

第十一屆 (BITS11)
海峽兩岸生物啟發的理論科學研討會
台灣南投縣 中興大學惠蓀林場
2018年7月3-6日

六種複雜疾病基因組的功能與分子異常之比較：共同、相反、與唯一性

李弘謙
中原大學物理系
中央大學醫工系
2018年7月3日

背景和目的

許多複雜疾病的分子基礎，包括2型糖尿病，神經退化性疾病和各類癌症，都非常複雜，尚未完全了解。

文獻中尚未見比較多種複雜疾病功能與分子異常的異同的報告。

這些疾病都有豐富的公共基因組數據庫。

這裡深度分析了六種不同複雜疾病的基因表達數據，以鑑定疾病中共有的功能和分子特徵以及每種疾病獨有的功能和分子異常特徵。

研究樣本 Material

19 gene expression
microarray datasets of
six complex diseases:

Type 2 Diabetes (T2D, 3)

2型糖尿病

Kidney Disease (KD, 3)

慢性腎病

Alzheimer's (AD, 5)

艾茲海默症

Thyroid cancer (TC, 5)

甲狀腺癌

Gaucher Disease (GD, 2)

戈謝病

Multiple Sclerosis (MS, 1)

多發性硬化症

Name	Number of chips (test vs. control)	PCA score	Reference
T2D_blood_GSE15932	8 vs. 8	0.75	—
T2D_islets_GSE25724	6 vs. 7	0.85	(85)
T2D_glomeruli_GSE30122	9 vs. 13	0.94	(86)
KD_small_renal_GSE7869	5 vs. 3	1	(87)
KD_medium_renal_GSE7869	5 vs. 3	1	
KD_large_renal_GSE7869	3 vs. 3	1	
AD_EC_GSE5281	9 vs. 11	0.91	(88)
AD_HC_GSE5281	8 vs. 10	1	
AD_MTG_GSE5281	14 vs. 11	1	
AD_PC_GSE5281	9 vs. 10	0.95	
AD_SFG_GSE5281	12 vs. 8	1	
TC_FC_GSE27155	13 vs. 4	0.88	(89)
TC_FA_GSE27155	10 vs. 4	0.95	
TC_OA_GSE27155	7 vs. 4	1	
TC_OC_GSE27155	8 vs. 4	1	
TC_AC_GSE27155	4 vs. 4	1	
GD_type1_fibroblasts_GSE21899	5 vs. 4	1	—
GD_type3_fibroblasts_GSE21899	5 vs. 4	1	
MS_blood_GSE17393	8 vs. 7	0.81	(90)

方法 I

- 使用複雜疾病的公共微陣列數據進行分析



- 使用分子標籤 (MSig DB) 表徵疾病的基因圖譜



- 使用GSEA分析法找出加富分子標籤



- 從顯著的加富分子標籤篩選高出現頻率的“前沿基因” (leading edge genes (LEGs)) 代表疾病。

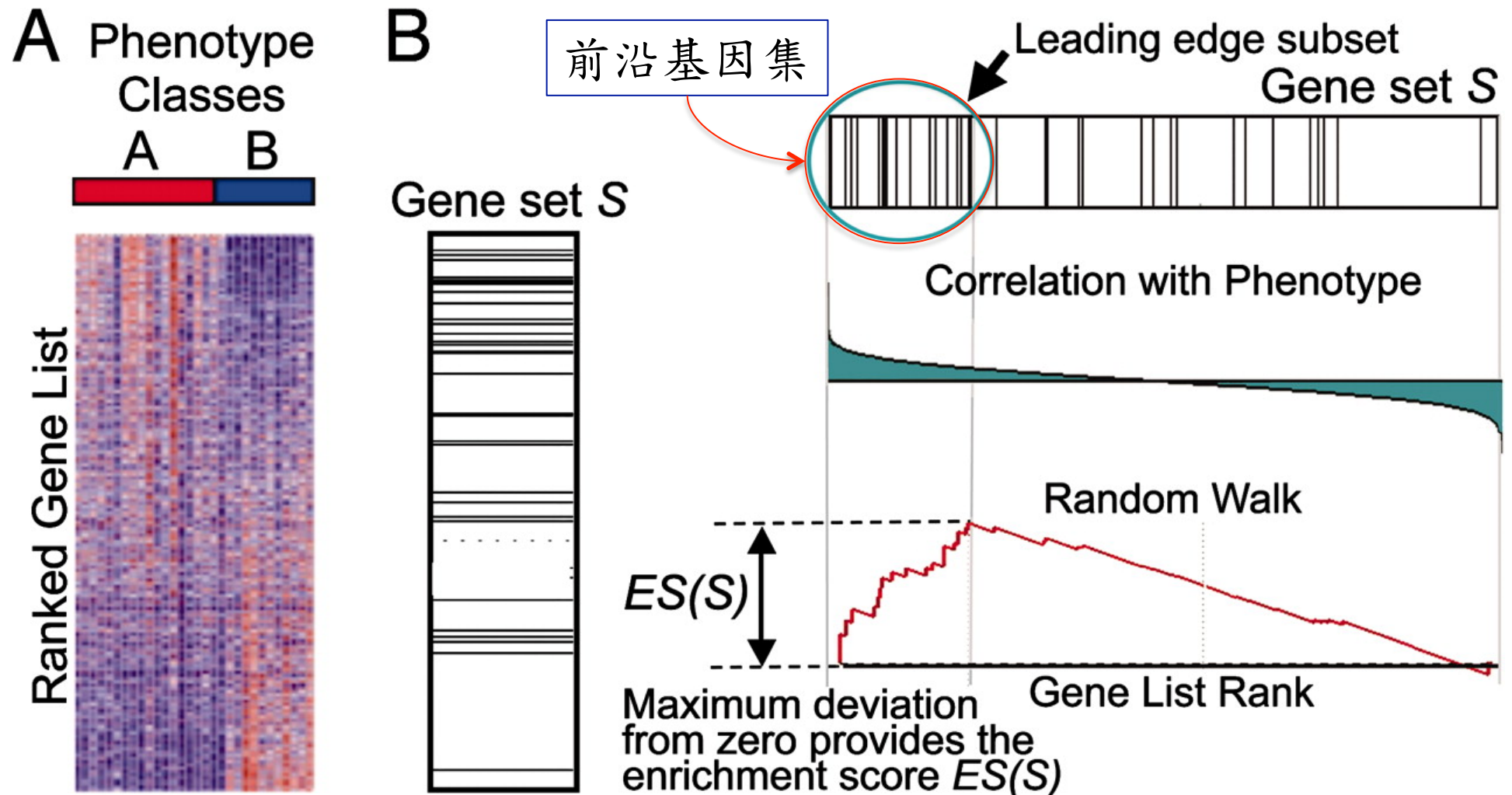
方法 II

- 依據各疾病高頻率前沿基因集
再用GSEA識別疾病過度/
未激活的 KEGG 生物功能名詞
- 比較 疾病結果 鑑定複雜疾病的共同和獨特
功能和唯一分子特徵



GSEA - Gene Set Enrichment Analysis

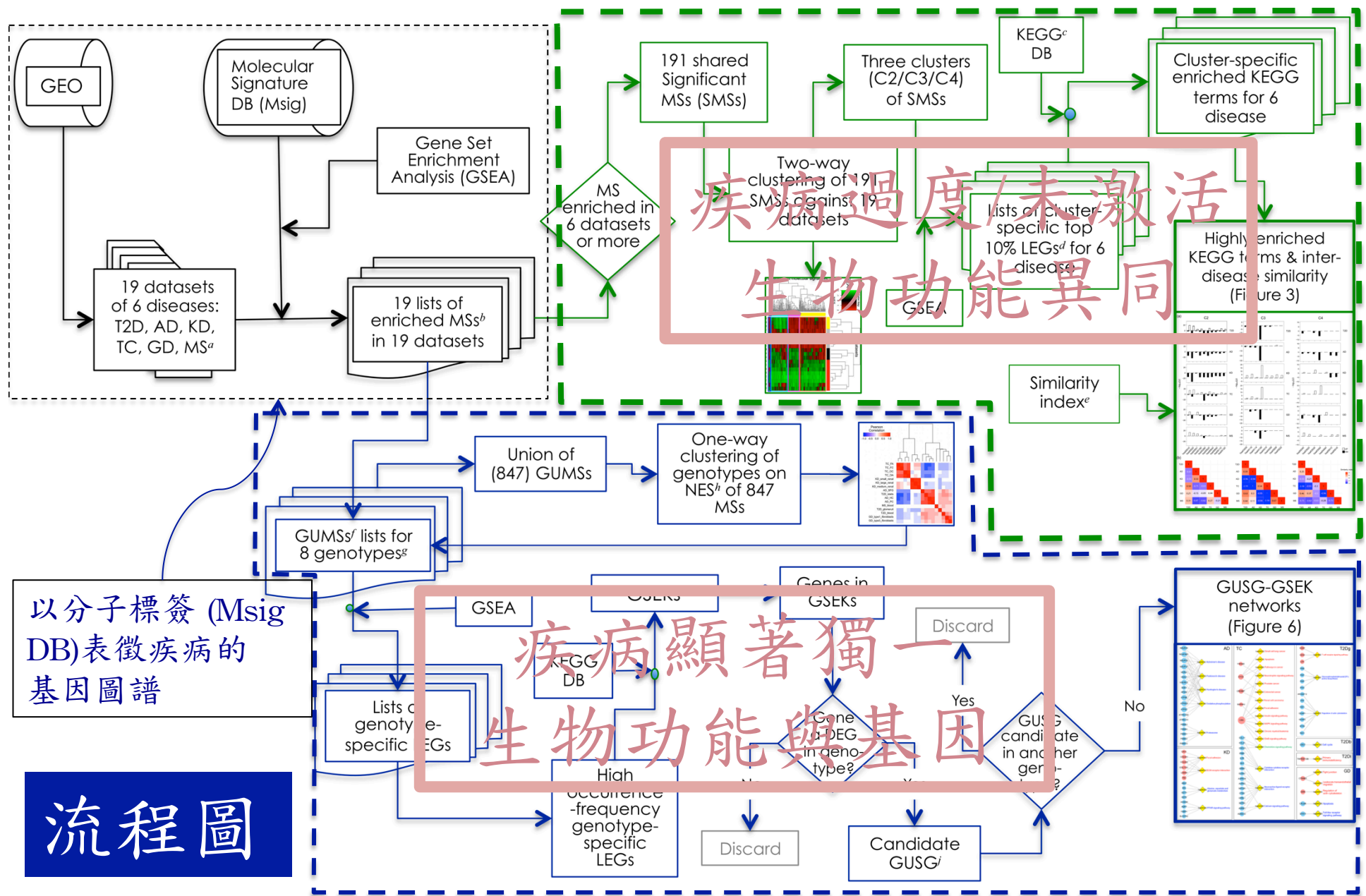
基因集加富分析



Subramanian A et al. PNAS 2005;102:15545-15550

©2005 by National Academy of Sciences

PNAS



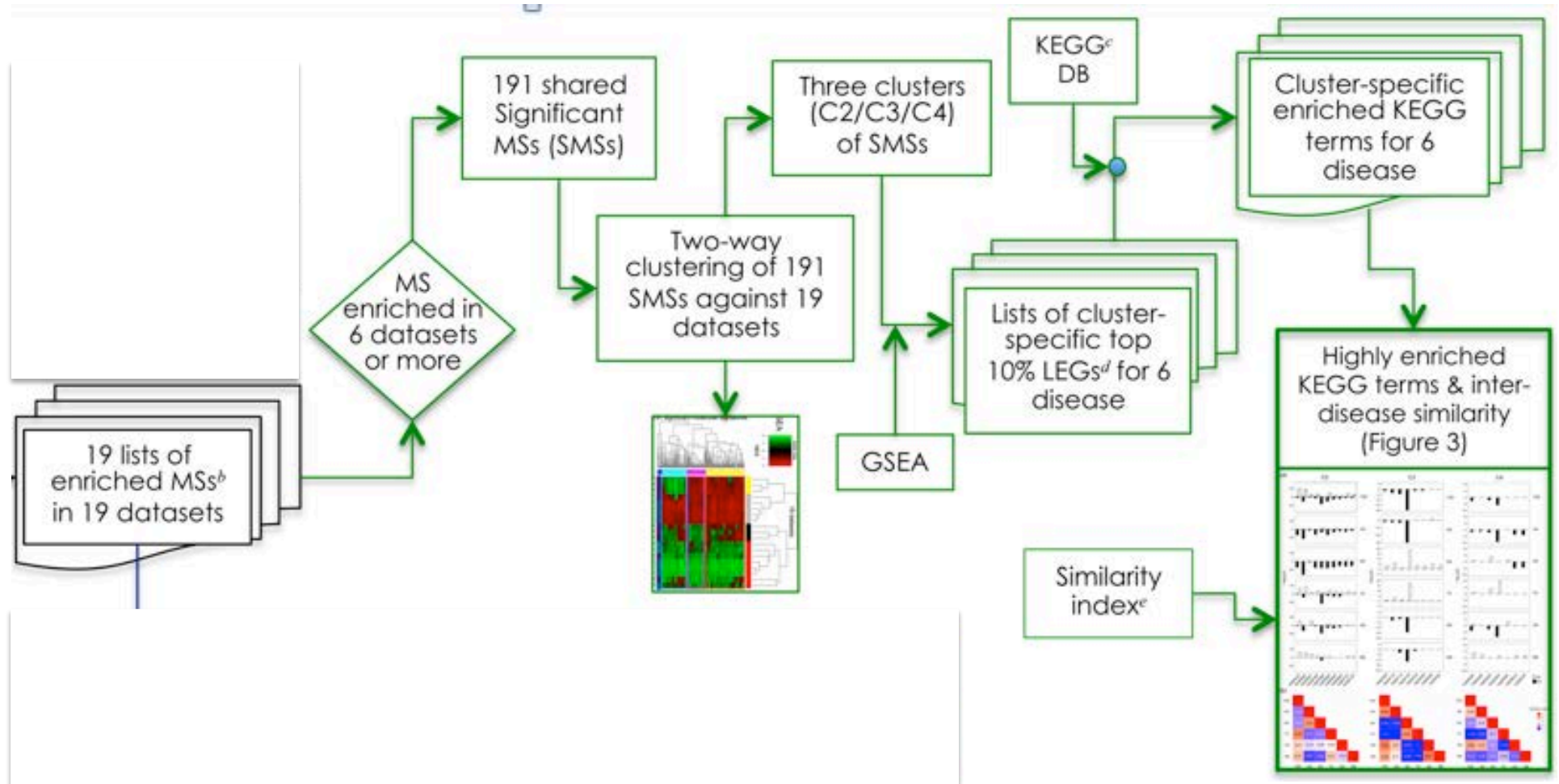
^aType 2 diabetes, Alzheimer's disease, kidney disease, thyroid cancer, Gaucher's disease, multiple sclerosis. ^bMS, molecular signature. ^cKEGG, Kyoto Encyclopedia of Genes and Genomes. ^dLEG, leading-edge genes, as defined in GSEA. ^eSimilarity index, see Methods. ^fGUMS, genotype-unique MS. ^g8 genotypes, T2D_blood, T2D_islets, T2D_glomeruli, AD, KD, TC, GD, MS, from 16 datasets. ^hNES, normalized enrichment score as defined in GSEA. ⁱGSEK, genotype-specific enriched KEGG term. ^jGUSG, genotype-unique significant gene.

部分一

鑑定複雜疾病
過度/未激活
生物功能的異同

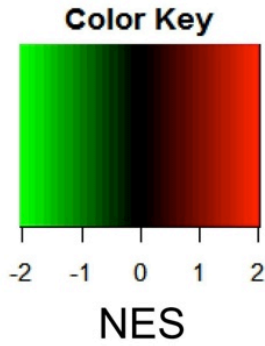
流程圖

認定191條部分共享顯著分子標籤

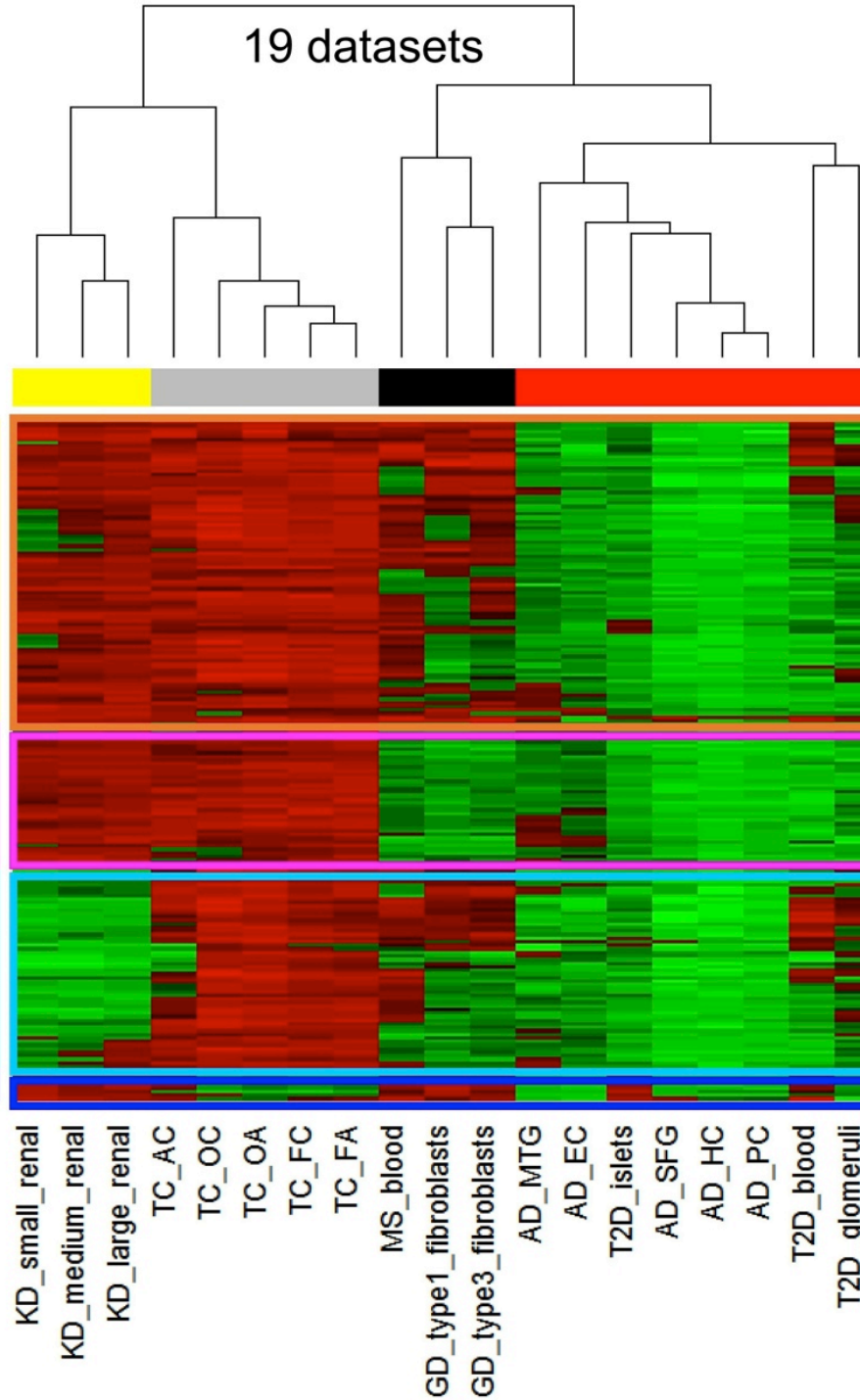
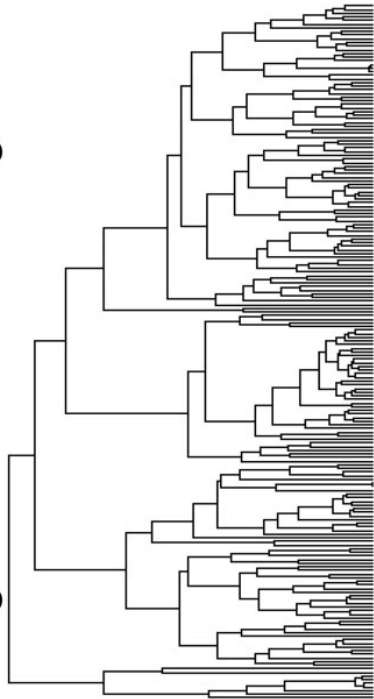


^aType 2 diabetes, Alzheimer's disease, kidney disease, thyroid cancer, Gaucher's disease, multiple sclerosis. ^bMS, molecular signature. ^cKEGG, Kyoto Encyclopedia of Genes and Genomes. ^dLEG, leading-edge genes, as defined in GSEA. ^eSimilarity index, see Methods. ^fGUMS, genotype-unique MS. ^g8 genotypes, T2D_blood, T2D_islets, T2D_glomeruli, AD, KD, TC, GD, MS, from 16 datasets. ^hNES, normalized enrichment score as defined in GSEA. ⁱGSEK, genotype-specific enriched KEGG term. ^jGUSG, genotype-unique significant gene.

GSEA



1/3 NOM p-val < 0.05
191 significant molecular signatures



以191條部分共享顯著分子標籤共集對19組數據組分群，數據組有兩大叢集：TC/KD，以及 AD/T2D/GD/MS。
同時分子標籤分為四個叢集，其中C1很小，之後忽略。

Cluster 4 (85 SMSs)
基因調控、細胞週期

Cluster 3 (41 SMSs)
蛋白質複合體、
病毒細菌感染

Cluster 2 (56 SMSs)
小分子、新陳代謝

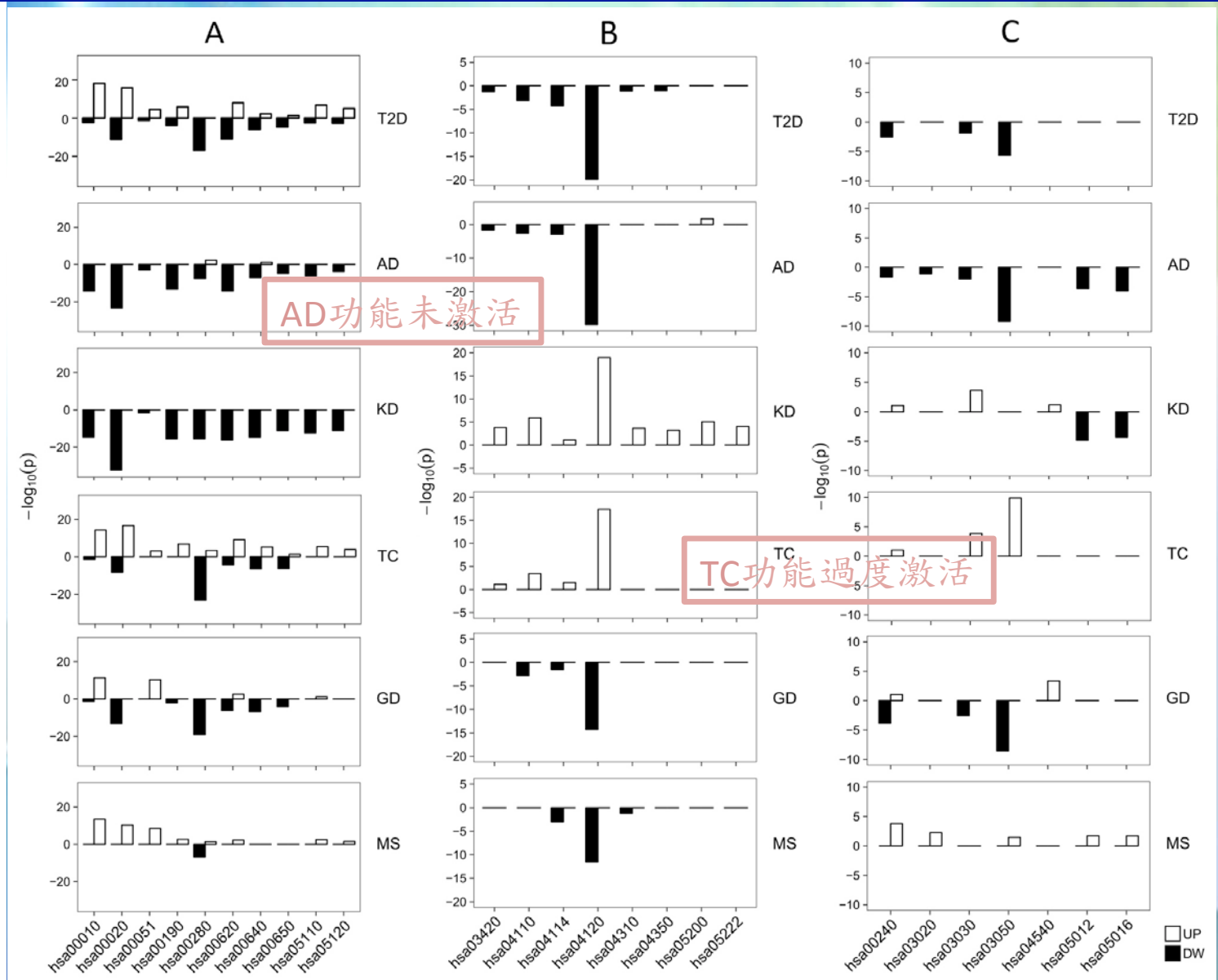
Cluster 1 (9 SMSs)

六種複雜疾病在三類功能過度/未激活程度之異同

A - KEGG terms enriched in C2 genes.
 Hsa00010: Glycolysis/Gluconeogenesis,
 hsa00020: Citrate cycle (TCA cycle), hsa00051: Fructose and mannose metabolism, hsa00190: Oxidative phosphorylation,
 hsa00280: Valine, leucine and isoleucine degradation, hsa00620: Pyruvate metabolism, hsa00640: Propanoate metabolism, hsa00650: Butanoate metabolism,
 hsa05110: Vibrio cholerae infection, hsa05120: Epithelial cell signaling in Helicobacter pylori infection.

B - Hsa03420: Nucleotide excision repair, hsa04110: Cell cycle, hsa04114: Oocyte meiosis, hsa04120: Ubiquitin mediated proteolysis, hsa04310: Wnt signaling pathway, hsa04350: TGF-beta signaling pathway, hsa05200: Pathways in cancer, hsa05222: Small cell lung cancer.

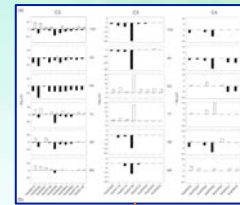
C - Hsa00240: Pyrimidine metabolism, hsa03020: RNA polymerase, hsa03030: DNA replication, hsa03050: Proteasome, hsa04540: Gap junction, hsa05012: Parkinson's disease, hsa05016: Huntington's disease.



六種複雜疾病在三類功能活性相似指數

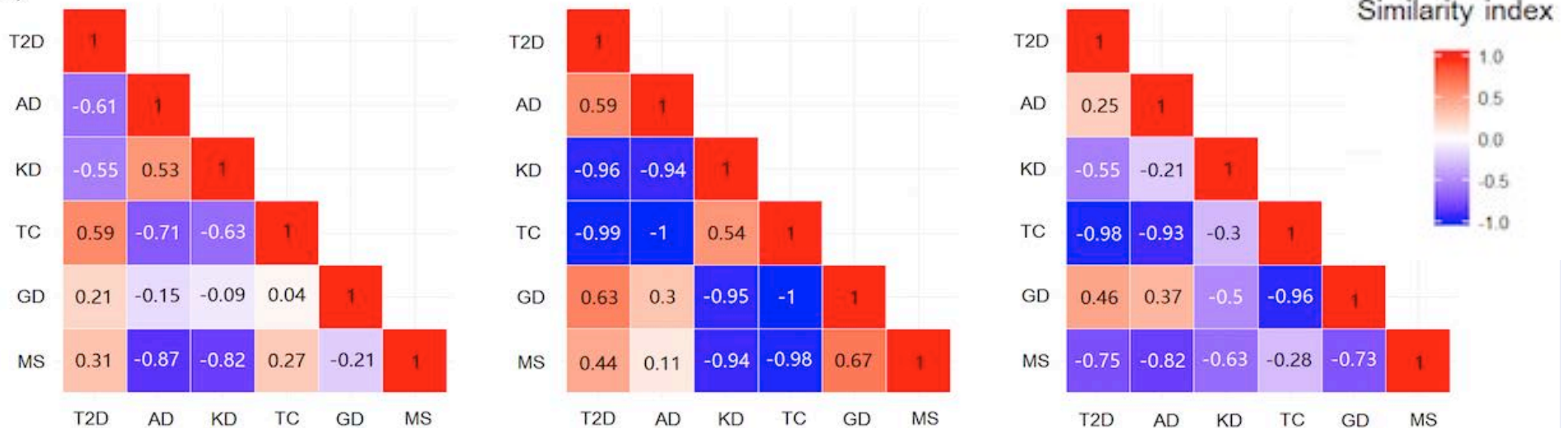
小分子、新陳代謝類

蛋白質複合體、
病毒細菌感染類



基因調控、細胞週期類

(b)



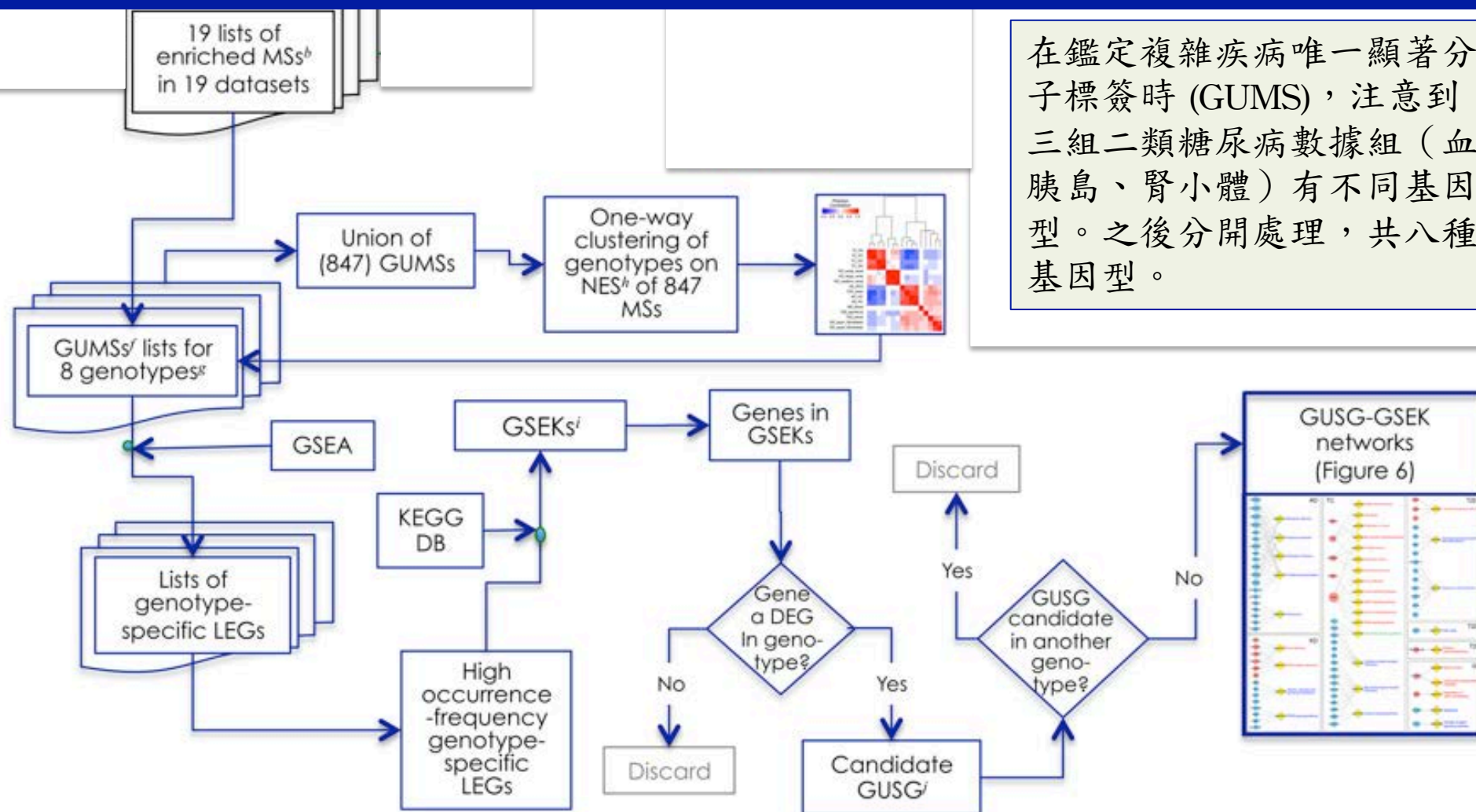
全域性相反：AD-TC、T2D-KD、KD-MS，全域性相似：T2D-GD，部分相似部分相反：如 AD-T2D、TC-T2D。

部分二

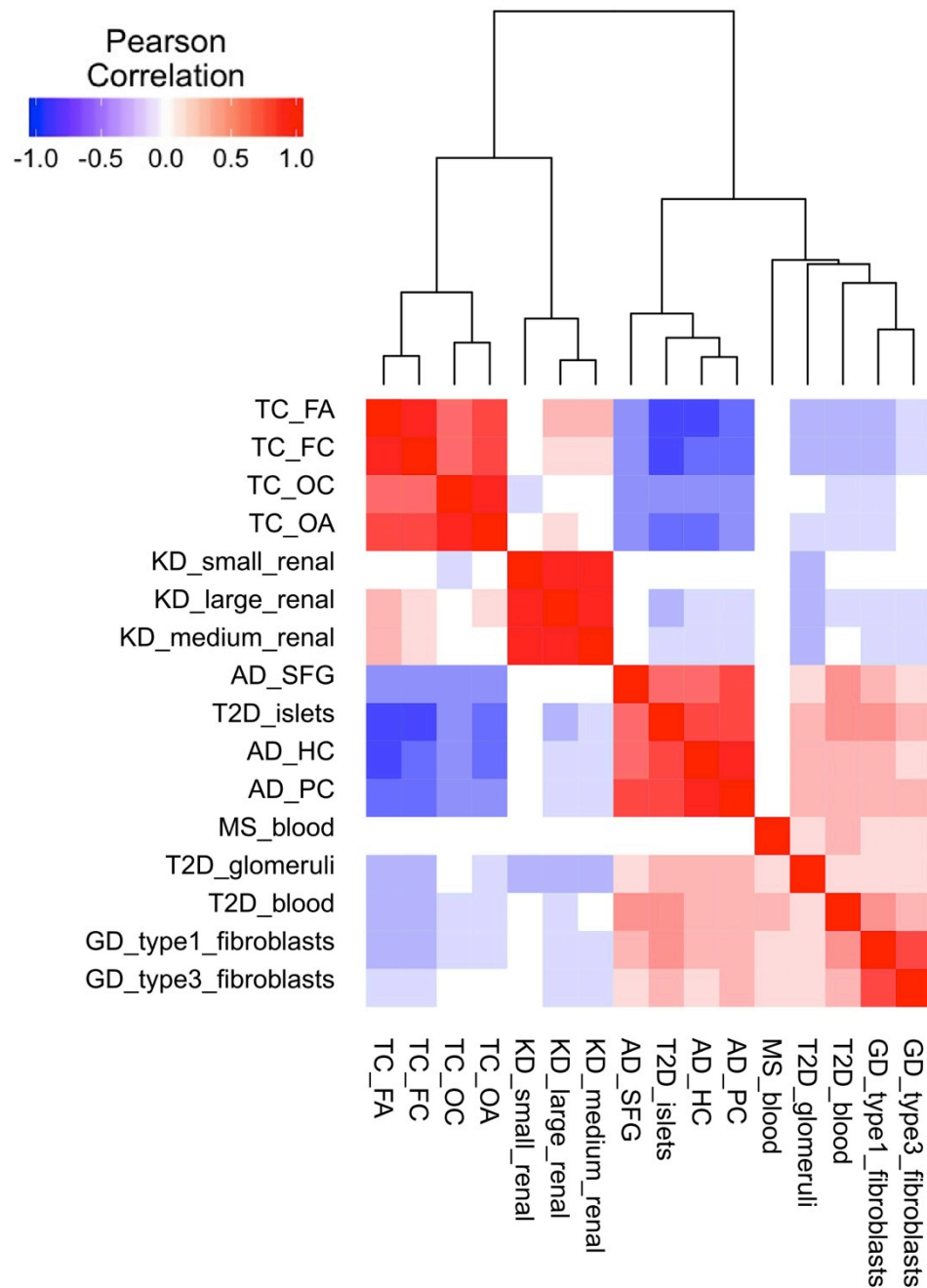
鑑定複雜疾病（基因型）
獨特生物功能與
唯一顯著基因

鑑定複雜疾病基因型 獨特生物功能與唯一顯著基因

在鑑定複雜疾病唯一顯著分子標籤時 (GUMS)，注意到三組二類糖尿病數據組（血胰島、腎小體）有不同基因型。之後分開處理，共八種基因型。



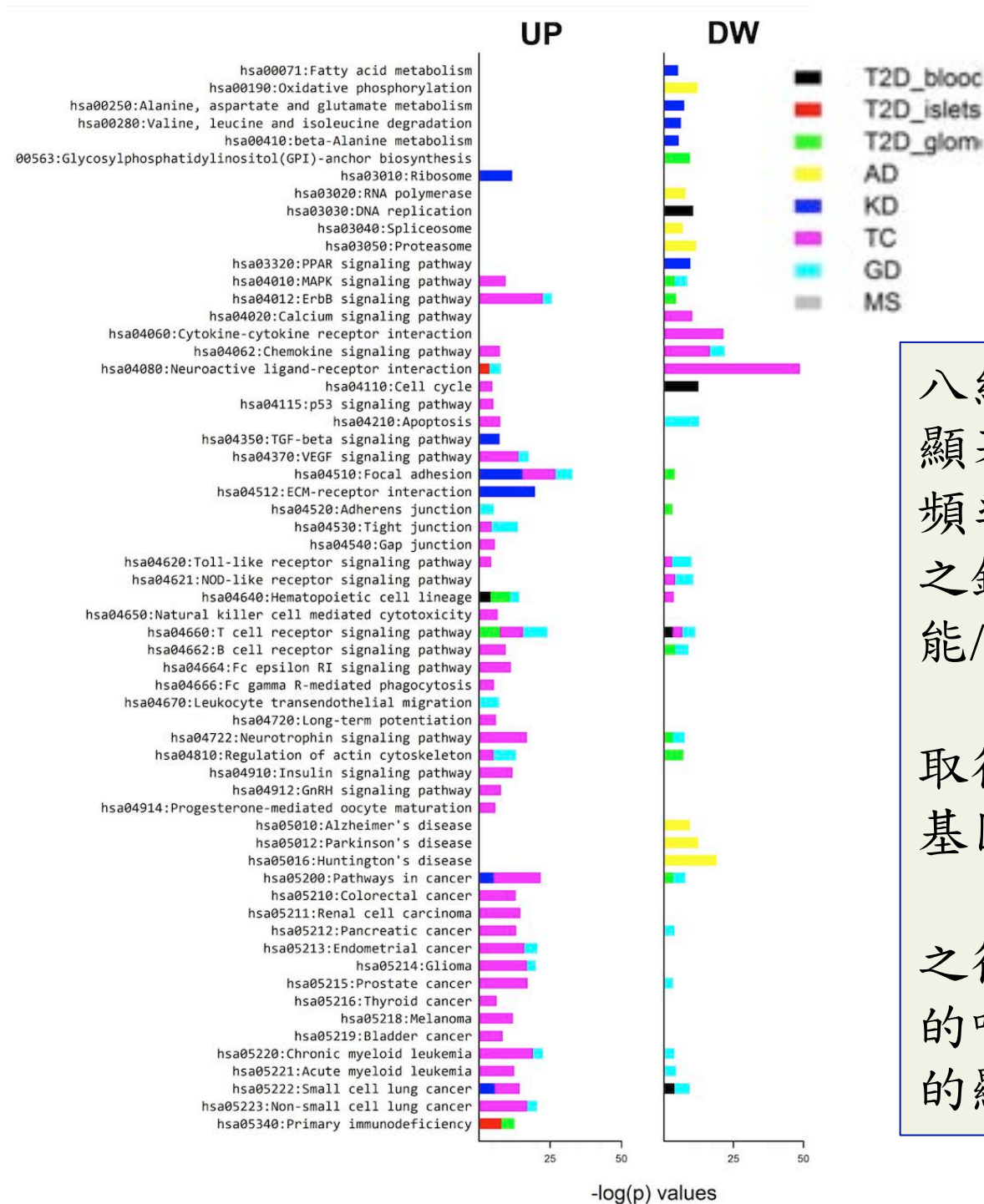
^aType 2 diabetes, Alzheimer's disease, kidney disease, thyroid cancer, Gaucher's disease, multiple sclerosis. ^bMS, molecular signature. ^cKEGG, Kyoto Encyclopedia of Genes and Genomes. ^dLEG, leading-edge genes, as defined in GSEA. ^e8 genotypes, T2D_blood, T2D_islets, T2D_glomeruli, AD, KD, TC, GD, MS, from 16 datasets. ^hNES, normalized enrichment score as defined in GSEA. ⁱGSEK, genotype-specific enriched KEGG term. ^jGUSG, genotype-unique significant gene.



以八組基因型唯一顯著分子標籤
共集對16組數據組分群，除
T2D_islets（二類糖尿病_胰島）
混入AD中之外，八組基因型分
群清晰。

仍分兩大類：TC/KD，以及 AD/
T2D/GD/MS。其中 T2D_blood
與 GD 相近。

呼應之前疾病分群結果



八組基因型各自用唯一顯著分子標籤集萃取高頻率前沿基因集，再以之鑑定 KEGG 中顯著功能/路徑。

取得之顯著功能/路徑非基因型唯一。

之後再認定每一基因型的唯一對應某功能/路徑的顯著基因。

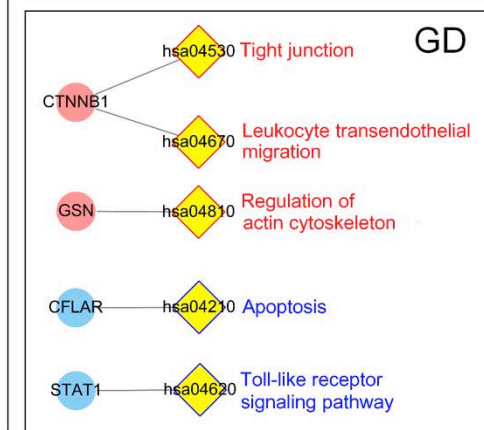
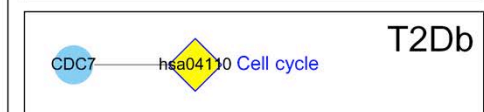
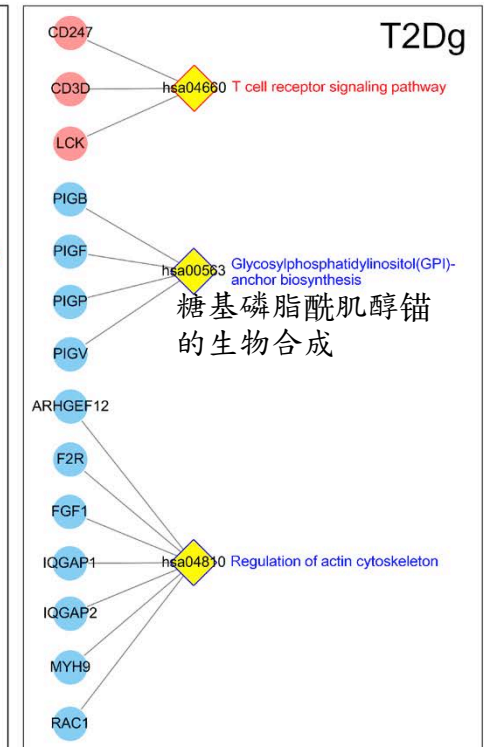
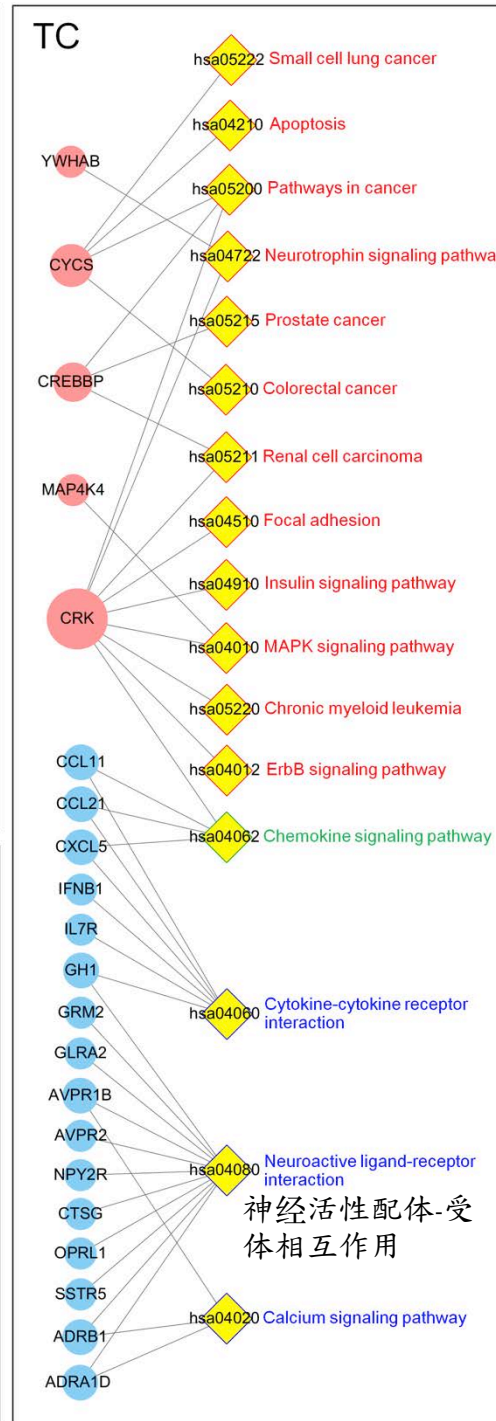
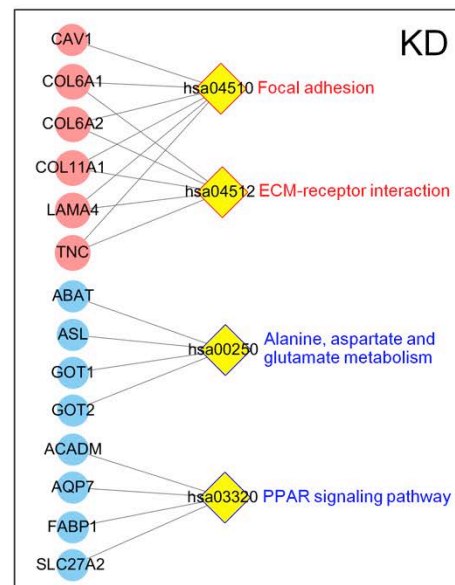
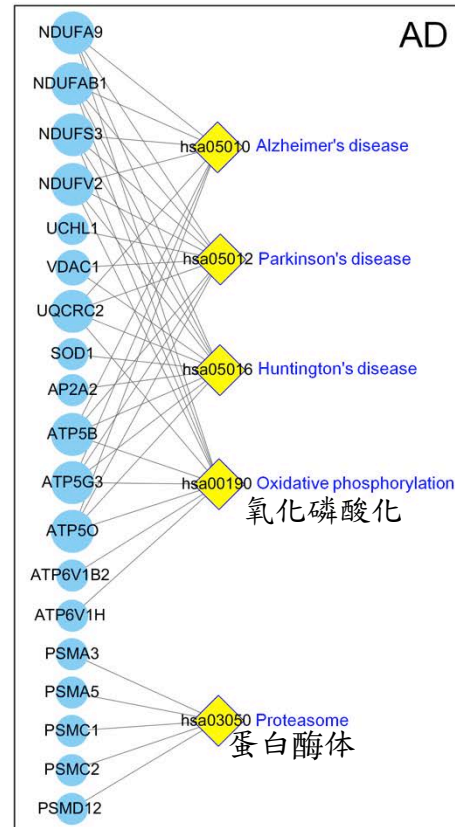
基因型特定功能/
路徑-唯一顯著基
因網絡

紅色上調、藍色
下調

共認定74個基因
型（MS除外）唯
一顯著基因

圖中所有的疾病-
路徑-基因組合都
是良好的實驗驗
證對象。

許多疾病-基因組
合已經有文獻報
導。



總結

- 複雜疾病的異常功能剪影有明顯的相互關係：有全域性相反如 AD-TC、T2D-KD，全域性相似如T2D-GD，部分相似部分相反如 AD-T2D
- 三組二類糖尿病數據組（血、胰島、腎小體）為不同基因型
- 建立了基因型特定功能/路徑-唯一顯著基因網絡，認定74個基因型唯一顯著基因
- 提供了許多疾病-路徑-基因組合的實驗驗證對象
- 對六種複雜疾病功能、分子層次的了解提供訊息，有助於複雜疾病的診斷與治療，對臨床的共病現象提供功能、分子層次的理解。

感謝



國立中央大學
National Central University



中原大學



國泰綜合醫院
Cathay General Hospital

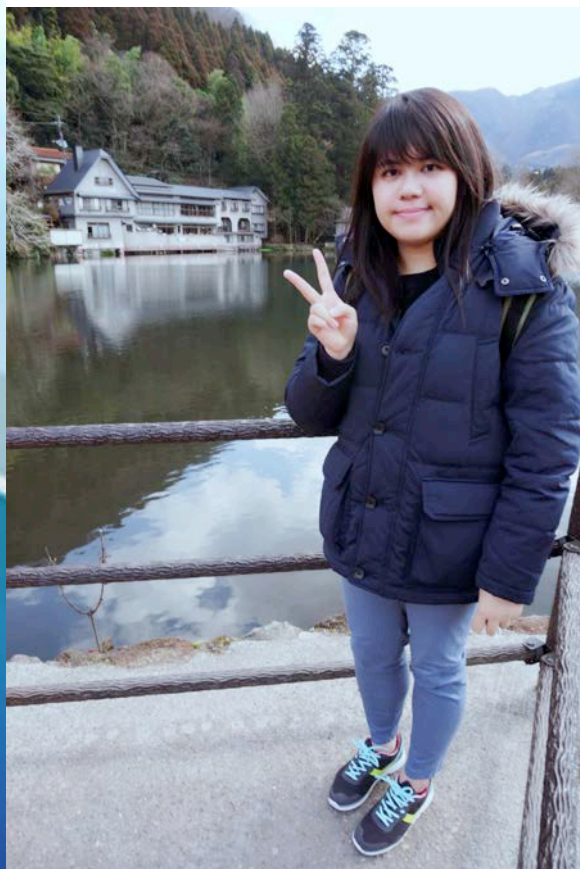


LANDSEED HOSPITAL
TAIWAN

聯新國際醫療壠新醫院

實際工作者

田雅馨博士前
中央大學生醫工程系
系統生物所



彭奕憲博士前
中央大學生醫工程系
系統生物所





感謝聆聽、請多指教