

出國報告（進修）

2024 美國國家衛生研究院精準醫學 進修心得

服務機關：高雄榮民總醫院/健康管理中心

姓名職稱：吳品潔/主治醫師

派赴國家：美國

出國期間：2023/08/01-2024/07/31

報告日期：2024/09/27

摘要

近年來，精準醫療領域迅速發展，臨床應用範疇也隨之不斷擴大，逐漸成為現代醫學的重要核心之一。隨著基因定序技術成本的顯著下降及其效能的日益提升，基因定序在臨床診斷、治療選擇及疾病風險評估中的應用越來越廣泛。為了更深入地了解這一領域的最新發展，職前往美國國家衛生研究院進修一年學習定序技術的應用及相關研究。此次進修學習基因資料在精準醫療中的應用，並了解現階段此技術與數據所面臨的限制與挑戰。這些學習將有助於我在臨床研究與健康管理上的應用。

關鍵字 (至少二組)。

精準醫學、基因定序

目次

一、目的.....	4
二、過程.....	4
三、心得及建議.....	19
附錄.....	22

一、目的

精準醫學旨在根據個人的基因、環境和生活方式的差異，制定個別化的預防、診斷和治療方案。與傳統醫療通常針對大多數人的共性，採用標準化的治療方式不同。近年來精準醫學得益於定序工具、數據分析技術和人工智慧的進步而快速發展。隨著基因檢測成本的降低和醫療研究數據的累積，已從實驗室研究擴大至臨床應用，除了應用於疾病的診斷和治療方面，也用於幫助識別個人患病的風險，從而制定個別化的預防計劃。美國國家衛生研究院（National Institutes of Health, NIH）是全球的首屈一指的生物醫學研究機構，職任職於健康管理中心，希望透過在美國國家衛生研究院的進修，學習基因資料分析處理與疾病風險評估，將精準醫學運用到健康管理與研究上。

二、過程

此次進修的單位是美國國家衛生研究院的神經疾病與中風研究所(NINDS)，向范揚政教授及他的團隊學習。范教授是 NINDS 資訊技術與生物資訊部門的主任（Intramural IT and Bioinformatics Program Director）。他負責的任務包含帶領生物資訊研究團隊，協助國家神經疾病和中風研究所（NINDS）的生物資訊相關研究工作，例如：實驗設計、次世代基因定序、定序之序列數據處理與統計數據分析等領域。他也同時領導人工智慧團隊，在機器學習、自然語言處理與大型語言模型等領域的技術研究應用。此次進修的領域與范教授的生物資訊團隊相關，因此透過范教授的安排開啟了此次的學習之旅。

2023 年 8 月-9 月

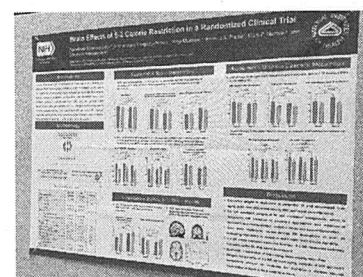
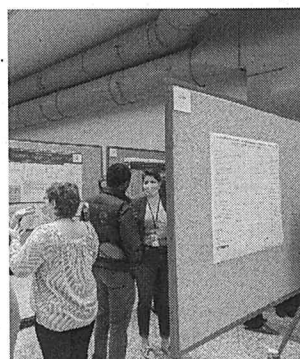
參加每週團隊會議

范教授每週都會召開團隊例行會議，會議的內容除了介紹部門內新進成員外，也讓團隊成員報告各自負責的計劃進展。若在執行計劃的過程中遇到問題，大家可以在會議中提出，透過集體討論共同找出解決問題的可行方案。由於我當時身為新進人員，最初主要是聆聽其他團隊成員的報告，觀察學習大家的工作方式。在幾次會議後，我在會議上向范教授表達了自己想學習精準醫學的興趣，特別是在次世代基因定序、定序後的數據處理以及基因風險評估這些領

域。范教授提供了多種學習資源，包括院內舉辦的相關研討會、FAES(The Foundation for Advanced Education in the Sciences)所提供的專業課程，以及如何利用基因資料庫進行研究等。此外，也協助我與生物資訊研究團隊中的專家聯繫，指導我基因數據處理。透過這些多元的學習管道，我得以逐步學習精準醫學相關內容。

NIH Research Festival

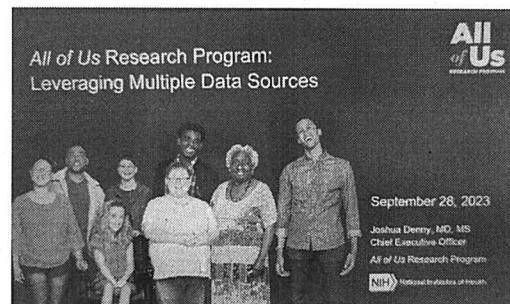
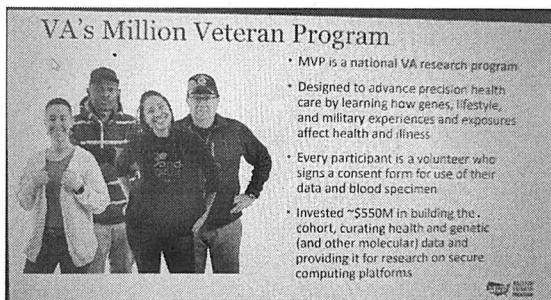
NIH Research Festival 是 NIH 定期舉辦的院內研究計劃的最新進展與成果，並促進研究者之間的學術交流與合作。在這個 Research Festival 中，展示許多與生物醫學相關的研究項目。由於約翰霍普金斯醫學院坐落於巴爾的摩，距離 NIH 僅約一小時的車程，因此地理上的便利性促使許多來自約翰霍普金斯醫學院的醫師、老師以及學生參與。海報內容多偏向基礎研究，但也有部分為臨床研究報告。



參加 NIH 舉辦的工作坊: Big Data Integration: Unlocking the Potential for Enhanced Epidemiological Research

這個工作坊是由 NIH 的國家心肺血液研究所 (NHLBI)、國家癌症研究所 (NCI) 及國家糖尿病、消化及腎病研究所 (NIDDK) 共同舉辦的研討會，主要是分享如何整合大數據以促進流行病學研究。隨著醫學研究領域數據量的迅速增長與數據處理能力的進步，流行病學研究大幅擴展。這些數據可用於風險預測、慢性病診斷及治療。然而，管理這些龐大且多樣化的數據面臨重大挑戰，特別是在整合不同來源的數據、數據品質與可靠性皆是重要議題。在這會議中介紹使用大數據庫所做的研究，包含 VA's Million Veteran Program (MVP), All of Us, The Atherosclerosis Risk in Communities (ARIC) study 等。MVP 是一項全國性的研究計劃，旨在探討基因、生活方式、軍事經歷和暴露因素如何影響退伍軍人。All of Us 旨在收集並研究

來自美國境內百萬人以上的數據，其資料庫內容涵蓋電子病歷、問卷、理學檢查、生化資料、穿戴式裝置之資料與基因資料。ARIC 於 1985 年開始，是一個以社區居民為主的前瞻性計畫，目的是了解心血管疾病及其危險因素在一般人群中的發展過程，這個研究追蹤至今仍持續進行，陸續增加各類型的資料型態。



2023 年 10 月-11 月

FAES 課程-Leveraging High-Throughput Sequencing in your Research 與 Bioinformatics for Analysis of Data Generated by Next Generation Sequencing

於每週的團隊會議，教授建議可以上 FAES (The Foundation for Advanced Education in the Sciences) 的課程。FAES 成立於 1959 年，是美國國家衛生研究院內的一個非營利組織，主要提供科學教育。它提供的學程涵蓋廣泛的主題，包括生物醫學、程式、數據分析及管理等。我參加 FAES 開立的 Leveraging High-Throughput Sequencing in your Research 與 Bioinformatics for Analysis of Data Generated by Next Generation Sequencing。

Leveraging High-Throughput Sequencing in your Research 這個學程是介紹高通量定序技術 (high-throughput sequencing) 的理論與實際應用，涵蓋現代定序技術 (如 Illumina、PacBio、Oxford Nanopore) 的介紹及其在 RNA 與 DNA 研究中的應用，並分析不同定序方法的優點與限制。

Bioinformatics for Analysis of Data Generated by Next Generation Sequencing 這門課程中，介紹如何分析由各種定序技術 (如 DNA 定序、RNA 定序和 CHIP 定序) 所產生的數據並應用到生物醫學研究中，如基因表現分析及與醫學相關的序列變異識別。課程中老師會先給予講座和閱讀資料提供背景知識，之後為實際操作。老師提供真實數據的下載方式，我們利用這些數據實際操作定序分析的主要步驟。以 DNA 序列為例，下載數據後將數據上傳至 Galaxy

平台，在這平台中先做定序品質分析，試圖將品質低的資料過濾，留下品質好的資料往下做 alignment. Alignment 是將我們的數據 mapping 到參考基因，若是人類的基因，就使用人類的參考基因作為模板進行 mapping. 再做 variant calling.

圖片:擷取自上課講義

FAES 的課程多為數週，除了上課外，也需要繳交作業與測驗，在課程完成後，會告知是否通過，並給予證明。

Confirmation of Completion

Student:	Pin Chieh Wu
Course:	GENE 530 - Section B (Three Weeks) Leveraging High-Throughput Sequencing in your Research
Instruction Method:	
Duration:	25/Oct/2023 to 14/Nov/2023
Total AUs:	1.0

It is hereby confirmed that the above-named individual has been accepted in studies at this institution as indicated above.

NHGRI Didactics genetics counseling

這是美國國家人類基因組研究院 (NHGRI, National Human Genome Research Institute) 定期召開的會議。NHGRI 是 NIH 的其中一個部門，主要是基因組學領域的基礎研究，開發新技術來研究基因組、揭示遺傳變異與疾病的關聯性、以及將基因組學技術應用於臨床醫療。經教授的幫忙，我得以每週參加。

這會議形式多樣，包含期刊選讀，探討近期發表的學術文章、個案報告、或基因組學技術實際應用於臨床個案的經驗等。參加者主要是醫師、基因諮詢專家、以及進修學者。通常是由 NIH 的醫師報告所屬團隊的個案，做個案分享與討論。透過這個會議，讓我對於基因組學應用於病患的診斷和治療計劃中有較多認識。

NHGRI Clinical Update Meeting

這個定期會議也同樣是由 NHGRI 所舉辦，會議內容為該單位內每週的住院病人或就診病人的討論。於 NIH 就診的病人皆是透過參與研究計畫而來，全球有符合條件者都可以提出申請，因此個案大多為較不常見的疾病或其他醫院的困難個案轉介。由於個案遍及全球，因此在 NIH 內也免費提供飯店供病人住宿。在 NHGRI 的病人大多與基因組學和遺傳學相關，NIH 的專家們

透過這些患者的臨床表現、基因資料，探討基因變異對健康的影響。同時也會納入患者家人之病史與基因資料，一同評估。他們的團隊中也包含基因諮詢師與臨床遺傳學家，提供基因遺傳諮詢，疾病診斷。

NHLBI Obesity, Nutrition, and Physical Activity Seminar: The Relationship between Body Weight, Adiposity, and Cardiovascular Risk among South Asian Americans

在這場研討會中，Namratha Kandula 醫師介紹了「南亞裔美國人體重、體脂與心血管風險之間的關係」的研究。體重與多種已知的心血管風險因子密切相關，但不同種族和族群之間，體重與心血管風險的關聯存在差異。南亞裔美國人即使在體重偏低的情況下，仍面臨較高的心血管疾病風險。雖然他們的 BMI 不高，但相較於其他族群，內臟脂肪比例較高，肌肉質量則相對較低。即便在調整 BMI 後，內臟脂肪仍然是血管鈣化的獨立危險因子。南亞裔美國人的肝臟脂肪含量普遍高於歐裔。未來研究可通過 multi-omics 方法進一步探討肥胖與內臟脂肪在南亞裔美國人心血管疾病的相關機轉，為降低南亞裔美國人的心血管疾病風險提供更多證據。

2023 年 12 月

All of Us 資料庫申請

除了持續參加定期的課程與會議之外，開始 All of Us 資料庫的學習。All of Us 是 NIH 的研究計畫，目標為收集超過一百萬人的健康數據，包含來自不同種族、民族、年齡、地區和健康狀況的參與者。到 2023 年 2 月 15 日為止，All of Us 共有 409,420 位參與者。收集內容包含基因、病史、生活方式和環境暴露。因資料庫中含有多樣的資料型態與基因資料，因此向教授提出使用 All of Us 的想法，希望能使用這個資料庫實際操作分析基因資料。教授非常幫忙，立即連絡單位負責人員，申請 All of Us 的使用權。

使用 All of Us 前須先閱讀並通過資料保密與倫理之測驗。測驗完成後才能進入資料庫。All of Us 的資料依使用權限分為兩大類，一類是 Registered Tier，一類是 Controlled Tier。Registered Tier 包括來自電子記錄 (EHRs)、可穿戴裝置和問卷資料，以及參與者在註冊時測量的身體量測數據。Controlled Tier 除了上述資料外，還包含基因資料。我申請的是

Controlled Tier。

在資料的取得與分析過程中，資料庫內的數據本身是無法直接下載的，能夠下載的僅是經過分析後的結果或圖檔。Register Tier 資料提供了點選式選擇功能，使用者可以勾選需要的資料。然而，由於資料量龐大，且這些資料來自多家醫療機構或保險公司，因此需要花費一定的時間熟悉資料庫，才能有效找到所需的族群及其相關數據。

關於基因資料的部分，研究者必須使用平台上支援的程式語言來下載這些資料，同時，資料分析也需要依賴程式進行。在 All of Us 平台中，最常使用的程式語言包括 R, Python, SAS 等。平台提供了線上教學影片，介紹常用功能的操作。此外，每週還提供線上諮詢服務，使用者可以在會議前上傳問題，並在會議中獲得針對操作問題的回覆，但這些諮詢主要集中在介面操作，並非實驗設計。

由於上述學習資源較為分散，再加上缺乏使用程式處理資料的經驗，因此教授請求 Hirotaka Iwaki 醫師提供協助。Hirotaka 是一位經驗豐富的神經內科醫師與研究員，隸屬於 NIH Intramural Research Program Center for Alzheimer's and Related Dementias (CARD) 實驗室。他在基因大數據分析方面具有豐富的經驗，並擔任數據分析工作小組的共同主席。

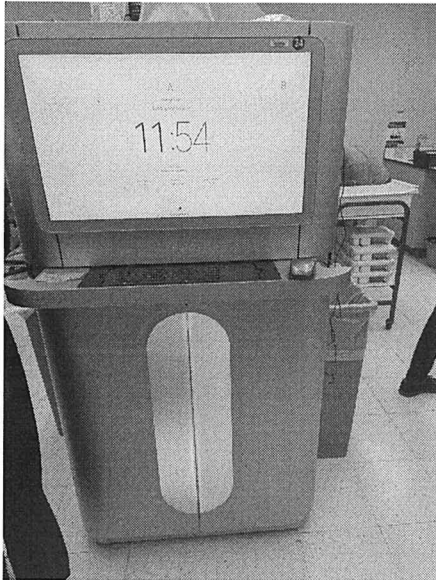
2023/12/5 參加 2023 年終歲末大華府地區研究學人交流座談餐會

這天參加駐美國台北經濟文化代表處科技組(華盛頓特區)為大華府地區舉辦的研究學人交流餐會。這次餐會的參與者包含華盛頓特區科技組楊琇雅組長及她的團隊、中研院莊德茂院士、何英剛院士、美國國家標準與技術研究院宋麗品博士、章以本教授、陳俊宏教授、李偉智教授、國科會龍門計畫與千里馬計畫來美的進修人員、博士後研究人員、在 NIH 的進修人員等。院士、教授與博士們皆是長期在美促進台美交流的專家，透過餐會與大家交流職涯規劃。

DNA sequencing and genomics core 實驗室

這個 DNA 定序與基因組核心實驗室成立於 2010 年，提供 NIH 研究人員 DNA 定序和基因組分析服務，能夠分析大規模的 DNA 數據。實驗室內會將定序的資料作處理，再以研究人員可分析之資料型態傳回給研究人員。實驗室內的儀器不斷更新，從最新型高吞吐量基因定序系統 NovaSeq X, 中高等吞吐量基因定序系統 NovaSeq 6000, 到小吞吐量基因定序系統

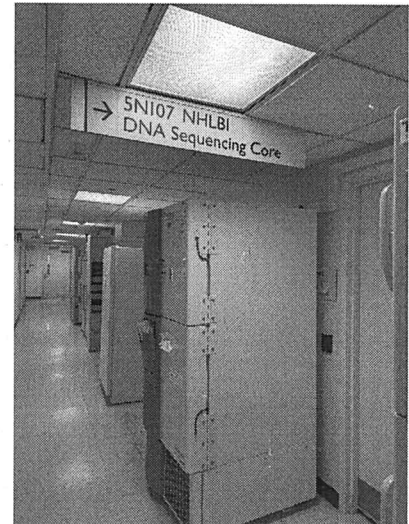
MiniSeq 及 iSeq 100 皆有。實驗室的博士提到，使用 NovaSeq X 或 NocaSeq 6000 對人類全基因定序，約 48 小時內可以完成。且 NovaSeq X per dual flow cell run 運行時，一次最多可生成 16 TB 的數據，生成 520 億個 reads。這些儀器的購入價、維護成本都非常昂貴，所以儀器幾乎都處於運轉的狀態。



(NovaSeq X)



(NocaSeq 6000)



FAES 課程 Bioinformatics for Analysis of Data Generated by Next Generation Sequencing
完成

Confirmation of Completion

Student: Pin Chieh Wu
Course: BIOF 521 - Session B Bioinformatics for Analysis of Data Generated by Next Generation Sequencing
Instruction Method:
Duration: 25/Oct/2023 to 12/Dec/2023
Total AUs: 2.0

It is hereby confirmed that the above-named individual has been accepted in studies at this institution as indicated above.

2024 年 1 月

The Prospect of Precision Medicine for All of Us 研討會

這個研討會主要兩位專家：American Society of Human Genetics 前任主席 Dr. Biesecker，以及 Dr. Denny (NIH All of Us 研究計劃的首席執行官，並擔任 NHGRI 精準醫學負責人) 的分享。對於疾病的鑑別診斷存在許多挑戰，尤其是罕病。一位醫師很難同時熟悉多種罕病，且並不是所有帶有該基因變異者就會得病，也非每一個患者都有相同的症狀，這時就需要一些工具來協助診斷。過往由於基因定序成本高，病患大多先由就診的醫師決定做那些基因檢測，直到找到答案為止。而現在由於基因定序成本的下降，診斷的流程可以做改變。個人可於年輕時即先做好基因的定序資料，對於無症狀且對日後健康有大影響的疾病可以提早做預防與治療，後續若出現疾病症狀，即可用過往已經定序過的資料去做比對，找出致病的基因變異，這樣可以減少疾病診斷的時間。此外大數據的基因研究，可讓基因與疾病間的關聯性更明確。秉持這個想法，NIH 持續有計畫做此方面的努力。All of Us 即是秉著這個概念這個研究計畫。除了收集資料外，參與者會收到個人的基因報告，以提升參與者的意願。

Using Informatics and Generative AI to Support Patients' Understanding of Lab Test Results

這個研究是針對老年患者了解自身身體檢查數據結果時的困難，並藉由分析結果設計 LabGenie 這個工具，希望能提升老年患者對數據檢查結果的理解。整個研究分為三個階段進行，第一階段透過調查患有慢性病的患者群體，瞭解影響他們理解檢測結果的關鍵因素；第二階段利用數據標註和大型語言模型，深入分析健康相關文章中的關鍵字與術語；第三階段則評估了四種生成式人工智慧模型（包括 GPT-4、MedAlpaca、Llama 2 和 ORCA_mini）在回答患者問題上的準確性與實用性。由這項研究可以看到生成式人工智慧在醫學中的應用。AI 技術不僅能快速處理和分析大量的醫療信息，還能根據個別患者的背景提供客制化的解答，這不僅改善了患者的數據解讀能力，還促進了個人健康管理。

FAES 課程- Introduction to Medical Genetics II: Clinical Concepts

這個 14 週的課程主要是介紹基因相關的多種疾病與症候群的臨床表現。課程內容按照人體不同系統進行分類，如循環系統、呼吸系統等，並邀請相關科別的專科醫師針對各系統進行講解。課程從胚胎學開始，說明胚胎發育過程中的問題導致胎兒出生時出現器官缺陷，這類發育異常往往與特定的基因變異有關。一些基因已被證實在胚胎發育過程中對器官的形成具有重要作用，當這些基因發生變異時，就會引起該器官的發育問題。課程會詳細介紹這些基因與器官發育異常之間的關聯。

有時，基因的變異不僅影響單一器官，還會涉及多個器官，導致綜合性的症候群，這類基因變異所引發的症候群也在課程中介紹。除了發育異常，課程也會介紹與單一基因變異相關的腫瘤。人體各系統相關的基因變異與症狀資訊繁多，學習這些知識除了需要掌握理論，臨床應用也十分重要，透過理論的了解與臨床使用，能鞏固知識並提升對病患的診斷能力。因此，這門課有許多準備參加基因諮詢師考試的學員。

2024 年 2 月

Introduction of terra data repository 工作坊

Terra Data Repository(TDR) 是一個數據儲存與分析的雲端平台，主要用來支援生物醫學和基因組學領域的大數據研究。它與 Google Cloud Platform (GCP) 和 Microsoft Azure 有緊密的合作，透過 GCP 和 Azure 來提供數據儲存、計算資源和數據管理服務。All of Us Research Program、NCI Genomic Data Commons、AnVIL (Analysis, Visualization, and Informatics Lab-space)、dbGaP (Database of Genotypes and Phenotypes)等多個 NIH 的計畫皆可運用此平台做大數據的資料處理。

這個工作坊除了介紹 TDR 的功能外，也同時介紹 Jupyter Notebook。Jupyter Notebook 是一免費的軟體，適用於多種程式語言的數據計算，為 Terra 平台內建的數據分析工具，允許使用者直接在雲端中對數據進行分析和建模。All of Us 便是透過 TDR 平台與 GCP 合作，儲存大量的資料，並讓使用者使用 Jupyter Notebook 做數據分析，並可即時查看分析的結果。工作坊中示範了如何使用 Jupyter Notebook 的基本功能。

NHLBI Obesity, Nutrition, and Physical Activity Seminar: Dietary Biomarkers and Cardioprotective Mechanisms of the DASH Diet

這個研討會在探討肥胖與 DASH 飲食相關的 biomarkers. DASH 飲食 (Dietary Approaches to Stop Hypertension) 是一針對預防和控制高血壓的飲食模式，已被廣泛證實有助於降低血壓並促進心血管健康。這種飲食強調多攝取蔬菜、水果、全穀類、低脂乳製品及瘦肉，並限制高鹽、高糖和飽和脂肪的攝入量。以往做的研究大多以受試者自我回報的飲食結果作為數據，然而這些回報確實會導致資料庫資料並不十分準確，也因此想要發展較客觀的數據來評估飲食的狀態。代謝體是可做為飲食客觀數據的方式之一，因此其收集血液、尿液等檢體做了代謝產物的分析，並找出代謝產物與食物種類的關係，使得之後有機會可以使用這些 biomarkers 做為飲食的客觀數據。

Prioritizing genetics to reduce existing health disparities

這場會議介紹將基因資料(Polygenic Scores, PGS)與電子健康記錄 (EHR) 數據相結合，包含透過結合門診和住院中收集的實驗室檢測結果，有機會可以發現複雜疾病新的生物標誌物 (biomarkers)。另外，隨著種族的不同，透過這些資料，也可以試著找出不同種族間的正常值範圍，提供更精準的數值診斷。然而這些數據雖具有巨大的潛力，但其在過去的基因研究中應用卻有限，主要原因在於數據質量不一致及高維醫療數據的分析與解釋難度。這些挑戰包括數據輸入錯誤、不一致的單位和參考範圍，以及重複記錄的檢測結果，這些因素增加了數據處理的複雜性。即便如此，EHR 實驗室數據仍然是臨床研究中的重要資源，特別是在遺傳學研究領域。隨著技術和數據處理方法的不斷改進，這些數據在未來具有發展性。

2024 年 3-4 月

All of Us 資料庫的學習與使用

從這個月開始到 5 月初，All of Us 舉辦一系列教學講座。首先，講座介紹 All of Us Research Program 的目標，並指導學員如何註冊並進入 All of Us Researcher Workbench。課程說明平台上數據的編排方式，幫助使用者理解資料的來源和選擇的邏輯。課程也指導學員如何篩選數據，設置篩選條件，並學習如何在 Researcher Workbench 進行編碼。課程涵蓋基

礎編碼的教學，展示如何利用編碼提取與分析資料。然而，要順利操作平台，需要具備基本的程式知識。

過去兩個月中，我逐步學習了 R 和 Python 的基礎概念，並在 Hirotaka Iwaki 醫師的協助下，了解基因資料的分析流程。基因資料須與參與者的基本資料和臨床數據結合進行分析，而在 All of Us 平台上，這兩類資料必須分開提取後再合併分析。通過上述課程的學習，我逐漸掌握了如何選取並使用臨床資料，但提取基因資料是一個更具挑戰性的步驟。除了需要了解基因數據的儲存方式，使用正確的編碼也是一大難關。在基因資料處理過程中，團隊多數人使用 PLINK 軟體來進行基因數據的分析，因此，我須對 PLINK 有一定的了解。掌握 R、Python 和 PLINK 的基礎後，才更能向團隊的專家學習。另外，透過團隊的介紹，我聯繫了 CARD 中心的 Kristin Levine。她是一位資料科學家，她以她的經驗指導我如何使用 Plink 提取基因資料。她的專業協助讓我在提取基因數據的過程中受益良多。

參加 Metabolomics Research in the NIEHS Intramural Research Program

代謝體指的是生物體內所有小分子代謝物的總和，這些代謝物包括糖類、脂質、氨基酸和核苷酸等，參與了生物體的各种生命活動。代謝體學是近年來迅速發展的研究領域之一，它與基因組學密切相關。基因組學提供了生物體的基因藍圖，而代謝體學則展示了基因在不同環境條件下的實際表現。此次會議邀請了來自 NIEHS（美國國家環境健康科學研究所）的學者，分享在 NIH 的研究。研究中運用了高解析度的質譜技術來分析人體代謝體，幫助深入了解人體內的化學變化，並探索環境中小分子化學物質如何影響健康。此外，也專注於改進無標靶分析技術，以便更準確地解讀數據，並辨識出之前未被發現的化學物質。代謝體的重要性在於它能夠反映基因表達和蛋白質功能的最終結果，是生命動態過程的表現。

參加 All of Us Researchers Convention

All of Us Researchers Convention 是一個為期兩天，以分享研究成果為目的的學術會議，重點在於使用 All of Us Research Program 所提供的數據進行研究，並在會議中分享研究經驗與發現。在這個會議中，來自不同領域的研究者展示了他們如何利用 All of Us 的龐大數據集來探索基因、環境和生活方式對健康的影響。因為 All of Us 計劃涵蓋多種族參與

者，這些數據反映真實世界中的多種族群，與不同族群間的差異。這些研究利用基因數據、臨床記錄以及生活方式信息來進行深度分析，提供研究多元方向。此外，會議中的經驗分享，研究人員們討論了在使用 All of Us 數據時面臨的挑戰，如數據清理、隱私保護以及如何將基因數據與電子健康記錄（EHR）有效結合。這個會議讓我對 All of Us 的數據資源的利用有更多的認識。

Researchers Convention | **All of Us**

April 3-4, 2024

Registration for the Researchers Convention is Now Open!
The *All of Us* Researchers Convention will be held April 3-4, 2024. The free, virtual event is open to researchers across all disciplines and career stages. The Researchers Convention provides an opportunity for researchers who use *All of Us* data to showcase their work to others who share their interests in precision medicine.

Register at: ResearchAllOfUs.org/2024Convention

The Black Greek Letter Consortium (BGLC) is a community partner of the *All of Us* Researchers Program.

2024 年 5 月

Transforming Medicine with AI: From TrialGPT to GeneAgent

這個演講在介紹 chatGPT 與人工智慧在生物醫學的應用。TrialGPT 是一種利用人工智慧技術來優化臨床試驗匹配的創新工具，主要根據患者的臨床數據智能化地配對正在進行的臨床試驗，幫助臨床醫生更高效地為患者找到合適的試驗。這對於癌症等需要精準醫療的領域尤為重要，因為找到合適的臨床試驗往往是一個耗時且複雜的過程。TrialGPT 使用自然語言處理技術來分析患者的醫療數據，再經分析數據，減少手動篩選的需求。同時，TrialGPT 也運用自然語言處理技術辨識臨床試驗的條件和要求，根據患者的個體特徵自動匹配最相關的試驗。如此加速從臨床試驗數據庫中篩選出合適試驗的過程，並降低了錯誤匹配的風險。這一工具是與美國國家癌症研究所（NCI）合作開發的。

FAES: Introduction to Medical Genetics II: Clinical Concepts 課程完成與心得

透過這門課程，我認識不同系統的基因性遺傳疾病，並認識到這些疾病在患者中典型和不典型的症狀表現。課程中也探討了基因檢測所帶來的倫理問題，這部分內容讓我印象深刻。當檢測結果顯示某位父母攜帶致病基因時，醫療團隊如何將這些資訊告知父母，往往成為一個具挑戰性的問題。攜帶致病基因的父母可能會感到強烈的自責和悲傷，也可能因此而受到指責，尤其是當這些基因變異導致孩子出現嚴重的健康問題時，其心理負擔可想而知。課程分享遺傳諮詢師在提供心理支持和專業建議時的重要性。在課程的最後一堂，特別邀請了遺傳疾病患者親自到場分享經歷。透過患者的親身講述，更能體會這些疾病對其生活的巨大影響。

Confirmation of Completion

Student: Pin Chieh Wu

Course: GENE 500-II - Session A and B Introduction to Medical Genetics II: Clinical Concepts

Instruction Method:

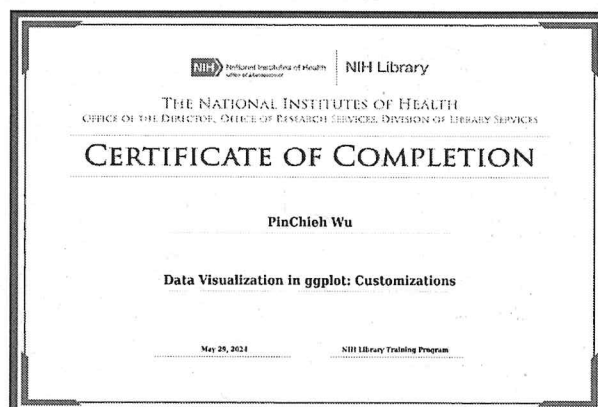
Duration: 31/Jan/2024 to 14/May/2024

Total AUs: 1.0

It is hereby confirmed that the above-named individual has been accepted in studies at this institution as indicated above.

參加 NIH 圖書館舉辦的課程: Data Visualization in ggplot: Customizations

這門課程主要是 R 語言 ggplot 的實際操作，ggplot 是 R 語言的繪圖套件，教師從基礎的繪圖開始，逐步加深至個人化的圖表製作，利用顏色、形狀和大小等視覺元素來強化數據的傳達效果。學習如何修改圖表的主題和配色，以適應不同的需求，使圖表更加清晰易讀。課程最後介紹如何將圖表儲存為想要之解析度的圖片格式，以符合各種需求。



2024 年 6 月

參加 Return of Individual Research Results to Participants in Observational Cohort Studies 研討會

這場為期兩天的研討會由 NHGRI、NICHHD、NIDDK 以及 All of Us Research Program 聯合舉辦，主要討論如何向參與者回報其個人基因與基因組研究的結果。會議回顧了自 2006 年以來，從最初對告知受試者基因測序結果的反對和擔憂，逐步演變至如今將告知可操作的遺傳資訊視為一項倫理責任。基因報告回報的結果分為兩大類：主要發現與次要發現。主要發現是指基因檢測中針對特定疾病或研究問題所預期得到的結果，這些發現通常是檢測的主要目標，醫療或研究團隊對此有清晰的處理流程。次要發現則是在檢測過程中偶然發現的、與最初測試目的無關的基因變異。雖然次要發現並非研究或檢測的目標，但可能具有重要的臨床意義，對個人健康產生潛在影響。因此，如何處理與回報這些次要發現成為了研究中的一大挑戰。在回報結果的過程中，參與者有知情權，然而，在告知時須遵循「行善」與「不傷害」的原則，以避免參與者承受不必要的心理或經濟負擔。

Single Cell and Spatial Seminar: "Single Cell and Spatial Technologies: Applications to Disease"

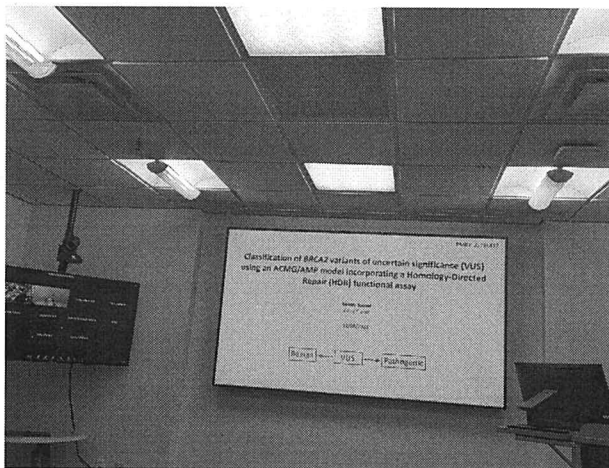
這會議介紹單細胞與空間技術主要是解析細胞層次的資訊和組織微環境的空間結構。單細胞技術專注於單一細胞的分析，非傳統方法分析大量細胞混合樣本，能夠提供細胞間的異質性及其在疾病進展或發育過程中的動態變化。空間技術則結合了基因組學與細胞的空間位置，提供基因表達的空間分布圖，從而提供細胞之間的相互作用及其對周圍微環境的影響。這兩項技術被用在癌症、免疫疾病、神經系統等，提供在疾病診斷、藥物開發等領域的精細資訊。

NHGRI Didactics genetics counseling

這個月，NIH 為完成培訓的臨床研究員(fellows)舉辦了畢業典禮，這也代表著這個定期會議告一段落。由於參與這會議的個案報告或期刊選讀分享大多來自該單位的 fellows，六月是他們完成培訓的時間點。八月時陸續會有新的研究員加入培訓。

在基因資料的判讀過程中，學習的重點之一是理解已知基因變異與臨床疾病之間的關聯，

並探討潛在變異對疾病的可能影響。ACMG（美國醫學遺傳學和基因組學學會）提出一套標準化指引對基因變異進行分類。根據指引，基因變異被分為五個主要類別：致病性、可能致病性、不確定性、可能良性及良性。這些分類依據變異對基因功能的影響和與疾病的相關性來進行判斷。ClinVar 和 dbSNP 等數據庫是查詢 variants 良惡性的常用工具，這些數據庫提供大量基因變異的已知信息，並依新出現的證據不斷更新資料庫的內容。參加這個定期的會議，使我對基因變異分類及其臨床應用有進一步的認識，對運用 ClinVar 和 dbSNP 等工具也較熟悉。



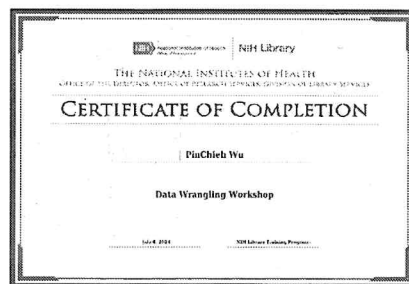
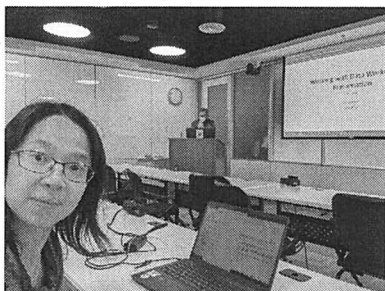
NHGRI Didactics genetics counseling 會議室

2024 年 7 月

進入最後一個月。這個月陽明交通大學生物醫學資訊研究所的研究生前來 NIH 學習。他們是碩一或碩二的學生，我陪同他們參觀定序實驗室、聽團隊內的專家向他們介紹生物醫學研究相關的實驗方法及技術。

參加 NIH 圖書館舉辦的課程: Data Wrangling Workshop

這門課程主要是介紹 R tidyverse 套件，並介紹如何使用它來操作、分析和匯出數據。課程中探討 tidyverse 函數的各種選項，如 select、filter、pipes、mutate、head、is.na 和 export。



All of Us 資料庫的基因資料分析

透過 Kristin Levine 的協助，學會抓取資料庫內的基因資料，接著是學習使用 PLINK 做資料分析。PLINK 是一款開源的全基因組關聯性分析工具集，主要功能是基因型與表型數據的分析，常用於 GWAS(全基因組關聯研究)。PLINK 也提供多種基因型數據的轉換工具，能夠將數據轉換為不同常見基因分析工具所需的格式。透過 PLINK 實際操作，我開始熟悉如何處理基因型和表型數據，並進行相關的分析。例如如何利用 PLINK 進行數據的品質控制，篩選出不符合標準的 SNP，去除基因型呼叫率過低的樣本，去除等位基因頻率過低 (MAF 太低) 的 SNP、檢查 Hardy-Weinberg 平衡 (HWE)、利用統計模型分析基因變異 (SNP) 與表型之間的關聯等。這些分析過程讓我能夠探討基因變異與特定疾病或性狀之間的潛在關聯。

三、心得及建議 (包括改進作法)

(一)心得

感謝院部與科內的支持，讓我可以前往美國國家衛生研究院進修一年。這一年我深深感受到這裡開放多元的研究環境。NIH 鼓勵來自不同背景的研究人員相互學習與交流，從不同角度探討問題，這種跨領域的合作豐富我的研究視角。此外，NIH 強調人才的多元發展，提供了豐富的培訓資源來提升研究者的能力。內部提供多樣的教學資源，定期舉辦各類科學研究活動，使研究者得到指導與資源。另外，NIH 先進的儀器設備也為研究提供了強大的技術支援。活絡的研究氛圍帶給我衝擊。在出發之前，對於新事物接受相對封閉的我，這次的經驗讓我感受到先進研究的發展和持續學習的重要性與動力。

(二)建議

A. 協助健康管理中心引進生物資訊相關之檢測，提供受檢者更多元個人化的健檢規劃

隨著基因定序技術成本的逐步降低，基因定序服務已成為健康檢查項目中的一項可選擇項目。國內外許多公司和醫院相繼推出了基因檢查的健檢方案。隨著多基因模型的導入，除了單基因疾病檢查外，多基因疾病風險評估也逐漸成為預防醫學檢查的選項之一。在陳以書部長與余憲忠主任的帶領下，健康管理中心不斷調整並新增更符

合民眾需求的健檢套組，逐步引進如 AI 影音報告和腸內菌檢測等生物資訊相關的項目。

基因檢查已逐漸成為個人化健康管理的選擇，隨著越來越多研究成果的發表，這項檢查的參考價值也逐漸提高。透過此次進修所學，希望能找到合適的合作廠商，共同討論出與健康管理中心目標一致的檢查項目，讓健檢的民眾能夠透過基因檢查獲得個人化的健康資訊，從而為他們制定更貼近需求的健康管理計劃。此外，依據 IRB 規範，將基因資料妥善儲存於院內，進一步豐富院內的基因資料庫，為醫療與研究提供更完整的數據支持。

B. 院內資料庫的整合與擴充

隨著近年來運算資源與效能的顯著提升，我們有能力處理越來越多的數據。然而，資料整合依然是一個重大的挑戰。尤其是來自多種不同來源的資料，無論是問卷、電子病歷系統、影像資料，還是生物資訊與穿戴式裝置數據，這些資料的統整不僅需要大量儲存空間，還面臨著跨類型數據統合的難題。這些問題成為 NIH 專家們持續討論和解決的重點之一。

范教授的團隊正積極致力於應對這些挑戰。他們針對資料的整合、管理及運用提出多項策略，特別是在應對巨量數據串聯方面不斷探索創新方法。台灣的醫療系統中基因資料的量也在迅速增加，特別是針對癌症病患的次世代基因定序數據。此外，健康管理中心逐步增添了生物資訊相關的檢查數據，這些數據與傳統問卷、生化及影像資料的整合，對於個案的後續健康照護規劃以及相關研究將帶來極大助益。未來，若能成功將電子病歷資料與這些生物資訊數據整合，並設計出易於搜尋和使用的介面，將能大幅提升院內資料的應用價值。資料整合可從健康管理中心開始擴展到全院，這不僅能使資料內容更加豐富，也有助於整體醫療系統的精準化與個別化發展。范教授於這方面有豐富經驗，持續向教授與其團隊學習，並請教授經驗分享，逐步向此方向前進。

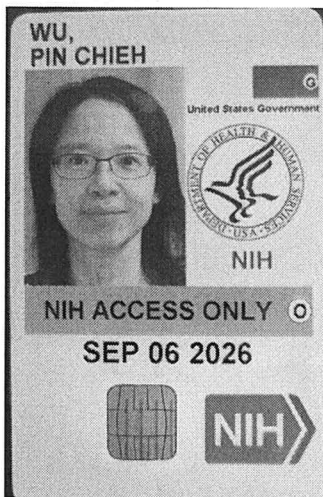
C. 跨界數據合作：提升院內使用外部資料庫的機會

院內資料庫是寶貴的資源，能提供個案完整且詳細的資訊，這對於臨床照護與研究而言具有極大的價值。而外部資料庫則擁有大量且多樣化族群的數據，可以為研究提供

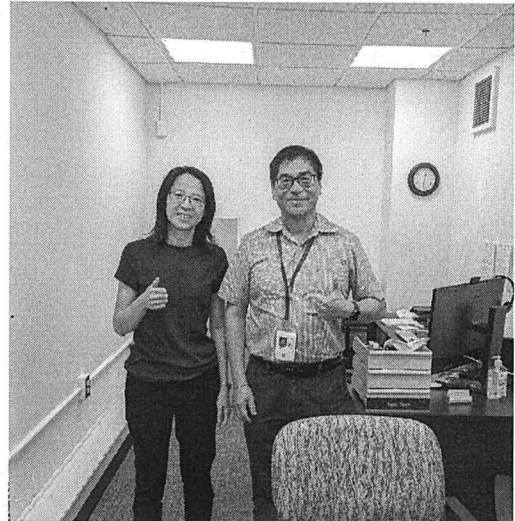
更加廣泛的視野與資料支持。兩者資料庫並行運作，能讓研究資源變得更加多元與豐富。NIH 支持的 All of Us 計劃建立了一個大型的資料庫，這個資料庫含基因組學、健康行為、環境因素等提供了豐富的數據，並可供全球研究機構進行合作使用。目前，台灣已有國立台灣大學、台大醫院和台北醫學大學與 All of Us 簽訂合約，成為合作夥伴。在完成合約簽訂後，這些機構的成員可以根據 All of Us 計劃和各自機構的規範，申請帳號並進行資料庫的使用。

透過與外部資料庫的合作，研究人員可以突破院內數據的限制，獲得更具多樣性的資料來源，這不僅能提升研究的廣度與深度，還能加速創新成果的誕生。因此建議若評估後可行，可以參考台大與北醫的模式，與 All of Us 簽約，獲取國際化的數據資源，進一步提升院內研究的廣度。

附錄



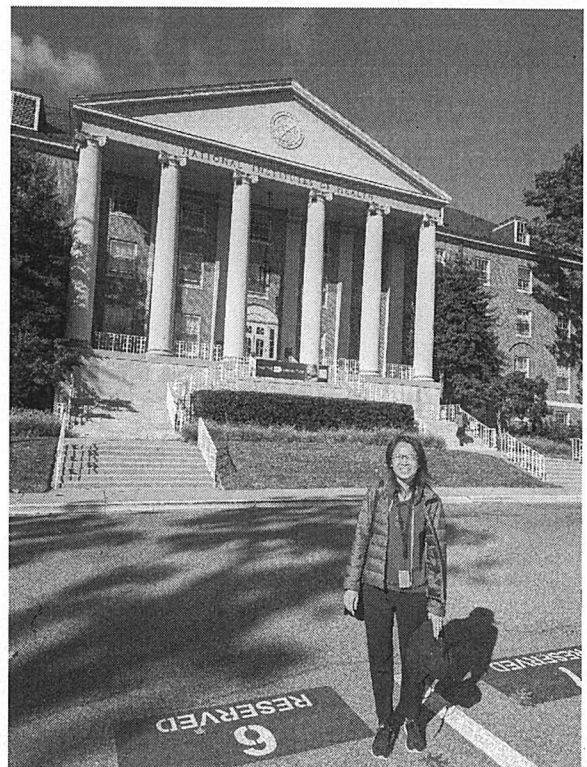
NIH 員工證



與范揚政教授合照



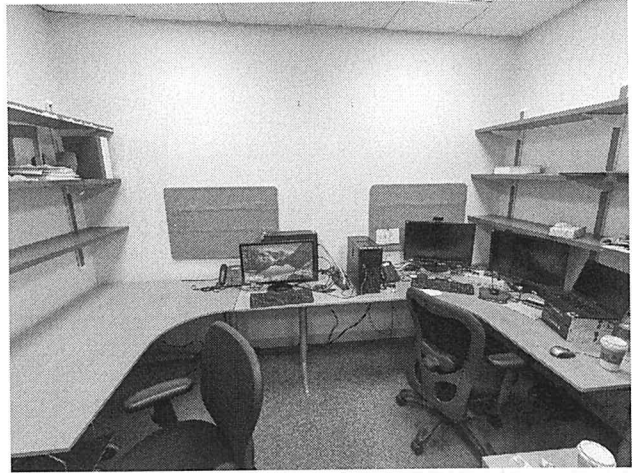
攝於 NIH Clinical Center
臨床研究醫學中心是以研究為導向的臨床醫院
也是辦公室的所在。



Building 1 是 NIH 的主要行政辦公樓，院
部所在。



研究辦公室外的走廊



研究辦公室



走出地鐵站，映入眼簾的是 All of Us 的大型宣傳布幔，格外醒目。該棟建築物為進入 NIH 的人員必須通過安檢之處。有員工證者，可從旁的閘門刷卡入內。



與 Kristin Levine 及中國醫藥大學附設醫院蔡昇達醫師合照



2023 年終歲末大華府地區研究學人交流座談餐會會後大合照

出國進修提報指標(人才培育申請時個人提報記錄檔)

單位	申請人	進修學校或機構	進修內容	指標1			指標2		
				名稱	計算方式	闕值	名稱	計算方式	闕值
健檢中心	吳品潔	美國麻州總醫院	骨質疏鬆疾病分子生物技術診斷與治療之研究與照護	經骨質密度檢測儀(DXA)診斷為骨質疏鬆但尚未發生骨折之患者接受持續照護之人數成長	(本季骨質疏鬆但尚未發生骨折之患者持續就診一年以上之人數-前一季患者持續就診一年以上之人數)/前一季患者持續就診一年以上之人數x100%	3%	骨質疏鬆疾病相關論文發表	每兩年骨質疏鬆疾病相關論文發表篇數	三篇/兩年