您好，我是統計魔法師楊欣洲。面對既有傳統疾病（例如：糖尿病、黃斑部病變、癌症、脂肪肝、骨質疏鬆等）和新興傳染病（例如：新冠肺炎、新型流感等）的嚴峻挑戰，我們與資料視覺大師陳君厚組成嶄新的魔法團隊，兵分多路展開對抗人類疾病的聖戰。除了攜帶先前開發的「追蹤藥水」（可定位致病基因）、「現形藥水」（可找出致病反應路徑）、「陷阱藥水」（可辨識藥物副作用）、「復原藥水」（可降低疾病風險）外，更進一步利用兩種新藥材：

(1)「醫學影像大數據」，調配出新的「透視藥水」，可以直接看穿罹病者不同臟器的醫學影像異常之處。

(2)「病毒全基因體定序大數據」，調配出新的「檢測藥水」，可以快速判斷傳染病病原的種類，以便疫苗開發和對症下藥。

這些聖戰正如火如荼展開中，歡迎加入我們魔法團隊，一起為人類健康而戰！

中央研究院統計科學研究所
楊欣洲老師實驗室
陳君厚老師智慧醫療計畫

Pbine 林胤均

P 值 (i.e., 統計顯著性) 整合法可以整合多方資訊做出有力的結論，已廣泛應用於生物資訊學與計算生物學，諸如：統合分析 (meta-analysis)、基因關聯分析、反應路徑 (pathway) 分析、藥物安全性評估、多模態腦成像分析等。

我們開發新穎的 *P* 值整合法，稱為 Pbine，進一步考慮不同資訊間的重要性和相關性，利用數值積分方法來進行統計推論。模擬研究結果顯示，Pbine 比起常用的獨立假設下的方法 (Fisher)、排列方法 (permutation)、去相關方法 (decorrelation)，在誤報錯誤、統計檢定力、計算效率、準確性等方面皆表現較好。

概念舉例說明：

目標：酗酒和癌症有無關係？

材料：收集各地有關酗酒及癌症之間關聯性的研究報告結果，可能部分結果顯示兩者之間有顯著相關，也可能部分結果顯示兩者無關。

方法：Pbine 透過結合 *P* 值與以下資訊來整合結果：

1. 報告間的相關性。(例：若兩研究使用的樣本或方法相似，則相關性較高)
2. 各報告的重要性。(例：若某研究使用的樣本數較多，可信度可能較高)

SMART² 梁佑任

一. 資料輸入
1. 質譜儀的原始檔
2. mzXML檔案
3. 波峰強度資料表

二. 資料檢視
1. 2D、3D圖譜
2. TIC圖

三. 波峰分析
1. 非目標行波峰
2. 目標型波峰

四. 資料前處理
1. 內標校正
2. 資料轉換
3. 資料標準化

五. 品質管控
1. 波峰篩選
2. 樣本篩選

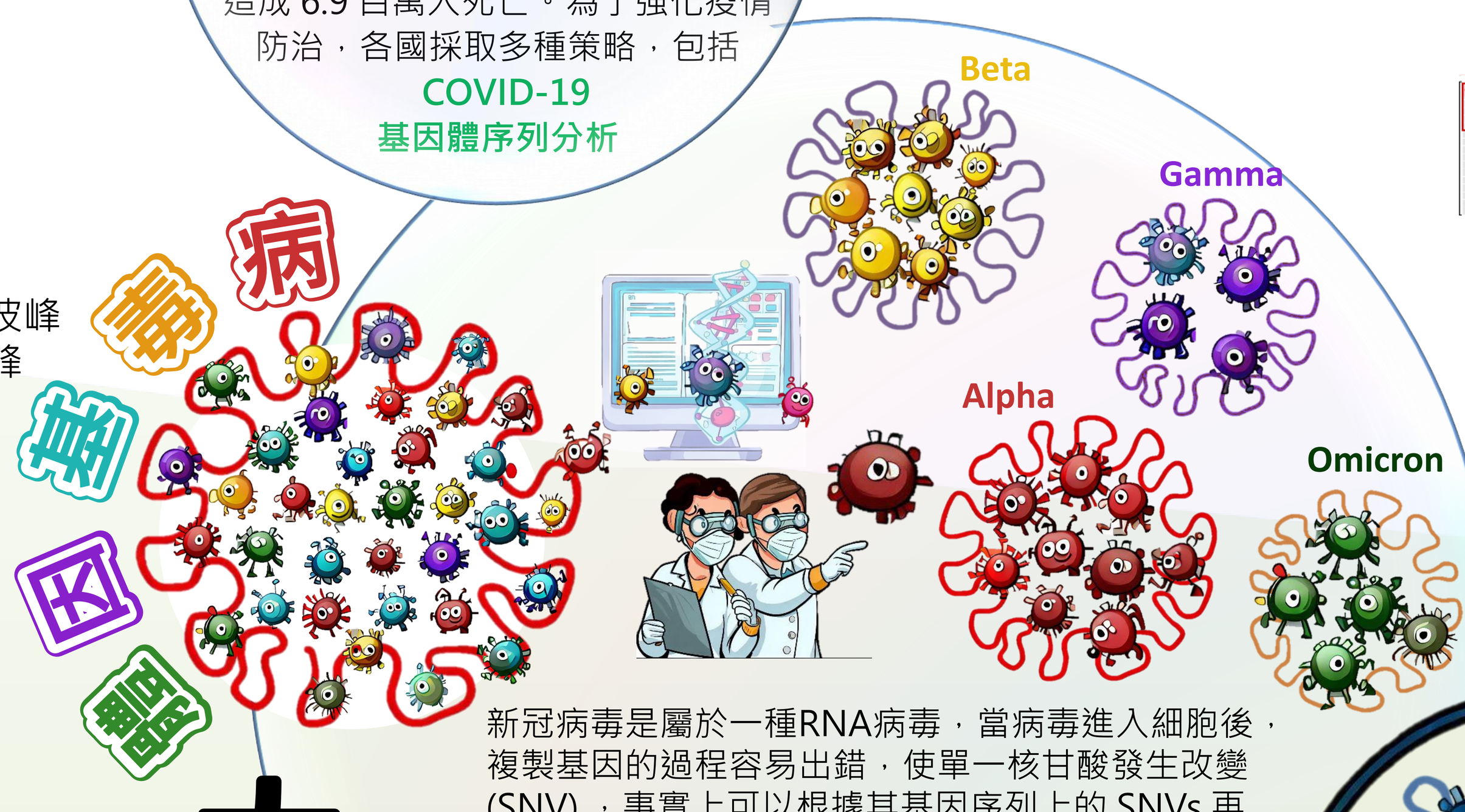
六. 批次效果分析
1. 主成分分析
2. 集群分析

七. 統計分析
1. ANCOVA
2. PLS/PLS-DA
3. IOPA

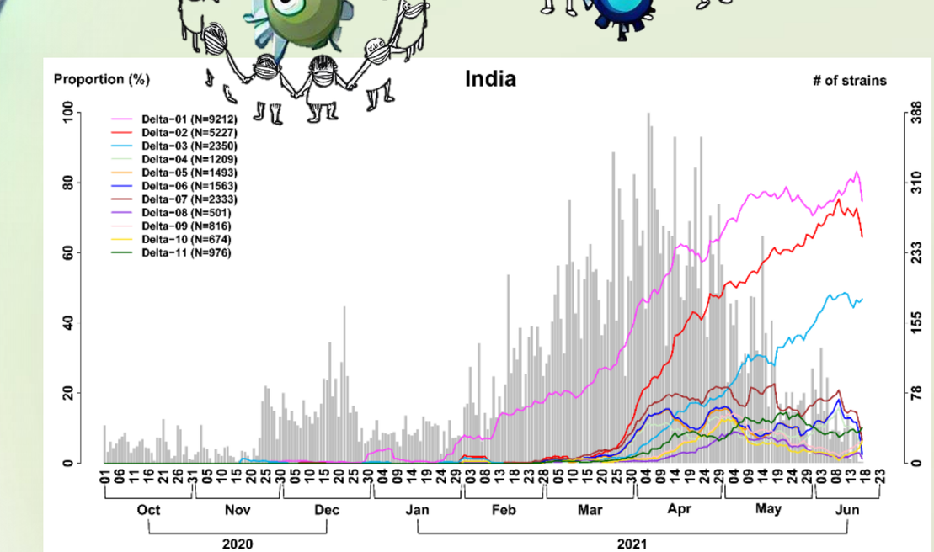
八. 事後分析
1. 波峰識別
2. 濃度校準

病毒基因體大數據研究 楊智婷

自 2019 年 12 月首次確診案例以來，COVID-19 便迅速在全球蔓延，世界衛生組織 (WHO) 隨即宣布提升其為國際公共衛生緊急事件。歷時逾兩年的疫情爆發，截至 2023 年 5 月，WHO 宣布 COVID-19 作為一個國際關注的突發公共衛生事件已經結束。全球已有 7.6 億人感染該病毒，造成 6.9 百萬人死亡。為了強化疫情防治，各國採取多種策略，包括 COVID-19 基因體序列分析



新冠病毒是屬於一種RNA病毒，當病毒進入細胞後，複製基因的過程容易出錯，使單一核苷酸發生改變 (SNV)，事實上可以根據其基因序列上的 SNVs 再細分成不同型態病毒株。本團隊開發了一套巨量基因體資料降維方法，使用僅由少數高相關 SNVs 組成的集合 (Correlated SNV set, CSS) 來代表大部分的病毒類型，能有效地在短時間內將成千上萬隻病毒株進行分類。不需考慮全基因序列長度 29,903 個核苷酸，可以使用少數幾百個 SNVs 組成的 CSSs 來解釋這一千萬株的新冠病毒株。



本團隊進一步發現同一型態的病毒株都有發展出子型，其子型傳播力也有所不同，結合最佳化理論，開發自動偵測具有傳播優勢及傳播劣勢的子型的方法，同時定義可能增加傳播優勢的 SNV (Transmission enhancer SNV) 與抑制或減緩傳播的 SNV (Transmission suppressor SNV)，用以解釋同一種病毒株不同子型動態傳播。

疫苗研發設計需要考慮到病毒的變異情形，新型態病毒可能會引起「免疫逃脫」，意旨原先在我們體內的抗體，無法辨認出新型變種病毒，這也是為什麼建議打第三劑疫苗的原因。

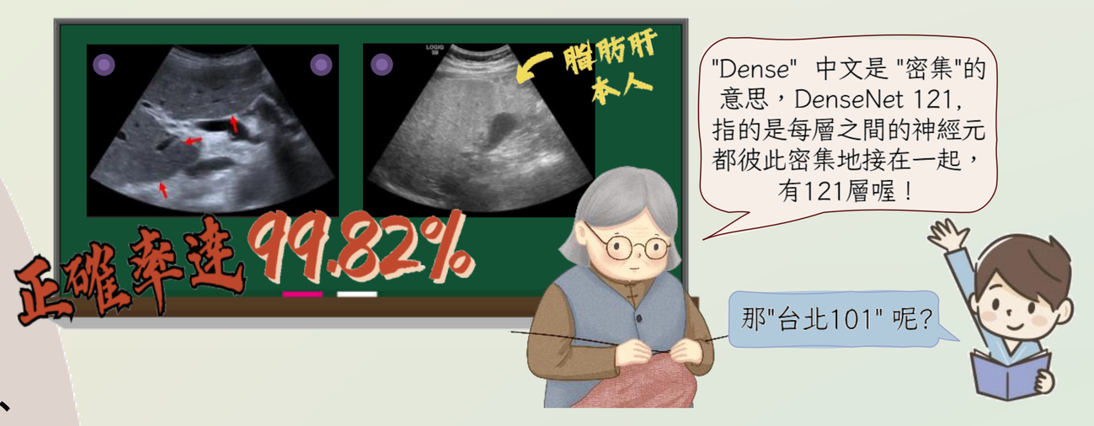
偵測變種病毒將病毒分類，能讓研究者更有效地了解病毒，進行藥物研發或快篩試劑開發。政府亦能針對不同變種病毒，採取不同防疫措施，超前部署將病毒擋於國門外。此外，除了比對確診者足跡，通過基因序列比對亦助於確診個案疫調，確認確診者間的關聯性。

為了持續監測病毒的變異情況，中研院的研究團隊建立「病毒變異全球即時監測網」，用於記錄病毒株隨著時間的變化以及持續追蹤新興的變異位點。https://sarscov2.sinica.edu.tw/

脂肪肝研究 李易儒

運用腹部超音波的影像研究

- 使用兩萬多張腹部超音波，去除影像之外的標註訊息。
- 腹部的臟器有很多，接著尋找有拍到肝臟的影像。
- 依據醫師給的文字說明檔，分別找出有脂肪肝和健康肝的影像作為訓練材料。
- 接著運用深度學習技術，建立 *DenseNet 121* 分類模型。



- 全世界非酒精性脂肪肝盛行率約25%。
- 台灣40歲以上成人，約60%有脂肪性肝臟疾病。
- 脂肪肝若置之不理，可能發展成不可逆的肝纖維化、肝硬化、甚至是肝癌。
- 脂肪肝沒有病徵，也沒有特定的藥物治療，僅能透過改變生活習慣改善。

運用基因體資料的GWAS研究

GWAS（基因組關聯研究）是一種科學方法，用於研究基因和特徵、疾病之間的關聯。透過比較有無脂肪肝的人之間的基因差異，有助於找出哪些基因變異與脂肪肝有關。

我們所找到的基因 PPNLA3，在過去被發現與非酒精性脂肪肝有關。

未來我們將計整合基因與影像，運用人工智慧達到精準醫療。

與國內各大醫學中心/醫療院所頂尖大學/研究單位跨域合作

① 醫學影像大數據 ② 基因體醫學與精準醫療大數據 ③ 各式體學大數據 ④ 環境與行為醫學大數據 ⑤ 各類型巨量數據品質管制

運用智慧科技為人類健康及疾病提供精準的

預防 診斷 治療 照護

我們關心的疾病有

眼科疾病 骨質疏鬆 乳癌 阿茲海默症 糖尿病 代謝疾病 大腸直腸癌 心血管疾病 等

資料分類/分群、影像分割、物件偵測、物件辨識、事件預測、自然語言處理等模型

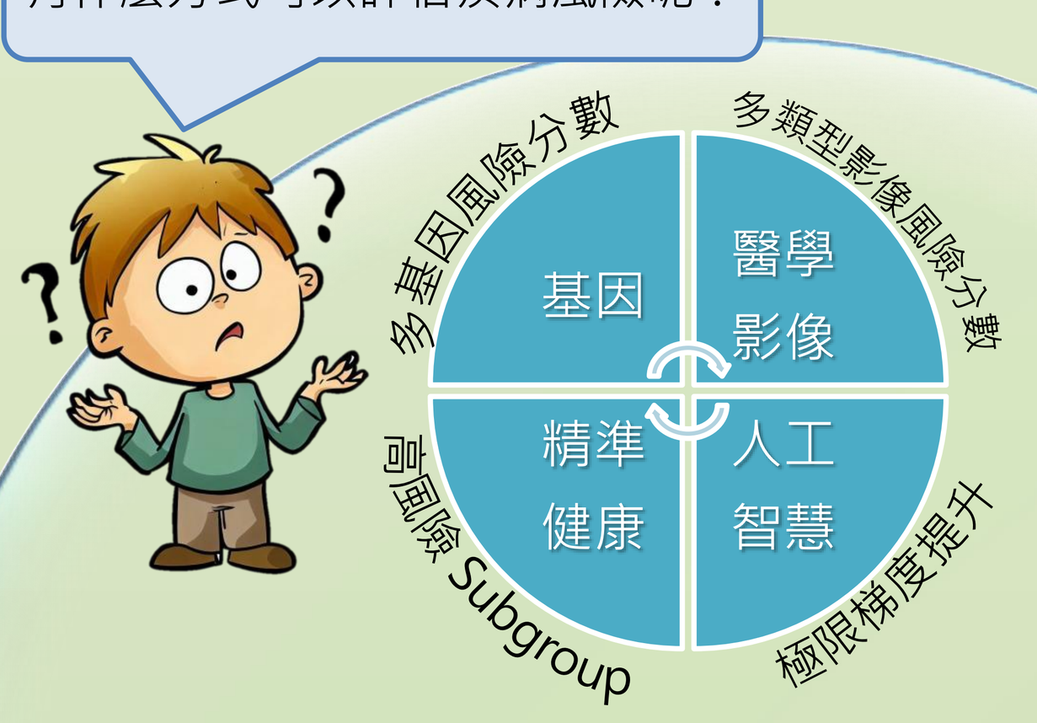
臺灣 Taiwan Biobank

生活及環境問卷 身體檢測資料 基因資料 醫學影像資料 Image Report

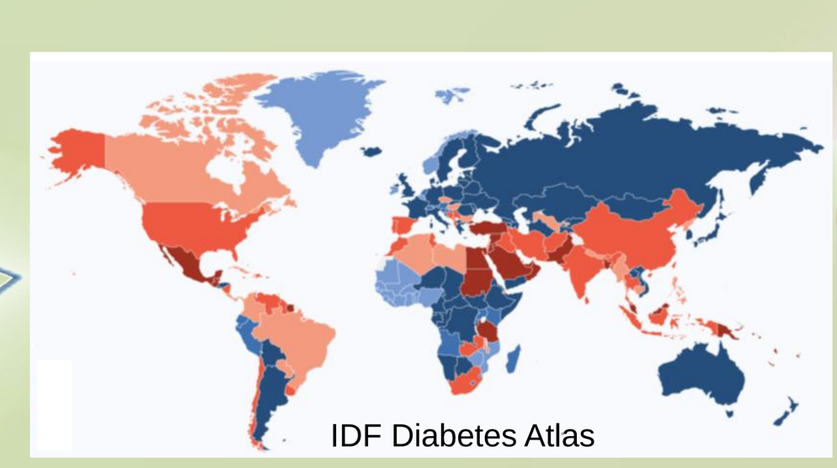
ABD	BMD	ECG	VAS	THY
腹部超音波	全身骨密度圖	心電圖	頸部超音波	甲狀腺超音波
8 項影像報告項目 例如：腎臟正 / 異常等	308 項影像報告項目 例如：T-score, Z-score	1 項影像報告項目 例如：ECG總結	54 項影像報告項目 例如：ICA、CCA等	15 項影像報告項目 例如：結節、大小等
以 DenseNet121 模型分類脂肪肝腹部超音波，正確率達 99.82%	以腹部及脊椎影像搭配 DenseNet121 模型分類骨質疏鬆，正確率達 98.35%	以心電圖分解技術搭配深度學習模型進行自我陳述疾病預測，火熱進行中	以彩色都普勒影像搭配 DenseNet121、AlexNet 模型分類異常頸動脈，進行中	以各式深度學習模型進行甲狀腺結節偵測，IOU達 91.0%

糖尿病精準醫療 黃怡嘉

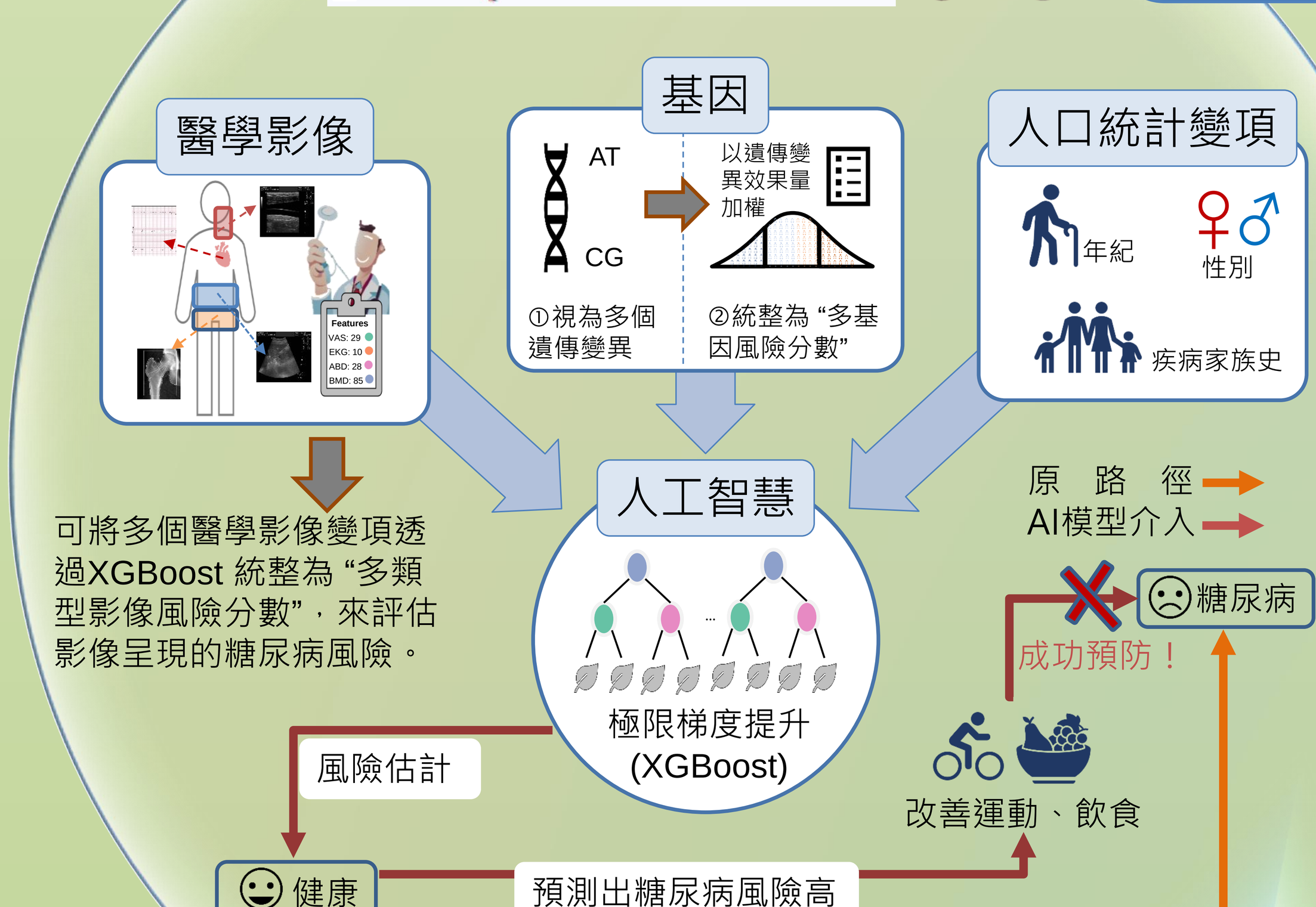
有什麼方式可以評估疾病風險呢？



簡介：我們分析臺灣人體生物資料庫 (Taiwan biobank; TWB) 大數據，建立多基因風險分數 (Polygenic risk score; PRS)、多類型影像風險分數 (Multi-image risk score; MRS)，並透過近年來非常熱門的人工智慧 (artificial intelligence; AI) 方法，建立第二型糖尿病(Type 2 Diabetes; T2D) 早期偵測模型，往精準健康 (Precision health) 邁進！



糖尿病為血液中糖份過高，其發生和多種因素有關，包含遺傳、肥胖、飲食等。在全世界盛行，如果能在還沒發生時，早期偵測風險，是不是就能預防或延緩發生？AI 應該有辦法做到這個吧！



風險特別高族群，需特別注意T2D預防方式！！

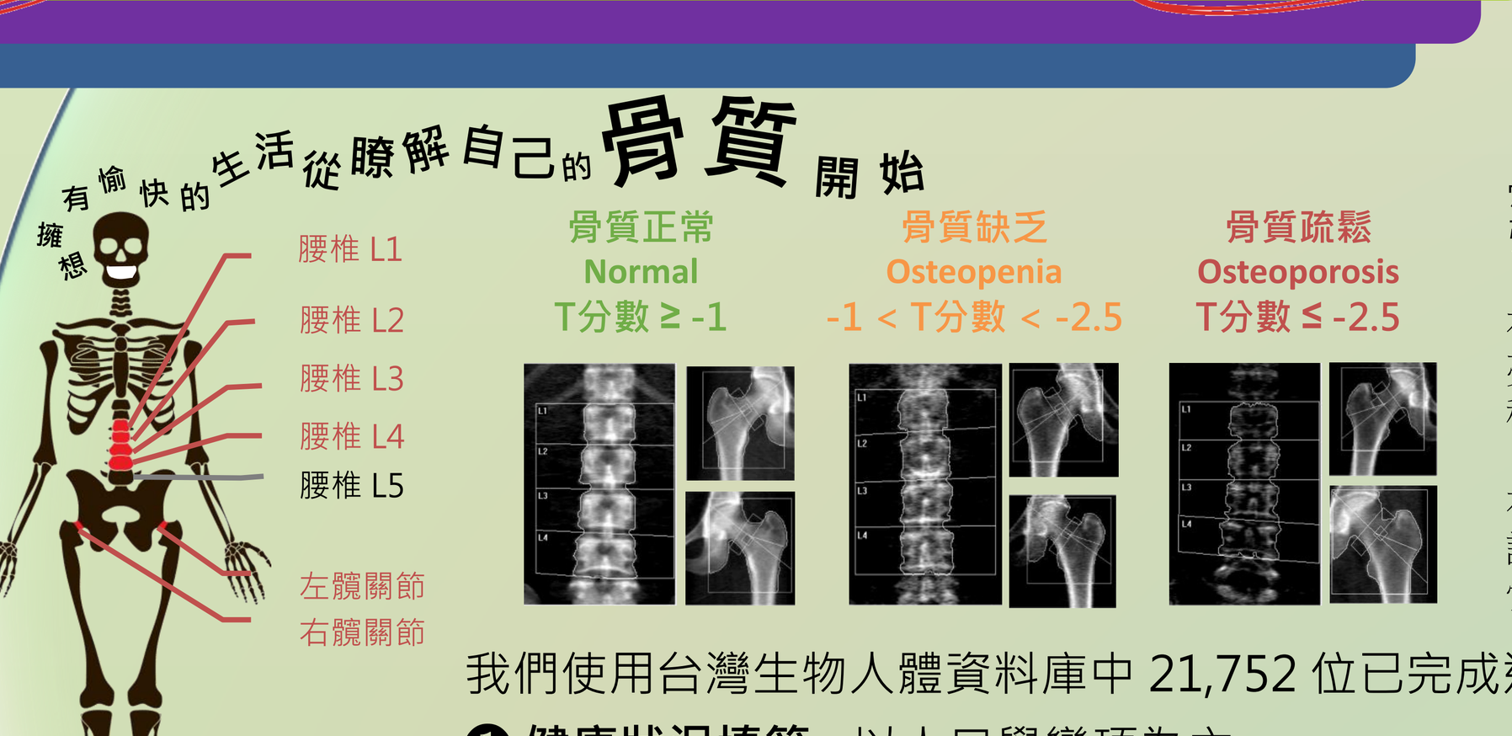
- 女性、年紀大於59歲、有T2D家族史且多基因風險分數高時。
- 男性、年紀大於54歲、有T2D家族史且多類型影像風險分數高時。

T2D 風險評估網頁

Website: https://hcyang.stat.sinica.edu.tw/software/T2D_web/header.php

骨質疏鬆研究 朱昱蓁

有哪些快的生活從瞭解自己的骨質開始



骨質疏鬆是由大規模骨質流失引起的疾病。

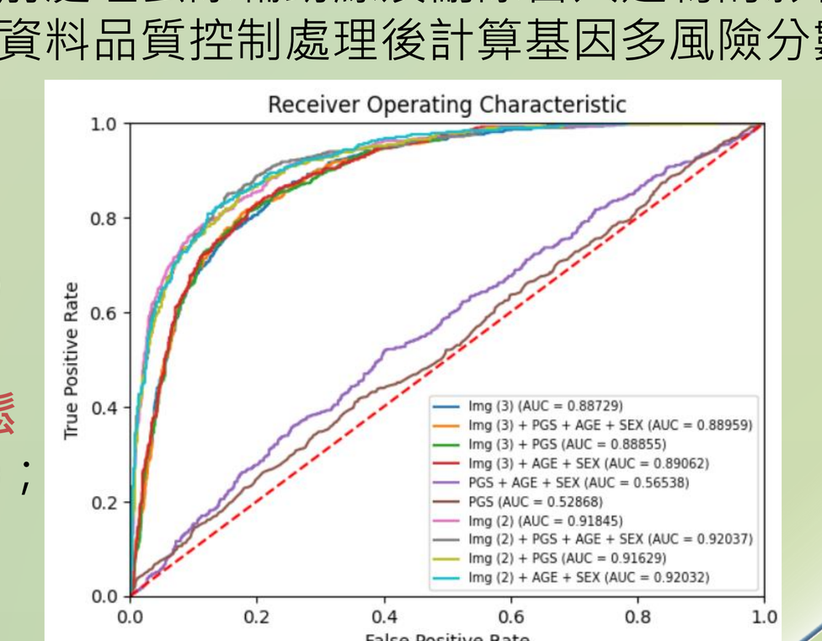
在 DXA 骨質影像上，可能藉由亮(骨頭)、暗(骨質疏鬆或低骨密度)程度瞭解骨質情形。

在臨床診斷上，需進一步藉由影像計算骨質密度，再使用T分數評估骨質情形。

我們使用台灣生物人體資料庫中 21,752 位已完成追蹤調查參與者的

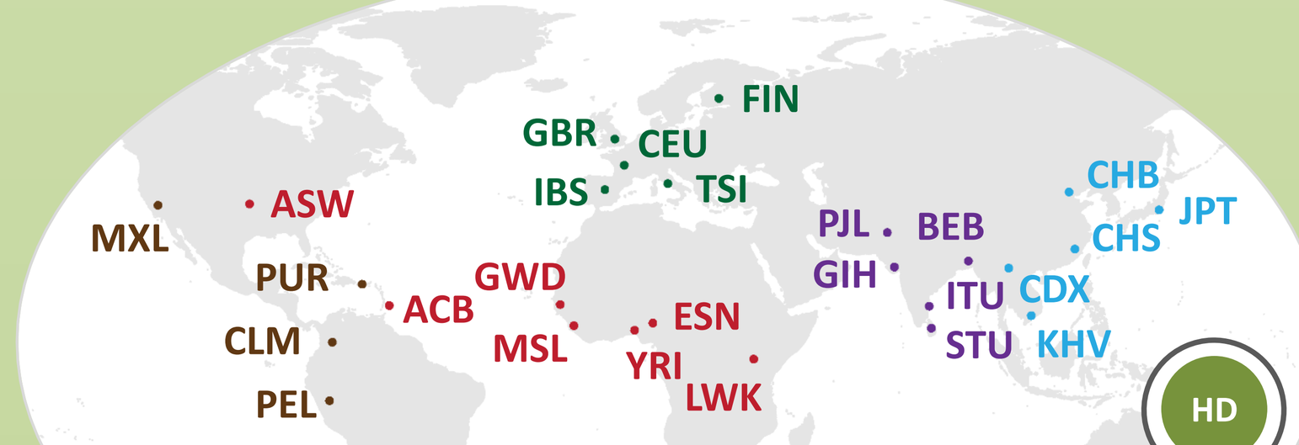
❶ 健康狀況填答　以人口學變項為主
❷ DXA 骨質影像　經過前處理去除輔助線及刪除含人造物的影像
❸ 基因資料　經過資料品質控制處理後計算基因多風險分數

- 資料集❷ 使用 DenseNet 121 預測骨質正常 骨質疏鬆 準確率達 98.35% (AUC: 99.75%)
- 資料集❷ 使用 DenseNet 121 預測骨質正常 骨質缺乏 骨質疏鬆 準確率只 85.55% (AUC: 91.85%)；再新增資料集❶❸，其改善幅度有限(右圖)



全球族群藥物基因體 陳佳煒

整合 Drug Bank、PharmGKB、PharmaADME、Biotransformation 中關於 PGx 的資訊，包含：功能註解 (function annotation)、藥物副作用 (ADR)、藥理效應 (FX)、藥物動力學 (PK)、藥效學 (PD) 等



使用千人基因組計劃 (1000 Genomes Project) 資料針對洲際間 (不同顏色) 或 族群間 (同顏色) 利用

1. Fisher's exact test 尋找 SNV 等位基因 (allele) 分佈有所差異者
2. Kruskal-Waills test 尋找基因同型合子不平衡 (homozygosity disequilibrium) 分佈有所差異者