

# 九七的兒女

## Q 大數據 Big Data

大數據是指收集龐大而複雜的數據集，其中包括大量的數據、數據管理能力、社交媒體分析和實時數據。大數據的數據容量、組織方式、來源、管理架構都跟傳統數據不一 [1]。今期通訊我們將會探討大數據和它與「九七的兒女」出生隊列研究的關係。

顧名思義，大數據是指傳統軟件和互聯網平台無法管理的大量數據。大數據近年在不同行業廣泛應用，如醫療衛生、金融財務、企業營銷和資訊科技等。Gartner Inc.公司分析師Doug Laney先生在2001年定義大數據的三個特徵 - 3V。它們分別是容量 (Volume)，速度 (Velocity)和多樣性 (Variety)。容量是指數據資料量大，可達到TB, PB, 甚至EB級數；速度是指數據輸入輸出的速度；而多樣性是指數據的類型廣闊，可以用非結構化的方式呈現，如圖像，影片等 [2,3]。

1 Terabyte (TB) = 1024 GB (Gigabyte)  
1 Pebabyte (PB) = 1024 TB  
1 Exabyte (EB) = 1024 PB  
例子：以一隻8 GB Universal Serial Bus (USB) 計算，1 TB = 128隻USB



### 醫療衛生中的大數據

大數據在醫療衛生界別最常見的來源是來自公共電子健康記錄（醫健通是最明顯的例子）。透過電子健康記錄，可以有系統地記錄、分析或預測人口健康。

在疫情期間，政府和學術機構都利用大數據（如確診數字、本地個案即時有效繁值率等）製作實時儀錶版，從而監察和預測疫情發展。因此，大數據在醫療衛生界別有著重要的角色去分析和預測民眾健康。

### 大數據與「九七的兒女」出生隊列

在「九七的兒女」出生隊列中，大數據通常指組學數據 (Omics data)，即集合生物份子進行研究。基因組學和蛋白質組學是「九七的兒女」團隊正進行研究的大數據例子。在醫學研究界別，生物樣本（如血液、頭髮等）是收集大數據的主要來源。研究員利用「九七的兒女」參加者早前在「健康跟進計劃」中提供的生物樣本轉換為大數據繼而用作研究用途。所以，您的參與對於未來公共衛生研究和發現有著重大的貢獻。



參考資料：

1. Ishwarappa, Anuradha J. A brief introduction on Big Data 5vs characteristics and Hadoop Technology. Procedia Computer Science. 2015;48:319-24.
2. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big Data in Healthcare: Management, analysis and future prospects. Journal of Big Data. 2019Jun19;6(1).
3. Rivera E. The V's of big data [Internet]. Marbella International University Centre; 2020 [cited 2023Feb28]. Available from: <https://miuc.org/vs-big-data/>

## 如何利用「九七的兒女」生物樣本作大數據分析？ 〔以基因分析為例〕



### 大數據與 「九七的兒女」



「九七的兒女」參加者提供的大數據不但有助研究員了解疾病的成因，同時也幫助我們找出健康行為（如母乳餵哺、吸煙等）與疾病的關聯，並透過研究成果向有關持份者提出干預措施，建議或政策（如禁煙等），從而預防或緩和疾病發展，促進民眾健康。



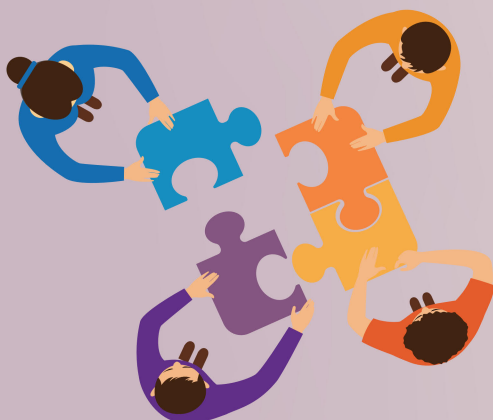
### 最終目標



#### 數據共享 / 加入相關基因數據研究聯盟

來自世界各地的生物樣本庫提供的大數據，並不包括個人資料。各地的研究員在授權下利用以不記名方式儲存的數據做相關人口健康的綜合分析、預測和研究，尋找疾病的成因和關聯。

全球生物樣本庫綜合分析計劃是一個有規模的研究聯盟，希望建立框架去推動各地生物樣本庫的協作。



大數據的出現是希望為人類帶來更準確和即時的資訊和分析。而大數據在醫療衛生界別更是在推動人口健康及提供更精準的醫療服務中扮演重要角色。

在醫學研究中的大數據只是作為研究工具或證據去證明或預測疾病。大數據分析才是研究的重心。在「九七的兒女」出生隊列中，團隊研究員只會利用當中的醫學數據，而不會知悉和利用當中的個人資料。「九七的兒女」研究也有相關的守則和指引供授權研究人員遵守。我們亦根據香港特別行政區法例第 486 章《個人資料（私隱）條例》的規定，「九七的兒女」參加者所提供的個人資料（如：姓名、電話、電郵及郵寄地址等），將授權我們作為聯絡通訊和研究用途，並絕對保密。



## 今年「九七的兒女」有甚麼工作可以分享？

我們已為「九七的兒女」的網站完成優化，並加入更多資訊。如果您未曾瀏覽，歡迎大家進入以下網站去發掘更多資訊：[《https://www.aprmay97.sph.hku.hk/》](https://www.aprmay97.sph.hku.hk/)！「九七的兒女」的 Instagram 亦會不定期更新，希望藉此與您們保持聯絡，將一些關於「九七的兒女」研究的歷史及與公共衛生相關的資訊帶給大家。歡迎大家私訊小編，告訴我們有關您的近況或想了解更多不同的健康資訊喔！

在研究方面，「九七的兒女」一直透過不同的研究方法或技術為生物樣本進行化驗，上年我們完成了代謝組學測定 (NMR Metabolomics assays)，而今年我們將繼續利用不同的組學數據進行實驗及分析，包括蛋白質組學、微生物組學、基因組學、代謝組學等。



## 甚麼是蛋白質組學？利用它進行了甚麼的測量？

蛋白質組學 (proteomics) 是其中一種組學數據 (omics data)，可找出蛋白質的結構和功能，以及其在細胞內的相互作用 [2]；它亦應用於醫學領域上，可透過分析血液中的蛋白質去識別生物標記 (biomarker) [3]。我們利用「九七的兒女」參加者於10歲時所量度的高度和體重去計算其體重指標 (BMI)，之後透過計算BMI的基因風險評分 (genetic risk score)，確認當中250個BMI遺傳預測因子中最高和最低的參加者。及後將這250個血漿樣本於2022年9月送到美國進行蛋白質組學的化驗。研究數據能夠提供大約7,000個蛋白質組學分析。團隊已於2022年年底收到相關的研究數據，我們正在進行處理及分析，並將會進行更多後續研究探討。

鳴謝：研究獲得大學教育資助委員會撥款基金（撥款編號17112622）的支持。



## 你們正在進行甚麼的研究工作？

研究團隊正與香港基因組中心 (Hong Kong Genome Institute) 合作，為「九七的兒女」的生物樣本進行全基因組測序 (Whole Genome Sequencing (WGS)) [4]。每個生物都有一個指定的遺傳密碼 (genetic code)，通常是由adenine (A), thymine (T), cytosine (C) 和guanine (G)組合而成。WGS 是透過測序 (Sequencing)、數據處理 (Data Processing) 及數據分析 (Data Analysis) 去找出不同生物中遺傳密碼的次序，從而透過這些已排列的次序去尋找可能與疾病相關的基因或致病基因的變異 [5]。



## 何博士，你可否解釋更多有關「微生物組學與肥胖的關聯」的研究嗎？



何寶婷博士  
博士後研究員

最近，我們在香港大學泛組學科研中心 (CPOS) 的協助下，從「九七的兒女」16至18歲健康跟進計劃收集的部分糞便樣本中提取出所有微生物基因，並採用先進的霰彈槍測序法 (shotgun sequencing) 將糞便中存在的微生物基因進行全面檢測。這個方法可以幫助研究團隊了解糞便樣本中微生物的多樣性，檢測微生物的相對數量，甚至可以用來推測與健康相關的微生物代謝信息。目前我們已經收到糞便微生物的基因序列，並取得初步分析結果。初步結果顯示，與體重正常人群的糞便相比，肥胖人群的糞便樣本具有較低的微生物物種多樣性，微生物構成並不一致（例如含有較多的氣味桿菌 *Odoribacter splanchnicus*），而在肥胖男性糞便中微生物差異更加明顯。團隊需要進行更多後續研究探討到底這種微生物差異會對疾病的發展產生甚麼作用。

## 駱博士，你最近研究「人類基因代謝組學與藥物的關聯」，能否告訴我們相關的研究結果？



駱珊博士  
助理教授（研究）

我在過去兩年主要負責「九七的兒女」基因組學數據的處理與分析。我們運用人類基因組學，以藥物靶點孟德爾隨機化設計 (Drug-target Mendelian randomization study)，模擬临床上常用的降血脂藥物「他汀類 (HMGCR抑制劑)」和「PCSK9」抑制劑，並推導其對250項代謝產物，人體測量學、血液指標、心腎肝的生物標記。我們初步的研究發現遺傳抑制的HMGCR和PCSK9可顯著降低膽固醇相關代謝產物的水平，例如載脂蛋白B (Apolipoprotein B)，並且降低冠狀動脈心臟病的風險。另外，遺傳抑制的HMGCR可降低亞油酸的水平，研究並未發現遺傳抑制的HMGCR和PCSK9對青少年有明顯的不良作用。這次研究有望為青少年的臨床隨機對照試驗提供助力。在無需將青少年暴露在潛在不良影響的藥物的情況下，帶來人類基因組學的科學證據。

鳴謝：研究獲得香港特別行政區前食物及衛生局（現為醫務衛生局）（撥款編號CFS-HKU1及17181271）和香港大學啟動資金的支持。

## 歐陽博士，「九七的兒女」來年會有甚麼計劃？



歐陽兆倫博士  
助理教授

我們希望將來可以繼續利用於「『九七的兒女』健康跟進計劃」中所收集到不同的生物樣本（如尿液和指甲）去進行不同類型的測量。多元化的數據不但可以豐富「九七的兒女」的數據庫，而且我們亦能透過這些數據進行不同的研究工作，為將來的公共衛生研究帶來新發現和知識。

1. Graves PR, Haystead TA. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev.* 2002 Mar;66(1):39-63; table of contents. doi: 10.1128/MMBR.66.1.39-63.2002. PMID: 11875137; PMCID: PMC120780.

2. National Cancer Institute. NCI Dictionary of Cancer Terms [Internet]. National Cancer Institute; [cited 2023Feb28]. Available from: <https://www.nccih.gov/publications/dictionaryofcancer/terms/dof/proteomics>

3. Al-Amrani S, Al-Jabri Z, Al-Ziaheri A, Alshkaili J, Al-Khabori M. Proteomics: Concepts and applications in human medicine. *World J Biol Chem.* 2021 Sep 27;12(5):57-69. doi: 10.4331/wjbc.v12.i5.57. PMID: 34630910; PMCID: PMC8473418.

4. Centers for Disease Control and Prevention. Whole genome sequencing [Internet]. PulseNet Methods & Protocols, The United States Government; 2022 [cited 2023Feb28]. Available from: <https://www.cdc.gov/pulsenet/pathogens/wgs.html>

5. Hong Kong Genome Institute. Whole Genome Sequencing [Internet]. Genomics - Hong Kong Genome Institute; 2022 [cited 2023Feb28]. Available from: <https://hkgp.org/en/discover-genomics/genomics/?tab=1>

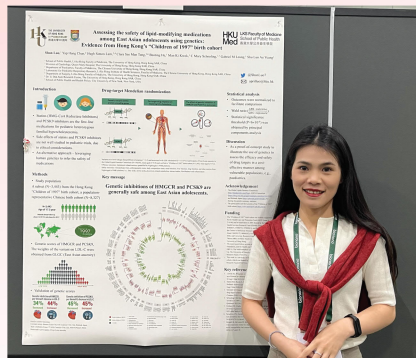




# 帶著「九七的兒女」的研究發現參加國際學術會議



在2019冠狀病毒病疫情逐漸平穩的情況下，世界各國的聯繫逐漸恢復。2022年下旬，兩位「九七的兒女」團隊成員前往美國洛杉磯參加了人類基因相關的國際學術會議，即由美國人類遺傳學學會 (American Society of Human Genetics)主辦的年度會議。以下是她們在學術會議後的分享見聞：



駱珊博士  
助理教授 (研究)

我很榮幸代表研究團隊參加學術會議，並向來自世界各地的研究者展示了香港「九七的兒女」出生隊列的研究發現。藉此機會，我亦榮幸代表「九七的兒女」出生隊列參與全球脂質遺傳學聯合會 (Global Lipids Genetics Consortium) 的討論和發言。現時的人類遺傳學和基因組學研究以西方人口為主導。美國人類遺傳學學會明確指出未來的研究發展方向將注重種族、性別、年齡等的多樣性以及基因與環境的相互作用。

感謝您們於16至18歲期間參與了「九七的兒女」健康跟進計劃，並提供了生物樣本和寶貴的數據。未來，我們將充分利用生物樣本，並匯聚多組學數據，與全球的研究團隊攜手，持續致力於人類基因組學的多樣性及全球公共衛生研究。



ASHG ANNUAL MEETING 2022  
ASHG.ORG

在會議中，我分享的主題是「糖尿病的遺傳傾向與代謝組學關聯」。我利用「九七的兒女」的數據，根據每位參加者的基因變體，計算了他們的基因風險評分。由於分數的高低代表糖尿病的風險，透過分析這個分數與血液中的代謝物的關聯，我們能夠辨別與之相關的生物標記，從而及早發現及預防糖尿病。目前這篇研究正準備投稿到學術期刊。

我在海報上利用可愛生動的圖像展示我的研究發現，例如用不同小動物的體型來類比美國和中國肥胖和糖尿病的發病率，以此突出「亞洲人群體重指標BMI較低，但糖尿病發病率仍然很高」的特點。結果，海報吸引了不少對流行病學視覺化感興趣的學者前來，讓我有機會進一步向他們介紹有關的研究內容，同時亦可了解對方的研究興趣，因此結識了很多新朋友。

作為一個研究者，評價一個海報最基本的標準應是研究假設是否新穎和有意義，方法是否嚴謹；但作為一個學術成果的推廣者，我覺得如何呈現自己的工作也很重要。例如在展示海報時，給予藝術和圖像化的發揮，讓表達更加實在和直接。我認為作為研究者，我們的使命是將研究成果普及化。希望透過「貼地」的語言，讓參與者和大眾了解和應用研究結論，從而促進學術界、參加者和大眾三方交流。



陳思宇小姐  
碩士研究生

## 團隊出版

何寶婷博士  
博士後研究員

### 成長發育模式與肺部功能及哮喘的關聯

一直以來，很多香港人都習慣祝福小朋友快高長大，但是身體快速成長是否必然對健康有利呢？為了解答這個問題，我們的研究小組採用一個最新的成長軌跡模型 (SITAR 模型，即通過平移和旋轉疊加建立生長曲綫的模型)，利用以前收集的「九七的兒女」出生隊列中參加者的身高體重及體重指標 (BMI) 的記錄，計算出參加者於嬰兒期及青春期的成長量 (較多或較少)，節奏 (較遲或較早) 以及速度 (較快或較慢)，並觀察這些成長數據與我們在16至18歲收集的肺部功能和哮喘病史的關聯。在刪除身高、經濟水平和其他因子的影響後，我們從研究結果中發現，如果有更早和更快的嬰兒期和青春期成長發育模式，青少年的肺活量水平亦比較高，但同時氣道功能反而較差，並且有更高的喘息及哮喘風險，這個現象在男性中更為顯著。此項研究結果與之前採取同樣生長軌跡模型的英國隊列研究結果一致，可能由於在身體成長速度較快的人群中，雖然肺容量亦快速成長，但氣道成長速度卻並不與肺容量成長速度匹配，導致不成比例的肺容量與氣道功能，因而相對更容易發生喘息甚至哮喘。此項研究刊登於國際流行病學學術期刊

《International Journal of Epidemiology》。我們正在探究這種較快和較早的成長模式是否與其他方面的健康與疾病有關，並有興趣知道你們對這項研究的看法。

He B, Li AM, Kwok MK, Au Yeung SL, Leung GM, Schooling CM. Association of growth patterns during infancy and puberty with lung function, wheezing and asthma in adolescents aged 17.5 years: evidence from "Children of 1997" Hong Kong Chinese Birth Cohort. International Journal of Epidemiology. 2022 Sep 24;dyac182. doi: 10.1093/ije/dyac182. (https://academic.oup.com/ije/advance-article-abstract/doi/10.1093/ije/dyac182/6713580/redirectedFrom=fulltext)

### 「九七的兒女」出生隊列研究小組

香港大學李嘉誠醫學院 公共衛生學院 香港薄扶林沙宣道7號白文信樓1樓

電話：(852) 3917 6824 傳真：(852) 3520 1945  
電郵：aprmay97@hku.hk 網頁：https://www.aprmay97.sph.hku.hk/

主辦機構 香港大學公共衛生學院 香港特別行政區政府衛生署

通訊編輯 簡朗軒先生 李浩宜小姐 駱珊博士 何寶婷博士 陳思宇小姐

私隱條例 根據相關特別行政區法例第486章《個人資料(隱私)條例》的規定，閣下所提供的個人資料(如：姓名、電話、電郵及郵寄地址等)，將授權我們只作聯絡通訊和研究用途，不會作商業用品，絕對保密。

如您希望不再收取本計劃的資訊，請以書面形式郵寄或傳真讓我們知悉。

如果您已更改地址、聯絡電話或電郵，請盡快通知我們，以便日後保持聯繫。



「九七的兒女」網頁



「九七的兒女」Facebook群組



「九七的兒女」研究小組多年來從不同的組織獲得研究經費。以下是歷年來獲得的撥款項目：

香港特別行政區醫務衛生局健康護理及促進基金  
香港特別行政區醫務衛生局控制傳染病研究基金  
香港特別行政區醫務衛生局醫療衛生研究基金  
大學教育資助委員會研究資助局  
WYNG基金會

圖片來源：本設計使用的圖片 (部分細節經過處理) 由 vecteezy.com 和 irasutoya.com 提供。本研究小組對以上設計者表示感謝。