length: 994bp

Graph

Query Region	Match Type	Repeat Name	Repeat Class	Hit Region
47..112	C	L3	LINE/CR1	4435..4368
588..713	C	L3	LINE/CR1	4347..4209

重複性序列在此區域的位置

重複性序列對應到保留序列的正股或負股

此段重複性序列的名稱

重複性序列對應到保留序列的位置

[New Query](#)

GBRED - 圖像列表

點擊圖中的重複性序列，可以
fasta格式下載其序列

>ENST00000371274|3UTR|588..713
 CTCACTGTCTCTTGATGCAGTCCAC
 TGCATAGTCATTCAGCTCTGCAAGT
 TACACACTTAGGTGAGTTCTGCCTC
 CTGCTAGCTTCCCGCACTGGTCCC
 AGACCTGCCTCCTAGAGCCCACAC
 AGA

ENST00000371274

Description of repetitive elements

Classified: Other
 Classified: Unknow
 Transcription Start site
 Transcription End site

Graph

Query Region	Match Type	Repeat Name	Repeat Class	Hit Region
47..112	C	L3	LINE/CR1	4435..4368
588..713	C	L3	LINE/CR1	4347..4209

藉由圖示呈現出重
 複性序列分布的相
 對位置

[New Query](#)

Thank you for your attention

