



大量分析處理及視覺化系統呈現多 層次資料——以瘧疾原蟲資料為例



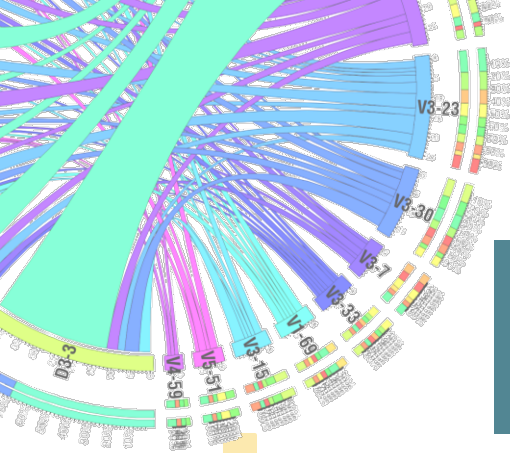
指導教授：廖崇碩 教授

組別：第一組

組員：9930171 陳姿瑾

9934008 龔 俐

9934055 謝維晨



研究目的與系統架構

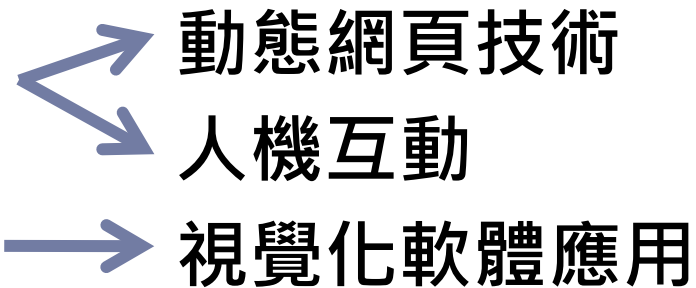
近年來因資訊蓬勃發展，各領域皆需應用龐大的數據資料，這些資料多而繁雜，需經分析歸納，並探討其多面向的關連性。

本專題研究目的為以視覺化手法分析巨量資料，並搭配網路服務平台的技術，將已視覺化的資料結合服務平台呈現。

系統
架構

服務平台的建立

巨量資訊視覺化





服務平台建立

動態網頁技術

動態網頁能與使用者進行互動，結合程式語言及資料庫。
其特色為具有後台管理員系統，管理員能及時更新網站內容。

本專題可將量化的資訊及時更新於網頁。

人機互動為系統與使用者之間的互動關係，
其介面設計須考慮使用者對系統的理解與否，
以增加使用者友好性。
依使用者的操作習慣來設計網頁介面，讓使用者達到高滿意度的心境。



為何要將資訊視覺化

視覺化工具介紹：Circos

Circos可將資料轉換成圓弧形圖像呈現，創造出多層次的圖形；
並同時比對多維數據，較容易讀出數據間的相對關係。



圓周的標籤為地點

圈內線條為不同貨品
連線至不同地點

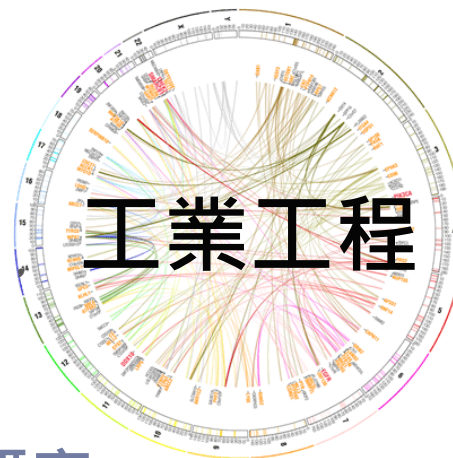


為何要將資訊視覺化

Circos於工業工程的應用

生產管理

應用於物流管理，分析城市間運輸量、貨運往來的頻率，用圖形呈現能一眼看出熱門路線分析兩地的交集性。



作業研究

求解最佳化問題，將數學方程式轉成圖形，藉以探討兩變數或多變數的關聯性。

人因工程

職業類別與工作時可能造成的傷害部位，其傷害程度分別為何，運用圖形整理，提供給工會、職訓中心、企業等參考。



為何要將資訊視覺化

A	B	C	D	E	F	G	H	I	J	K	L	M
Chr	Effect	Effect_Imp	Functional	Codon_Ch	Amino_Ac	Gene_Nam	Sample	Coverage	RefCov	VarCov	VarFreq	Sample_co
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	116,24,87,72,13,54,3,44,11,33,3,37.93%,45					44
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	20,32,91,2,11,25,71,1,9,7,20,44,3,45%,21.88					24
M76611	NON_SYNMODERATE	MISSENSE	Gct/Act	A1T	malmito_rr	ERR01289	22,144,35,3,42,14,4,2,12,69,16,8,54.55%,47					49
M76611	NON_SYNMODERATE	MISSENSE	gCt/gAt	A1D	malmito_rr	ERR01289	23,155,40,4,53,19,37,17,61,15,5,40.48%,23					45
M76611	NON_SYNMODERATE	MISSENSE	gCt/gGt	A1G	malmito_rr	ERR01991	167	124	44	20.66%		1
M76611	FRAME_S' HIGH	.	-C	-1?	malmito_rr	ERR01991	163,178	115,120	47,62	22.27%,25		2
M76611	NON_SYNMODERATE	MISSENSE	Atc/Gtc	I239V	mal_mito_	ERR01905	#####	0,0,2,0,0,0	#####	100%,100%		24
M76611	SYNONYM LOW	SILENT	ttC/ttT	F77	mal_mito_	ERR01991	#####	0,1,0	#####	100%,99.5		3
M76611	SYNONYM LOW	SILENT	Tta/Cta	L8	mal_mito_	ERR01991	#####	0,0,0	#####	99%,99.02		3
M76611	STOP_GAIHIGH	NONSENSE	taC/taA	Y6*	malmito_rr	ERR03770	#####	#####	#####	32.85%,32		2
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01905	#####	1,1,0,0,0,1	#####	99.96%,99		21
M76611	NON_SYNMODERATE	MISSENSE	aCc/aAc	T21N	mal_rna_1	ERR01289	140,60	97,36	43,22	30.28%,36		2
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	134,17,162,106,10,122,29,7,40,13	21.48%,41				35
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	35	27	8	22.86%		1
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	32	20	8	25%		1
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	30,124,63,16,76,36,1,11,34,14,6,36.67%,27					18
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	82,105,23,36,59,122,25,27,51,1,30.49%,25					42
M76611	INTERGENMODIFIER.	.	.	.	.	ERR02294	34	4	13	38.24%		1
M76611	INTERGENMODIFIER.	.	.	.	.	ERR01289	9,65,13,62,0,6,3,9,4,1,7,27,6,30,177.78%,41					44
Pf3D7_01	INTERGENMODIFIER.	.	.	.	.	ERR03770	22	14	8	36.36%		1
Pf3D7_01	INTERGENMODIFIER.	.	.	.	.	ERR01904	20	11	7	36.84%		1
Pf3D7_01	INTERGENMODIFIER.	.	.	.	.	ERR01904	33,14,19,1,21,8,12,7,410,6,7,6,1230.3%,42.8					10
Pf3D7_01	INTERGENMODIFIER.	.	.	.	.	ERR01544	29,30,29,3,20,23,21,2,8,7,8,7,7,27.59%,23					12

資料來源：長庚醫院



研究結果

瘧疾原蟲資料實做範例

分析瘧疾原蟲16種染色體(chromosome)在13種不同實驗下(effect)基因的變異程度。

Summary View	Chromosomal View
從 巨觀 角度探討檢體差異性	從 微觀 角度探討檢體差異性
整組基因組	單筆染色體
欄位為染色體	欄位為基因數目
以顏色深淺表示數值高低	
有四種類別(type)可供觀察	

Choose Types:

- ☐ Coverage
- ☒ RefCov
- ☐ VarCov
- ☐ VarFreq

研究結果

染色體

瘧疾原蟲資料實做範例

變異發生的比率高：

⇒ 原蟲易產生變異

⇒ 易產生具抗藥性之基因，
成為新型疾病的來源

圖中顯示的是檢體的染色體基因變異情況，
顏色越深表示變異發生的比率越多。



研究結果

瘧疾原蟲資料實做範例

此種呈現方法能提供**快速檢視與篩選**之用，對於特定基因位點也能透過彈出式選單即時呈現**詳細變異資訊**

Chr4:1,807,894
Ref/Var:G/A
Ti/Tv:Transition
VarFreq:99.79%
Coverage:1935
Zygosity:Hom

GeneSym:FGFR3
Region_Type:exonic
TranscriptID:NM_022965
CDS_Mut:c.G1617A
AA_Mut: p.T539T
Mut_Type: synonymous

dbSNP:rs7688609
1000G_Variant_Frequency:
American:0%
Asian:0%
African:0%
European:0%
COSMIC:---