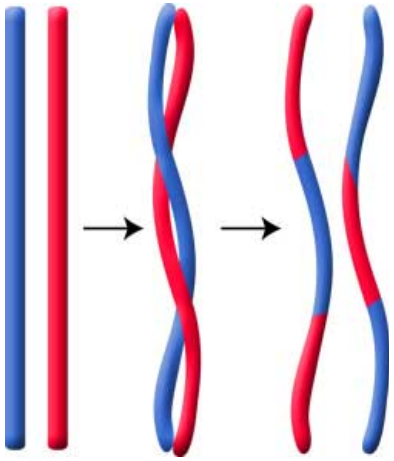
http://petang.cgu.edu.tw/bioinformatics/index.htm



Bioinformatics

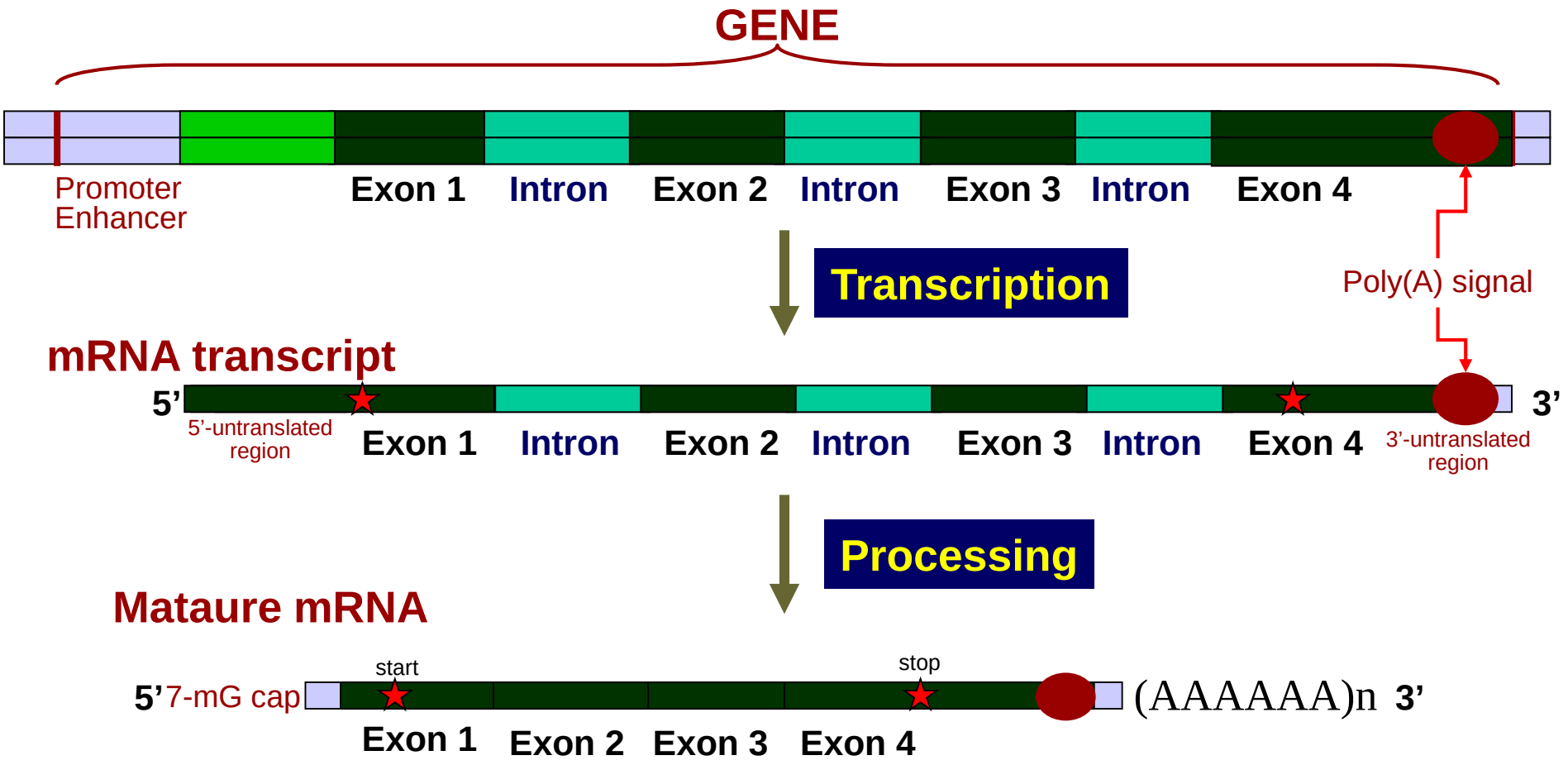
Lecture 6 – Nucleic Acids Sequence Analysis

AGTCCGCGAATACAGGCTCGGT AGTCCGCGAATACAGGCTCGGT

Petrus Tang, Ph.D. (鄧致剛)
Graduate Institute of Basic Medical Sciences
and
Bioinformatics Center, Chang Gung University.
petang@mail.cgu.edu.tw
EXT: 5136

助教：
蔡智宇(分機5690)

The Organization of an Eukaryotic Gene



Gene identification involves 4 main stages

Find the putative coding region(s) in the sequence

Open reading frame

Find non-coding features of interest in the sequence

CpG islands
Tandemly and dispersed repeats
Promoter regions (TATA box, cap signal, CCAAT-box)
Transcription factors, Poly-A sites

Determine the exon-intron organization

Branch point signal
CT(G,A)A(C,T)

5' and 3' splice sites:
AG/GUAAGU----------PyPyPyPyPyPyPyPy-CAG/G

Identify the gene

motif, signal and pattern
Blast, FASTA
Functional studies

Function	Command	GCG	SeqWEB
Sequence manipulation	Reverse 	+	+
ORF Searching	Frames 	+	+
	Map 	+	+
	Translate 	+	+
Mapping (restriction sites)	Map 	+	+
	(-minc)	+	+
	(-maxc)	+	+
	Mapsort 	+	+
	(-exclude)	+	-
	(-digest)	+	-
	Mapplot 	+	+
Mapping (transcription factors)	Map tfsites 	+	-

Select enzymes for MAP:

Type "*" to select all enzymes.
Type "**" to select all enzymes including isoschizomers.
Type individual names like "AluI" to select specific enzymes.
Type "?" to see this message and all available enzymes.
Type "??" to see the available enzymes AND their recognition sites.
Type "?A*" to see what enzymes start with "A."
Type "A*" to select all enzymes starting with "A."
Type parts of names like "Al*" to select all enzymes starting with "AL."
Type "~A*" to unselect all selected enzymes starting with "A."
Type "/*" to see what enzymes you have selected so far.
Type "#" to select no enzymes at all.

Press <Return> after each selection.
Press <Return> and nothing else to end your selections.
Spaces are allowed; upper and lower case are equivalent.

What to do next?

The predictions by these programs is just that: ***a prediction*****.**



NEVER TRUST A COMPUTER!

Exercise 105-06-1

Programs used in this exercise:

(1) Sequence manipulation – reverse

(3) ORF Searching – frames, map, translate

(4) Mapping (restriction sites)

- map (-minc, -maxc), mapsort(-exclude, -digest), mapplot, plasmidmap

(5) Mapping (transcription factor)

- map(tfsites).

Sequences used in this exercise:

fetch z18853 (*C.elegans* mRNA for capping protein alpha subunit.) cds:10-858

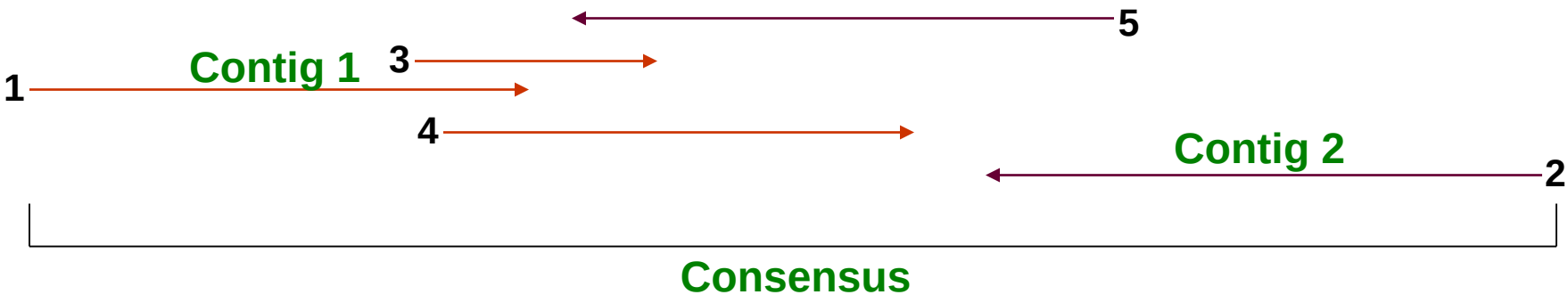
fetch x03795 (Human mRNA for platelet derived growth factor A-chain, PDGF-A) cds:388-1020.

BioEdit

Copyright © 1997-2013
Tom Hall
Ibis Biosciences
Carlsbad, CA 92008

Biological sequence alignment editor for Win95/98/NT/2K/XP/7

(1) Store fragment sequences;
(2) Recognize overlapping sequences and create aligned assemblies, called contigs;
(3) Display, edit and output the contigs for further analysis.



Assemble overlapping fragment sequences from a sequencing project.

http://www.mbio.ncsu.edu/bioedit/bioedit.html



Copyright © 1997-2013
Tom Hall
Ibis Biosciences
Carlsbad, CA 92008

Biological sequence alignment editor for Win95/98/NT/2K/XP/7

BioEdit is a biological sequence alignment editor written for Windows 95/98/NT/2000/XP/7. An intuitive multiple document interface with convenient features makes alignment and manipulation of sequences relatively easy on your desktop computer. Several sequence manipulation and analysis options and links to external analysis programs facilitate a working environment which allows you to view and manipulate sequences with simple point-and-click operations.

BioEdit.zip (Full install)
Bug fixes / changes
BioEdit General information
BioDoc.pdf (pdf format help doc)
View Screenshots

BioEdit's features include:

- Several modes of hand alignment
- Automated ClustalW alignment
- Automated Blast searches (local and WWW)
- Plasmid drawing and annotation
- Accessory application configuration
- Restriction mapping
- RNA comparative analysis tools
- Graphical matrix data viewing tools
- Shaded alignment figures
- Translation-based nucleic acid alignment
- ABI trace viewing, editing and printing
- Customizable ... other features

Note: BioEdit is no longer being regularly maintained, and the documentation is out of date and no longer maintained. BioEdit is free for use by any and all interested parties, but is supplied as is. Users must agree with the terms of the license.

Citeline.com

VALUABLE Citeline Internet Research Software RESOURCE

Enter disease, therapy or topic and click Search: Tips

Search

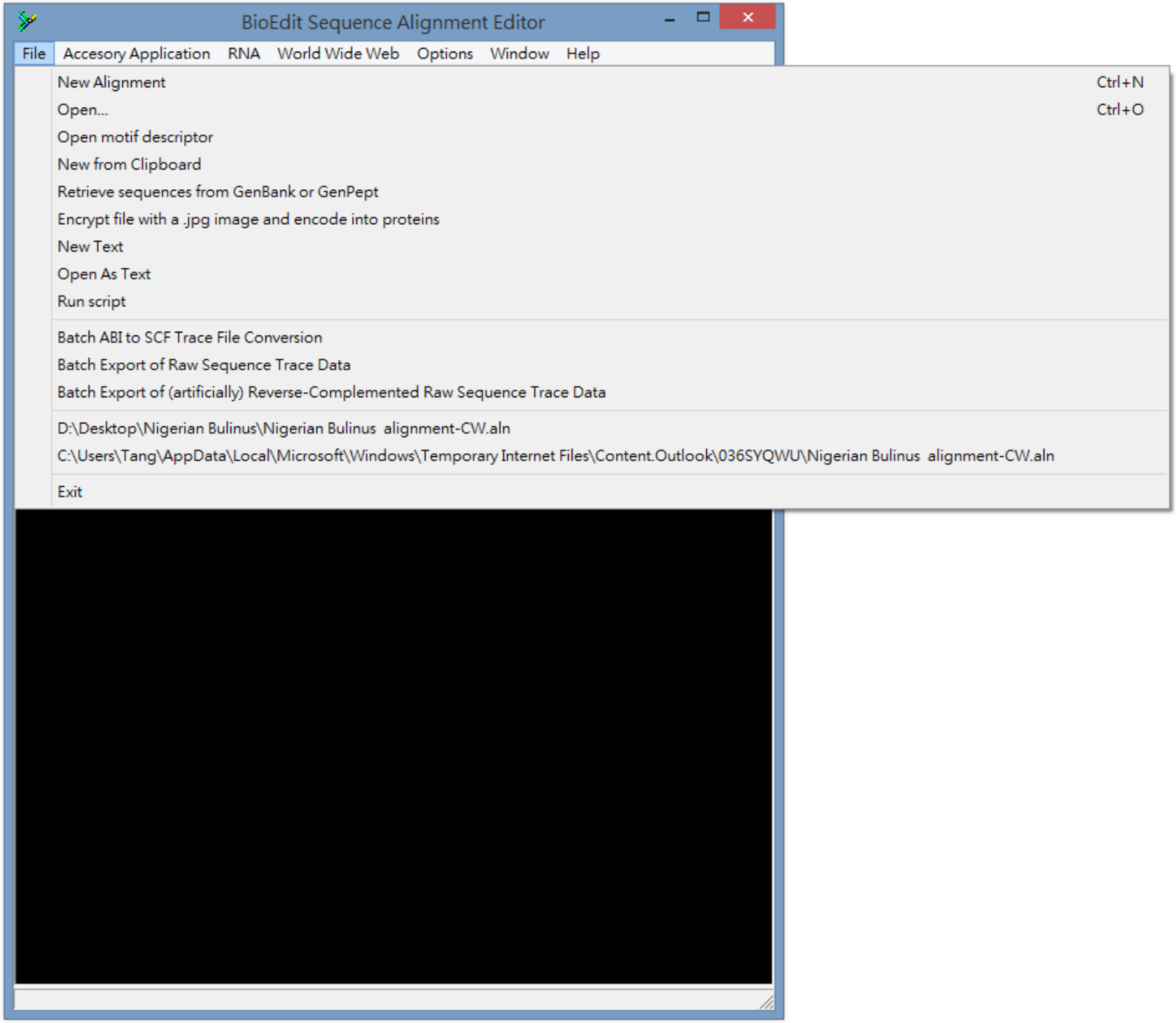
Copyright © 1999 Caredata.com, Inc.

Be updated on the software you use!
software.informer

BioEdit Software Informer

Current version 7.2.5
Last updated 12/11/2013
Tom Hall
Thomas.Hall@abbott.com

RNaseP databasee | ReadMe file | See screen shots | Accessory applications | Multi-file download | email Tom


BioEdit Sequence Alignment Editor
File
Accesory Application
RNA
World Wide Web
Options
Window
Help
New Alignment Ctrl+N
Open... Ctrl+O
Open motif descriptor
New from Clipboard
Retrieve sequences from GenBank or GenPept
Encrypt file with a .jpg image and encode into proteins
New Text
Open As Text
Run script
Batch ABI to SCF Trace File Conversion
Batch Export of Raw Sequence Trace Data
Batch Export of (artificially) Reverse-Complemented Raw Sequence Trace Data
D:\Desktop\Nigerian Bulinus\Nigerian Bulinus alignment-CW.aln
C:\Users\Tang\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Outlook\036SYQWU\Nigerian Bulinus alignment-CW.aln
Exit

Exercise 105-06-2

Download EX105_6Seq.exe
Decompress the file
-> Seq01.txt - seq10.txt

Use BioEdit to assembly the sequences.
Q1: What is the correct order of the assembled sequence?
Transfer the assembled sequence to GCG
Q2: Identity putative protein coding region
Q3: Is there any potential regulatory elements upstream of the gene?
Q4: Export the amino acid sequence
Q5: What is the identity with the human homologues protein?