

2025年臺灣國際科學展覽會

優勝作品專輯

作品編號	090001
參展科別	醫學與健康科學
作品名稱	建構CpG 甲基化老化預測模型並評估表觀遺傳 年齡與血液相關生化指數和潛在疾病的關聯性 Construction of an Aging Prediction Model Based on CpG Methylation and Evaluation of the Association Between Epigenetic Age, Blood-Related Biochemical Indices, and Potential Diseases
就讀學校	臺北市立第一女子高級中學
指導教師	陳倩瑜 陳怡芬
作者姓名	葉宸羽 邱品潔
關鍵詞	<u>CpG 甲基化、表觀遺傳年齡、老化</u>

作者簡介



我們是來自北一女中數理資優班的葉宸羽（左）與邱品潔（右）。非常幸運能加入陳倩瑜教授的實驗室，參與DNA甲基化、人體健康與老化相關的研究，為科技醫療貢獻微薄之力。感謝一路以來大家的支持與鼓勵，成就了今天的我們，共創珍貴回憶與未來無限可能。

2025 年臺灣國際科學展覽會

研究報告

區別：(編號由國立臺灣科學教育館統一填列)

科別：醫學與健康科

作品名稱：

建構 CpG 甲基化老化預測模型並評估表觀遺傳年齡與血液
相關生化指數和潛在疾病的關聯性

Construction of an Aging Prediction Model Based on CpG
Methylation and Evaluation of the Association Between
Epigenetic Age, Blood-Related Biochemical Indices, and
Potential Diseases

關鍵詞：CpG 甲基化、表觀遺傳年齡、老化

編號：

(編號由國立臺灣科學教育館統一填列)

摘要

老化是生命週期中逐漸喪失功能的過程，本研究以 86 萬個 CpG 位點的甲基化數據及血液生化指數和相關疾病的問卷資料為數據集，分析與老化顯著相關的甲基化位點，其中與年齡高度相關的位點符合已知的染色體和基因分布。基於此，本研究建構一個年齡預測模型 Model_Age (Random Forest, estimators=200)，並比較實際年齡與模型預測的甲基化年齡，透過回歸線將數據分為「表觀遺傳年輕」與「表觀遺傳年老」兩類。驗證結果顯示，這兩類在特定血液生化指數上存在顯著差異，接下來將會探討與疾病的關聯性，並檢驗台灣受試者與國際結果是否具有一致性。本研究透過表觀遺傳甲基化資訊，揭示老化過程中的生理變化，並預測可能的疾病風險，為老化相關血液疾病提供個人化診斷與治療策略，推動精準醫療的發展。

Abstract

Aging is a process of gradual functional decline throughout the lifecycle. This study utilizes a dataset of methylation data from 860,000 CpG sites, alongside biochemical blood indices and questionnaire data related to associated diseases, to analyze methylation sites significantly correlated with aging. The sites that exhibit a strong correlation with age align with known chromosomal and gene distributions. Based on this, we constructed an age prediction model, Model_Age (Random Forest, estimators=200), and compared the actual age with the methylation age predicted by the model. Using regression lines, the data was categorized into "epigenetically young" and "epigenetically old" groups. Validation results indicated significant differences in specific biochemical blood indices between these two categories. We will further explore the relationship with diseases and examine whether the results from Taiwanese subjects are consistent with international findings. This research utilizes epigenetic methylation information to reveal physiological changes during the aging process and predict potential disease risks, providing personalized diagnostic and treatment strategies for aging-related blood diseases and promoting the development of precision medicine.

壹、前言

一、研究動機

人類自古以來擔憂衰老和死亡，而想了解自己的壽命，科學家們也不斷探討其生物機制，如達海弗利克極限前無盡的分裂和端粒縮短使染色體衰老等。老化是無可避免且不可逆的漸進過程，會導致人體功能衰退，減弱適應性和抵抗力，更是多種病症（如癌症、糖尿病、心血管疾病、認知功能下降等）的誘發因子，這些都是我們希望能改善的。若能預測並追蹤個人老化程度，將能更有效控制疾病，實現精準化醫療。

近年來研究顯示，老化與表觀遺傳甲基化之間存在關聯。透過檢測 CpG 位點的甲基化程度，可以推算出表觀遺傳年齡。本研究旨在進一步驗證台灣資料中表觀遺傳時鐘模型的有效性，並透過血檢結果中的各類生化指數及疾病問卷調查資料，評估此模型在實際健康狀況上的合理性，為精準醫療提供參考，從而改善人們的生活品質和健康狀況。

二、研究目的

- (一) 找出與年齡相關性高的甲基化位點並推測表觀遺傳相關基因
- (二) 基於 CpG 位點甲基化程度建構年齡預測模型 Model_Age
- (三) 以實際年齡和預測年齡之差距篩選出表觀遺傳年輕與年老者
- (四) 評估表觀遺傳年輕與年老者之間的生化指數有否顯著性差異
- (五) 評估表觀遺傳年輕與年老者之間的問卷調查疾病結果有否顯著性差異
- (六) 驗證台灣資料可否得到國際發表之表觀遺傳相關結果

三、文獻探討

(一) 表觀遺傳學與甲基化

表觀遺傳是指在不改變 DNA 序列的條件下調控基因表現，在細胞分化和多種疾病中皆扮演關鍵角色，如癌症和免疫失調等[1, 2]。常見的表觀遺傳研究如影響分化的非編碼 RNA (non-coding RNA, ncRNA)、對 DNA 修復至關重要的組蛋白修飾以及表觀遺傳中研究最多的 DNA 甲基化等[3]。

其中，甲基化是指透過 DNA 甲基轉移酶的催化，將甲基轉移到 DNA 上，通常會抑制基因表現[4]。近年來研究指出，甲基化程度可做為老化的生物指標，用來建構表觀遺傳老化時鐘，其模型預測年齡稱為甲基化年齡 (DNA methylation age, DNAm) [5-7]。甲基化年齡可反映出受試者的潛在健康狀況，如研究顯示癌症患者的甲基化年齡較未患病衰老 36 歲之多[8]。

雖然已有許多研究證實此表觀遺傳時鐘，目前仍在探討其背後的生物學意義[9]，並持續尋找較佳演算法[10]，已有研究提及和年齡具有顯著相關性的甲基化位點[11-15]以及相關基因區域[15-18]記於附錄 B。

(二) 機器學習模型

1. 線性回歸 (Linear Regression)

常見的簡單模型，透過找到數據之間的最佳回歸直線，最小化預測值與實際值之間的誤差，簡單線性回歸為 $y = mx + b$ ，多元線性回歸為 $y = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n + b$ 。

2. 支持向量回歸 (Support Vector Machine, SVM)

基於支持向量機原理的回歸模型，讓多數資料位於一定誤差範圍內，具有容錯率，常見於房價預測、序列分析等[19]。

- (1) 線性核 (Linear Kernel)：公式為 $K(x, x') = x^T x'$ ，表示 x 和 x' 的內積，其中 T (transpose) 表示將向量的行和列互換。
- (2) 多項式核 (Polynomial Kernel)：公式為 $K(x, x') = (\gamma x^T x' + r)^d$ ，其中 d 是多項式的最高次數， γ 為影響模型對樣本之間距離敏感程度的縮放因子， r 是常數項。
- (3) 徑向基函數核 (Radial Basis Function Kernel, RBF)：公式為 $K(x, x') = \exp(-\gamma \|x - x'\|^2)$ ， $\exp$ 指自然指數函數 (exponential function)，使核函數的值固定於 0 到 1 之間， $\|x - x'\|$ 則為兩點之間的歐幾里得距離。

3. 隨機森林 (Random Forest)

透過多個決策樹進行回歸，在隨機樣本上訓練每棵樹，以所有樹的預測值取平均做為結果，常用於處理高維數據，因此適用於本研究之多特徵甲基化。

貳、 研究方法

一、研究設備與器材

由於本研究所用資料所需記憶體龐大，以下分述硬體規格、所需軟體以及數據相關資訊。

(一) 硬體

1. 電腦處理器 12th Gen Intel (R) Core (TM) i7-1260P 2.10 GHz
2. 電腦記憶體 16.0 GB
3. 外接硬碟 2TB

(二) 軟體

1. Visual Studio Code (version 1.94)
2. Anaconda (version)
3. Python 套件 (包含 Pandas、numpy、sklearn、datetime、Os、tqdm 和 matplotlib、selenium 等)
4. Excel

(三) 數據資料

取自於台灣人體生物資料庫的甲基化資料(申請案標號:TWBR10811-02),每人有約八十六萬位點之資料(最大位點數為 866897),另有疾病問卷調查資料,涵蓋個人基本資料、個人健康行為、生活環境、飲食狀況、女性相關問題和經濟狀況等,可供後續甲基化與健康狀態關聯性之探討,受試者共計 2088 人。

二、研究流程與方法

本研究流程如圖 1 所示,首先整理原始資料並製作資料集,接著利用皮爾森相關係數選出與年齡相關性高的甲基化位點,分析其基因分布位置,以找出支持此資料具有普遍性的生物學證據。然後構建年齡預測模型 Model_Age,並比較不同模型和參數的運作成本,找出在有限時間和空間下準確且適用於甲基化資料的模型。再來比較實際年齡與模型預測之甲基化年齡,以回歸線將資料分為表觀遺傳年輕與表觀遺傳年老兩類,透過分析受試者的血液檢查生化指數和問卷調查之疾病,評估指數和疾病等生化指標在兩資料集之間有否顯著性差異,驗證此表觀遺傳年輕與表觀遺傳年老的分類,以檢驗台灣受試者資料與國際發表結果是否具有一致性。

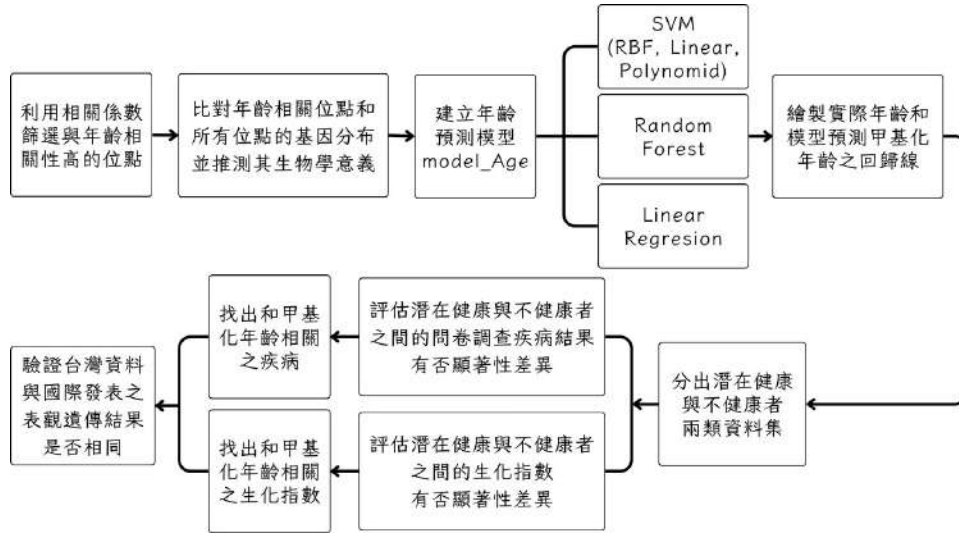


圖 1、研究流程圖（本研究作者自行繪製）

(一) 製作各位點甲基化程度與病患資料的資料集

甲基化資料集如圖 2 所示，儲存受試者 ID、年齡以及各位點的甲基化程度，其中甲基化程度以 β 值表示，介於 0 到 1 之間，0 代表此位點完全未甲基化，1 則代表完全甲基化。

	A	B	C	D	E	F	G	H	I	J	K	L
1	Methyl_ID	AGE	1	2	3	4	5	6	7	8	9	10
2	MET3_20160202A	57	0.574881	0.914957	0.878034	0.915587	0.93331	0.190466	0.805724	0.728545	0.812016	0.785027
3	MET3_20160202B	34	0.543203	0.875229	0.874281	0.90908	0.929698	0.20419	0.839746	0.737069	0.852474	0.825707
4	MET3_20160202C	57	0.536595	0.878625	0.873959	0.924217	0.91635	0.223354	0.806855	0.708881	0.807736	0.808185
5	MET3_20160202D	40	0.514183	0.88468	0.902714	0.904135	0.940286	0.197557	0.81473	0.720826	0.862127	0.799262
6	MET3_20160202E	44	0.523125	0.914009	0.856559	0.912781	0.921536	0.20605	0.809443	0.785398	0.838613	0.803416
7	MET3_20160202F	61	0.491495	0.88586	0.877662	0.908494	0.933431	0.181264	0.799794	0.735143	0.838175	0.795538
8	MET3_20160202G	62	0.561529	0.922558	0.878873	0.929349	0.920263	0.197093	0.812159	0.749063	0.732795	0.83654
9	MET3_20160202H	70	0.483483	0.85724	0.868213	0.938547	0.948834	0.176743	0.811512	0.769534	0.779383	0.799033
10	MET3_20160203A	61	0.496078	0.901143	0.885183	0.896328	0.925227	0.176169	0.783024	0.736605	0.853711	0.813007
11	MET3_20160203B	70	0.491949	0.889367	0.897515	0.91808	0.938095	0.298206	0.80401	0.760224	0.878479	0.81137
12	MET3_20160203C	58	0.502575	0.884156	0.86293	0.915875	0.931611	0.159398	0.767877	0.711206	0.840914	0.757511
13	MET3_20160203D	48	0.552256	0.875723	0.858676	0.90703	0.917638	0.215821	0.828004	0.72184	0.855988	0.790999
14	MET3_20160203E	66	0.48927	0.934864	0.893695	0.898949	0.933867	0.231307	0.799898	0.751155	0.828388	0.808392
15	MET3_20160203F	63	0.493298	0.898068	0.881524	0.908813	0.927398	0.154359	0.819567	0.726238	0.836199	0.795383
16	MET3_20160203G	58	0.471968	0.916593	0.850278	0.920302	0.930103	0.171499	0.815012	0.733273	0.833751	0.815502
17	MET3_20160203H	37	0.586373	0.925849	0.878834	0.900211	0.952905	0.120325	0.8377	0.737041	0.804988	0.816008
18	MET3_20160204A	62	0.579918	0.900685	0.871597	0.891245	0.942837	0.165264	0.799467	0.751636	0.865736	0.770271
19	MET3_20160204B	68	0.508681	0.904147	0.926515	0.913161	0.920432	0.180516	0.799205	0.740564	0.832936	0.780405
20	MET3_20160204C	66	0.484454	0.865144	0.866288	0.90938	0.936239	0.161196	0.810513	0.764367	0.862749	0.826079

圖 2、甲基化資料集之部分節錄（本研究作者自行製作）

(二) 利用皮爾森相關係數篩選較與年齡相關的前一千個甲基化位點

對於高達八十六萬多的位點資料，計算甲基化資料和年齡的皮爾森相關係數，篩選出受年齡影響較顯著的前一千個甲基化位點，取得其索引值。本研究用來評估相關性的皮爾森相關係數（Pearson correlation）公式為：

$$r(x, y) = \frac{COV(x, y)}{SxSy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

其中 r 為相關係數， COV 為共變數， $\bar{x}$ 為 x 的平均數， $\bar{y}$ 為 y 的平均數。皮爾森相關用於探討兩連續變數 (X, Y) 之間的線性相關，若兩變數之間相關係數絕對值較大，則表示 x 與 y 之間共變程度較大。

(三) 分析年齡相關性高之甲基化位點的基因位置分布

透過 Python 爬蟲擷取台灣人體生物資料庫的公開線上甲基化位點資料，並分析其基因分布，確認此資料中的位點符合已證實之特點，或是具有潛在生物學意義，以支持此筆資料能反映普遍性結果。

(四) 以甲基化程度建構年齡預測模型 Model_Age

取前一千個甲基化位點的資料 80% 作為訓練集資料，建構甲基化程度與年齡預測模型，再以其餘 20% 作為測試集評估模型的準確性。

(五) 比較不同模型的準確性

依資料特性選擇了常見的三種支持向量機 (RBF、linear 和 Polynomid SVM)、隨機森林 (Random Forest) 和線性回歸 (Linear Regression)，各運作十次取其誤差值與平均，選出表現最佳之模型。

(六) 比較實際年齡和預測年齡分出表觀遺傳年輕與年老者

透過模型預測結果，得知受試者甲基化程度反映的表觀遺傳年齡，實際年齡較年輕者列為「表觀遺傳年輕」，實際年齡較年長者列為「表觀遺傳年老」，分出表觀遺傳年輕與表觀遺傳年老的兩類數據集。

(七) 評估表觀遺傳年輕與年老者之間的生化指數有否顯著性差異

對於常見血液生化指數，確認表觀遺傳年輕與表觀遺傳年老兩資料集間有否顯著性差異，驗證以甲基化年齡和實際年齡分類的數據集。並且比較「和年齡相關性高的生化指數」以及「在兩資料集之間具顯著性差異（即和甲基化年齡相關性高）的生化指數」有何差異。

(八) 評估表觀遺傳年輕與年老者之間問卷調查疾病結果有否顯著性差異

對於常見血液相關疾病，確認表觀遺傳年輕與表觀遺傳年老兩資料集間有否顯著性差異，驗證以甲基化年齡和實際年齡分類的數據集。並且比較「和年齡相關性高的疾病」以及「在兩資料集之間具顯著性差異（即和甲基化年齡相關性高）的疾病」有何差異。

(九) 檢驗台灣受試者的資料是否與國際發表結果一致

以台灣人體生物資料庫的甲基化數據與國際期刊與論文發表的結果比對，評估兩者的一致性，並找出相關性高的生化指數和疾病與表觀遺傳的關聯性。

參、 研究結果與討論

一、 年齡相關性高之甲基化位點的基因位置分布

透過分析與年齡相關性高的前一千個甲基化位點在台灣人體生物資料庫中的相關資訊，可以發現這些位點在染色體、基因區域和 CpG 區域上的分布，並推測其生物學意義。

(一) 染色體

圖 3 呈現年齡相關性高的甲基化位點分布所占比率大於 4.5% 的染色體，所有位點中染色體占比由高至低分別是染色體 1、2、6、11、3、7、5、12、17、10，而年齡相關性高的一千個位點中，染色體分布比例由高至低則是染色體 1、6、2、3、17、7、5、11、19、8、12。

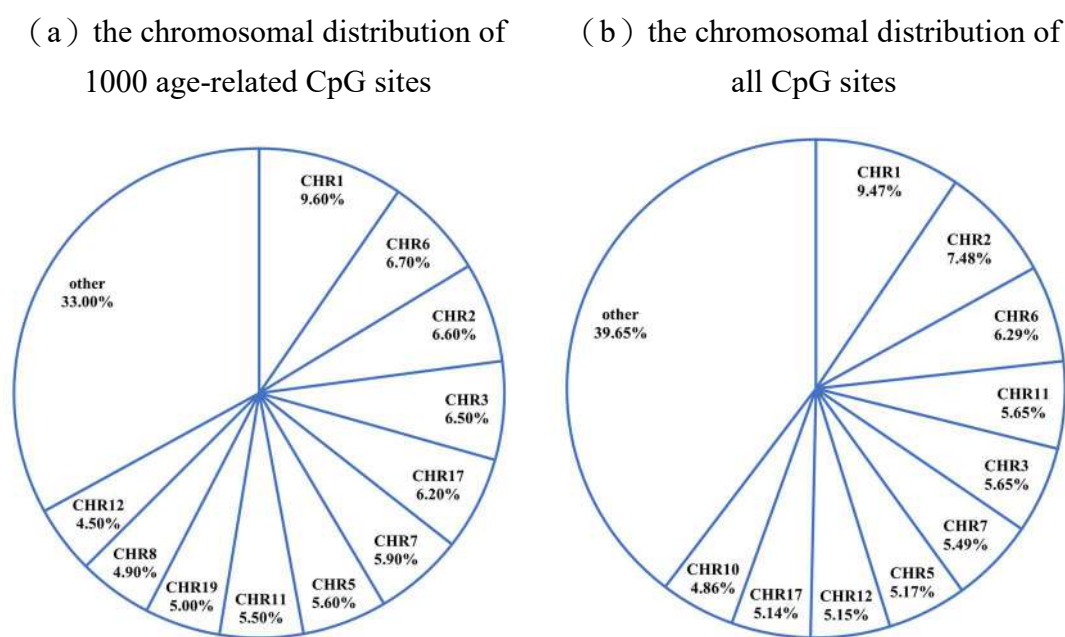


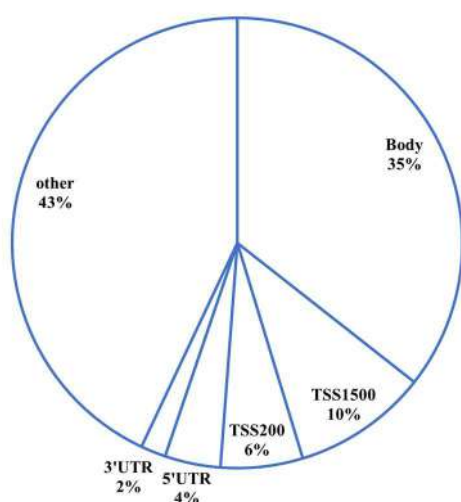
圖 3、年齡高相關性位點與所有位點的染色體分布比較（本研究作者自行繪製）

可發現在所有位點中，染色體 8 和 19 因占比未達 4.5% 沒有出現於圖表上，但年齡相關性高的甲基化位點分布中卻分別占了 4.90% 和 5.00%，因此推測相較於其他染色體，年齡相關性高的甲基化位點可能較多的呈現於染色體 8 和 19，而此表現可能源於染色體 8 和 19 上的老化相關基因，分別是染色體 8 上的成人型早衰症基因（WRN），以及染色體 19 上的阿茲海默基因（APOE）。相反的，染色體 10 雖然在所有位點中的占比達到 4.5%，卻沒有出現在年齡相關性高的甲基化位點圖表上，說明年齡相關性高的甲基化位點可能較少分布於染色體 10。

(二) 基因區域

圖 4 呈現年齡相關性高的甲基化位點分布所佔比例大於 2% 的基因區域，可發現染色體上主要基因區域占比皆為基因本體 (Body)、轉錄起始區 1500 (TSS1500)、轉錄起始區 200 (TSS200)、五端非轉譯區 (5'UTR) 和三端非轉譯區 (3'UTR)。

(a) the gene region distribution of 1000 age-related CpG sites



(b) the gene region distribution of all CpG sites

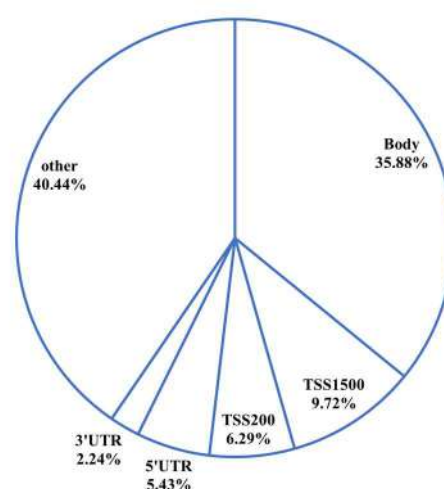


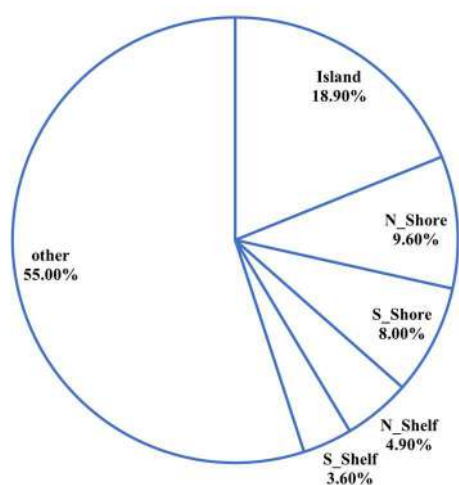
圖 4、年齡高相關性位點與所有位點的基因區域分布比較 (本研究作者自行繪製)

兩者中占比能大於 2% 的基因區域占比高低之順序排列相同，故比較不同區域在年齡相關性高的甲基化位點分布中是否占比較高，發現僅有轉錄起始區 1500 符合，縱使兩者差異甚小，相較於其他基因區域，年齡相關性高的甲基化位點在轉錄起始區 1500 分布較密集。

(三) CpG 區域

圖 5 呈現年齡相關性高的甲基化位點分布所佔比例大於 2% 的基因區域，將無資料者記為其他，依其餘資料可知 CpG 區域占比由高到低為 CpG 島 (Island，甲基化高密度區)、N 岸 (N_shore，靠近 CpG 島的區域)、S 岸 (S_shore，靠近 S 區架的區域)、N 區架 (N_shelf，位於 CpG 島周圍的低 CpG 區域)、S 區架 (S_shelf，位於 CpG 島另一側的低 CpG 區域)。

(a) the CpG region distribution of 1000 age-related CpG sites



(b) the CpG region distribution of all CpG sites

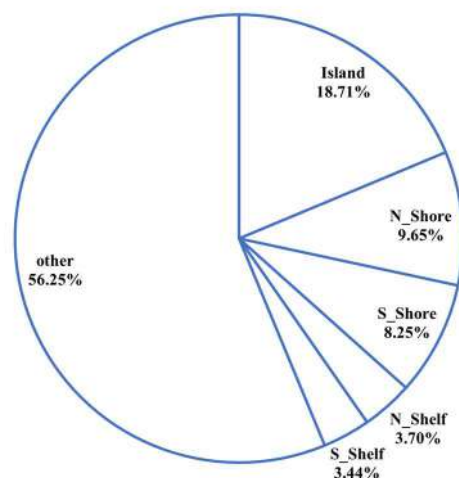


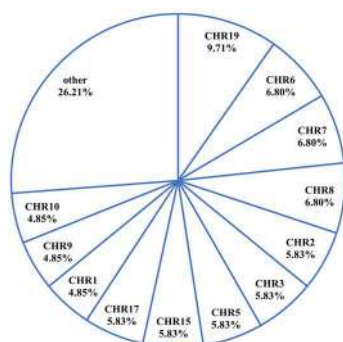
圖 5、年齡高相關性位點與所有位點的 CpG 區域分布比較 (本研究作者自行繪製)

兩者中 CpG 區域的占比高低之順序排列相同，故比較不同區域在年齡相關性高的甲基化位點分布中是否占較高，發現 CpG 島和兩個區架皆符合，推測年齡相關性高的甲基化位點在 CpG 島及區架上分布較密集。

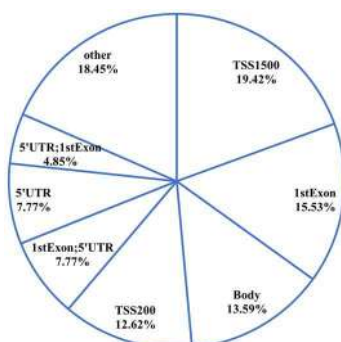
(四) 已證實年齡相關性高的甲基化位點之資料分析

接著本研究取用曾被文獻發表為「和年齡具有顯著相關性」的 103 個甲基化位點 (附錄 B) 和台灣人體生物資料庫的數據，分析此 103 個已被證實與年齡相關的位點，關於染色體、基因區域、CpG 區域的分布，如圖 6 所示。

(a) the chromosomal distribution of CpG sites



(b) the gene region distribution of CpG sites



(c) the CpG region distribution of CpG sites

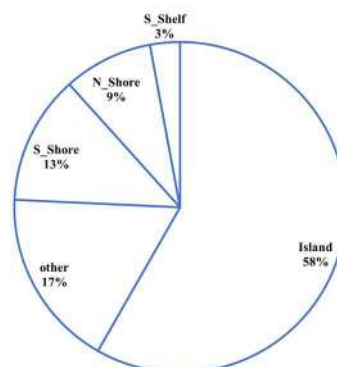


圖 6、文獻探討中年齡高相關性甲基化位點的染色體、基因區域和 CpG 區域分布比較 (本研究作者自行繪製)

染色體部分可證實在染色體 19 具有 23 個染色體之中最多年齡相關性高甲基化位點的分布，染色體 8 的占比也高達 6.80%，此外染色體 15 在本研究的資料中占比未達 4.5%而沒有反映出來。

基因區域部分可證實高年齡相關性甲基化位點在轉錄起始區 1500 上的占比遠高於所有甲基化位點的分布占比，另外此處的第一外顯子、轉錄起始區 200 和五端非轉譯區占比較所有甲基化位點高，為本研究沒有呈現的。

CpG 區域部分顯示此處的 CpG 島占比顯著高於在所有位點上所反映的，相較之下 S 岸僅稍微高於在所有位點上的占比。CpG 島本是甲基化位點最密集聚集之處，但此特性沒有表現在所有甲基化位點的 CpG 區域分布上，推測可能資料取樣的不同，本研究所用來自周邊血檢體，上述所提及研究多來自其他細胞組織。

透過比較前 1000 個年齡相關性高的甲基化位點、103 個文獻探討和年齡具有顯著性相關的位點以及本研究中所有位點，得出染色體 8 和 19 以及轉錄起始區 1500 中呈現較多年齡相關性高的甲基化位點，支持前 1000 個高年齡相關性的甲基化位點整體上符合已知特徵，或具有潛在生物學意義解釋。

二、生化指數與年齡相關性

生化指數來自周邊血檢驗結果，其中 Ref_1 表示最小標準值，Ref_2 為最大標準值，詳細資料如附錄 B 所示。生化指數與年齡的相關係數大於 0.01 者如圖 7 所示，相關係數取絕對值後大於 0.1 者視為與年齡呈現較高相關的生化指數，其中大於 0.1 者歸類為總體性正相關，小於 -0.01 者則歸類與總體性負相關，正負相關分別以藍色和橘色表示。

與年齡呈現總體性正相關的生化指數中，相關性由大到小依序為血中尿素氮、抗 B 型肝炎核心抗原抗體、糖化血色素、空腹血糖、總膽固醇和天門冬胺轉胺酶。與年齡呈現總體性負相關的生化指數中，相關性由大到小依序為血小板、白蛋白、抗 B 型肝炎核心抗原抗體、紅血球和白血球。

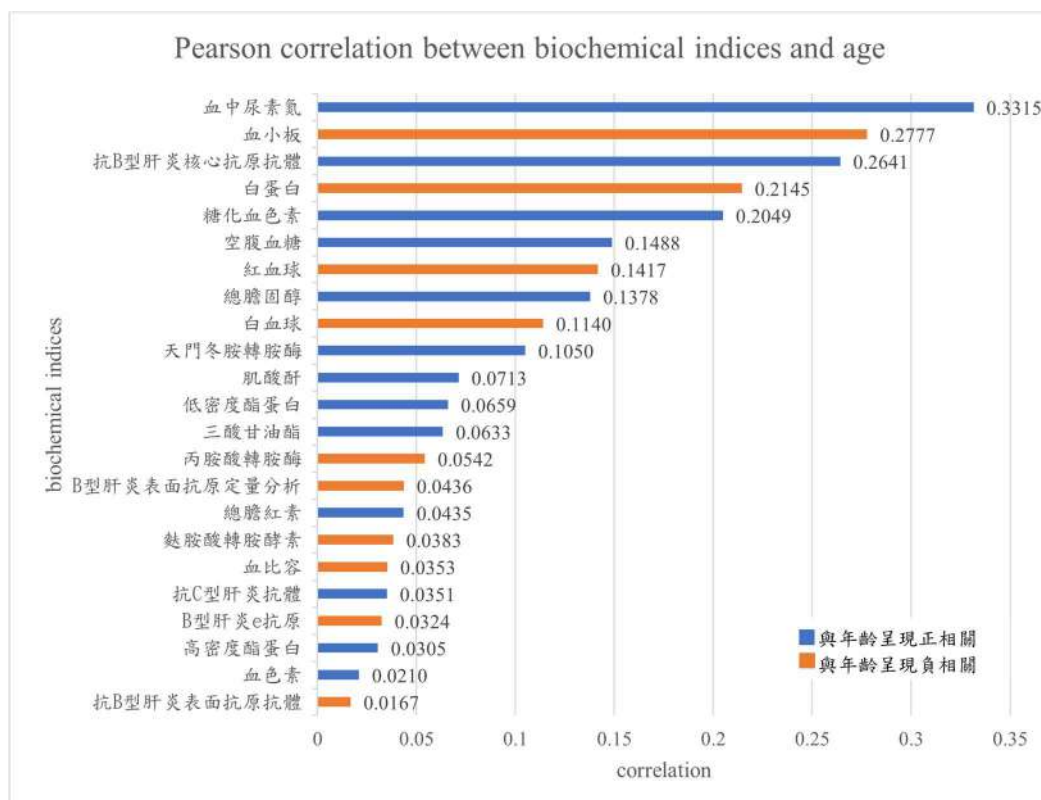


圖 7、生化指數和年齡相關係數（本研究作者自行繪製）

（一）與年齡呈現總體性正相關

圖 8 至圖 13 為與年齡呈現總體性正相關之圖表，正相關性由大到小排列分別為血中尿素氮、抗 B 型肝炎核心抗原抗體、糖化血色素、空腹血糖、總膽固醇和天門冬胺轉胺酶。

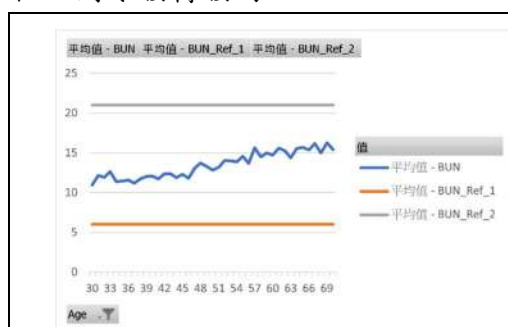


圖 8、血中尿素氮與年齡的關係
（本研究作者自行繪製）

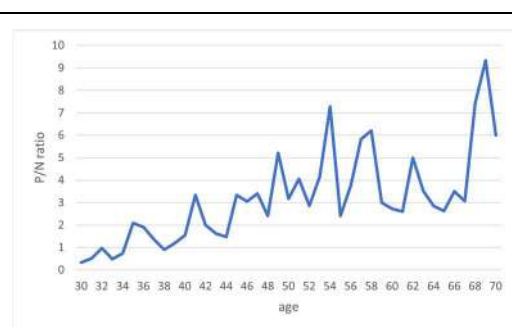


圖 9、抗 B 型肝炎核心抗原抗體
正負比 (positive/negative) 與年齡的關係
（本研究作者自行繪製）

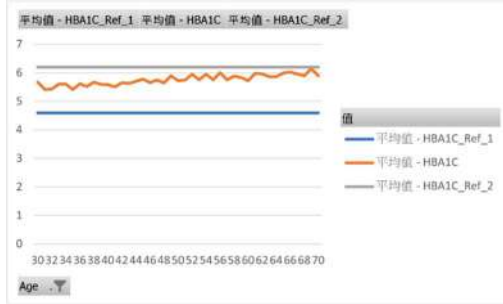


圖 10、糖化血色素與年齡的關係
(本研究作者自行繪製)

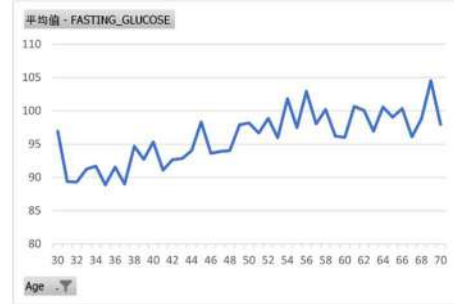


圖 11、空腹血糖與年齡的關係
(本研究作者自行繪製)

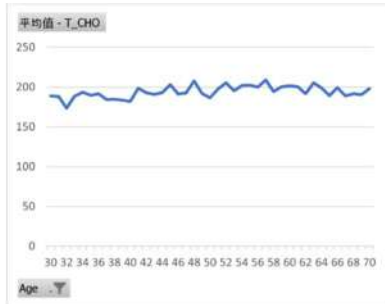


圖 12、總膽固醇與年齡的關係
(本研究作者自行繪製)

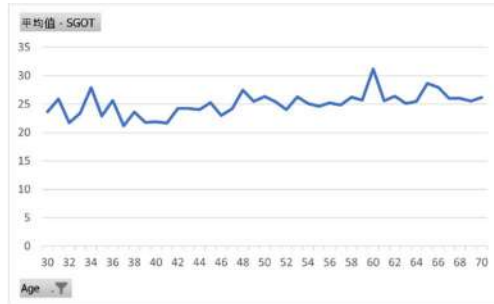


圖 13、天門冬胺轉胺酶與年齡的關係
(本研究作者自行繪製)

(二)與年齡呈現總體性負相關

圖 14 至圖 18 為與年齡呈現總體性負相關之圖表，負相關性由大到小排列分別為血小板、白蛋白、紅血球和白血球。

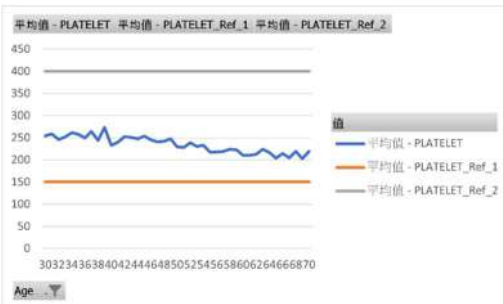


圖 14、血小板與年齡的關係
(本研究作者自行繪製)

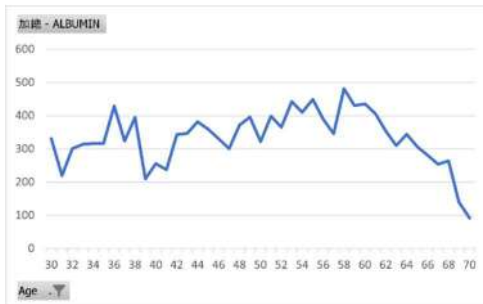


圖 15、白蛋白與年齡的關係
(本研究作者自行繪製)

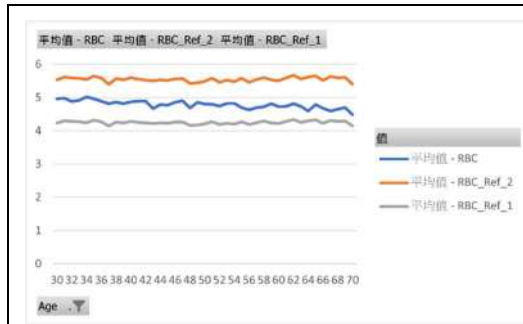


圖 16、紅血球與年齡的關係
(本研究作者自行繪製)

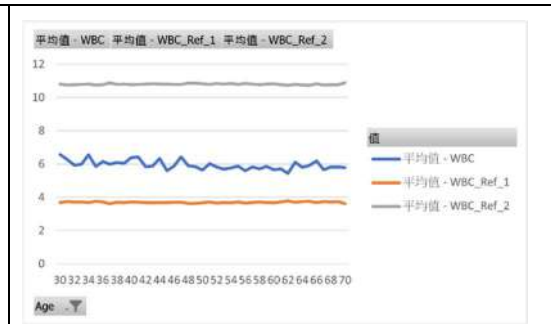


圖 17、白血球與年齡的關係
(本研究作者自行繪製)

三、Model_Age 甲基化年齡預測模型

本研究共訓練三類甲基化年齡預測模型，分別是支持向量機 (Support Vector Machine, SVM)、隨機森林 (Random Forest) 以及線性模型 (Linear)，以下為此三類模型的測試成本分析，詳細測試成本數據表記於附錄 A。

針對每種模型分別調整參數，如支持向量機的不同核函數 (kernel)、隨機森林中的樹數 (number of estimators) 和線性模型的學習率 (learning rate)，以尋找最佳組合。

在本研究的測試成本評估中，所呈現的平均測試成本結果為模型在相同硬體配置及測試環境中，重新切割數據並訓練模型後的測試成本平均值。為了確保結果的穩定性和可靠性，本研究針對該模型進行了 10 次獨立測試，並將所得成本進行平均處理，以減少隨機誤差對結果的影響。

(一) 支持向量機模型

如圖 18 所示，線性函數 (Linear)、多項式函數 (Polynomial) 和徑向基函數 (RBF) 三種支持向量機中，依平均測試成本結果而言，徑向基函數的表現最好，其次是線性函數和多項式函數。

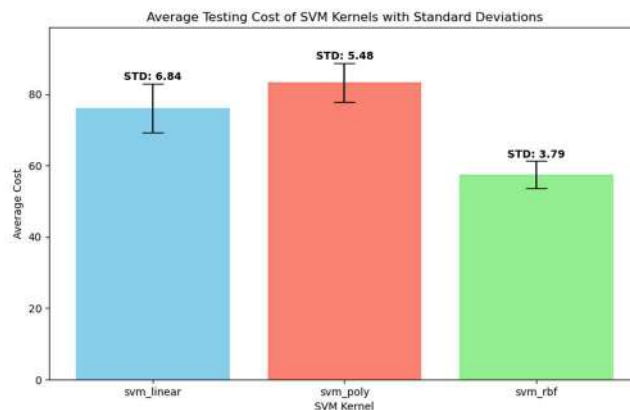


圖 18、使用不同核函數 (線性、多項式和徑向基) 建構支持向量機模型之測試成本比較
(本研究作者自行繪製)

(二) 隨機森林模型

如圖 19 所示，測試使用不同樹數（number of estimators）的隨機森林何者表現較好，依平均測試成本結果而言，樹數為 200 時表現最好，樹數降低或增加都使測試成本增高。

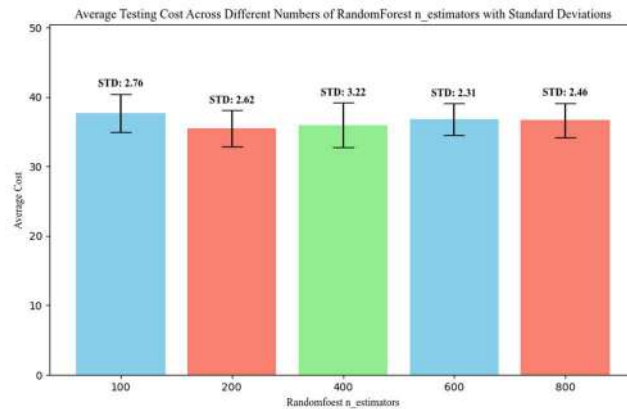


圖 19、使用不同樹數（number of estimators）建構隨機森林模型之測試成本比較
（本研究作者自行繪製）

(三) 線性模型

將迭代數（RunTimes）固定於 10000 次，並檢測以不同學習率訓練模型可得的測試成本，發現其數值與使用其他模型的測試成本相比顯著較大，考量到運行時間與記憶體的使用效率，此研究選擇不採用線性模型。

結果顯示比對三種核函數的支持向量機、不同樹數的隨機森林和線性模型之後，樹數（number of estimators）為 200 的隨機森林模型表現較佳，其平均測試成本為 35.6，平均訓練成本為 5.1，基於該結果本研究選擇此模型建構甲基化年齡預測模型。

四、實際年齡與模型預測年齡之關係

圖 20 呈現實際年齡與模型預測年齡之間的分布情形，並以 $y=x$ 之對角線切割，可以發現若以對角線切割，會使年輕者多被歸類於表觀遺傳年老，年老者多被歸類於表觀遺傳年輕，為避免實際年齡本身對生化指數和疾病的關聯對判別造成影響，因此不選用對角線作為兩資料集的分隔。

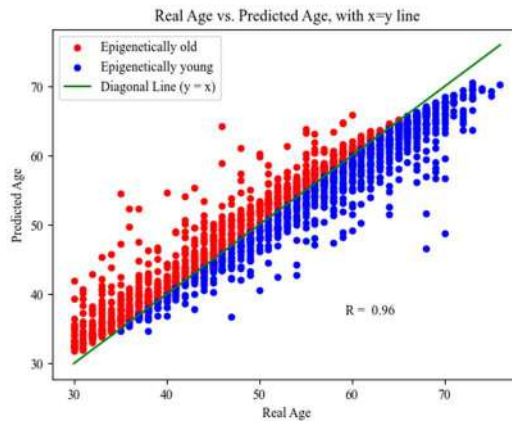


圖 20、實際年齡與 Model_Age 模型預測年齡之分布圖（使用對角線切割）
（本研究作者自行繪製）

圖 21 同樣呈現實際年齡與模型預測年齡之間的分布情形，並以回歸線將受試者數據切割為兩個資料集，相較於對角線切割的方式更不受實際年齡影響。

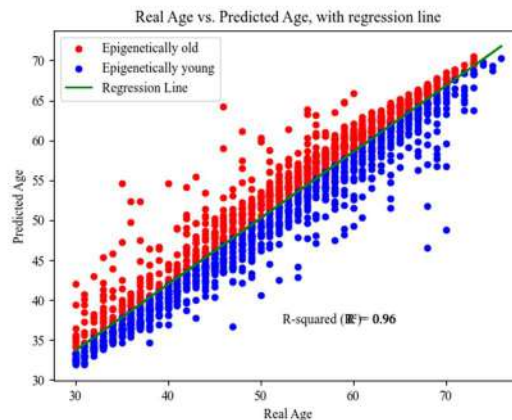


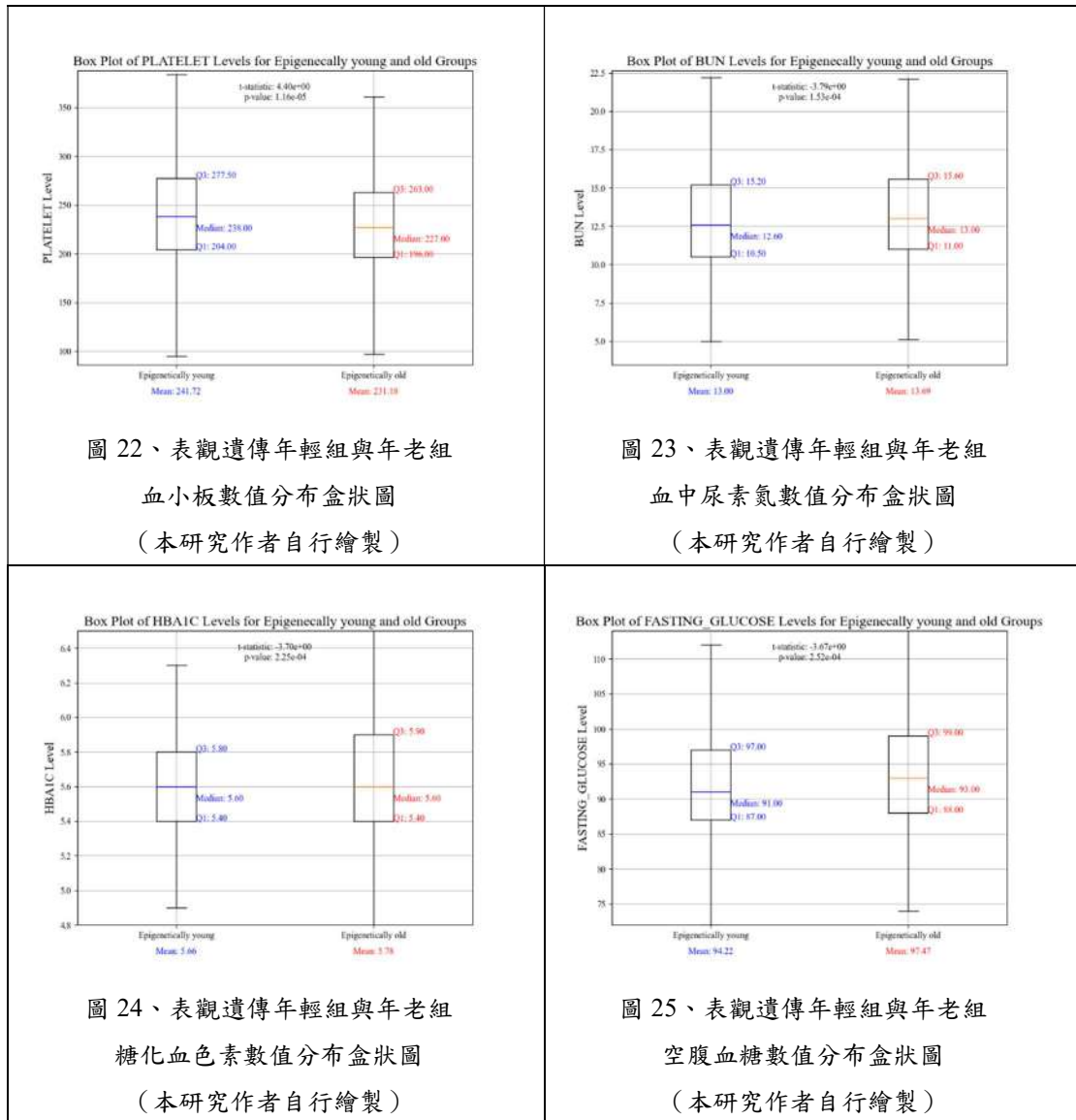
圖 21、實際年齡與 Model_Age 模型預測年齡之分布圖（使用回歸線切割）
（本研究作者自行繪製）

為了減少模型本身系統性的偏差，此研究以回歸線為分界線，將數據點分為兩個區域，從而對比表觀遺傳年輕與年老的個體特徵。

1. 回歸線上及右下方區域：該區域中的數據點主要分布在回歸線附近或其右下方，顯示這類個體的預測年齡與其實際年齡相當或稍低。這表明這些個體的老化速度較為緩慢或與實際年齡一致，可能具有較佳的健康狀態。因此，此區域的個體被歸類為**表觀遺傳年輕者 (Epigenetically young)**。
2. 回歸線左上方區域：分布在回歸線左上方的數據點表示模型預測年齡明顯高於實際年齡，反映出這類個體的老化速度相對較快，與實際年齡不相符。這暗示這些個體的身體狀況可能較差，老化進程較快，因而歸類為**表觀遺傳年老者 (Epigenetically old)**。

五、生化指數對於表觀遺傳年輕與表觀遺傳年老兩數據集之差異

使用 T 檢定和卡方檢定分析，圖 22 到圖 31 呈現表觀遺傳年輕與年老兩數據集之間呈現顯著性差異之數值分布盒狀圖。與甲基化年齡呈現總體性正相關的生化指數依相關性由大到小為抗 B 型肝炎核心抗原抗體、血中尿素氮、糖化血色素、空腹血糖、肌酸酐、血色素和總膽固醇。與甲基化年齡呈現總體性負相關的生化指數依相關性由大到小為血小板、白血球和白蛋白。



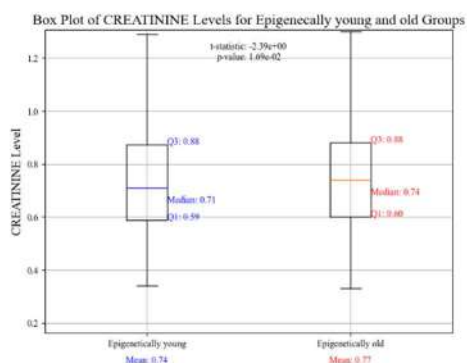


圖 26、表觀遺傳年輕組與年老組
肌酸酐數值分布盒狀圖
(本研究作者自行繪製)

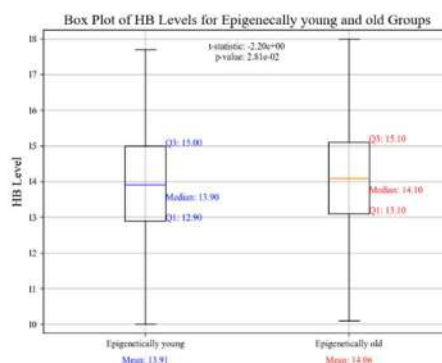


圖 27、表觀遺傳年輕組與年老組
血色素數值分布盒狀圖
(本研究作者自行繪製)

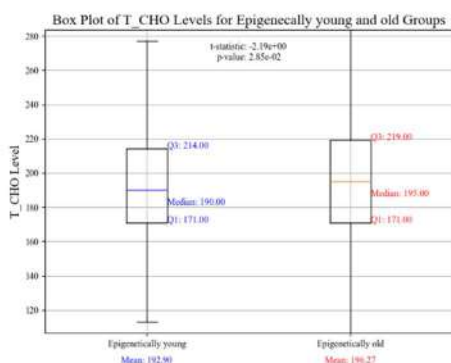


圖 28、表觀遺傳年輕組與年老組
總膽固醇數值分布盒狀圖
(本研究作者自行繪製)

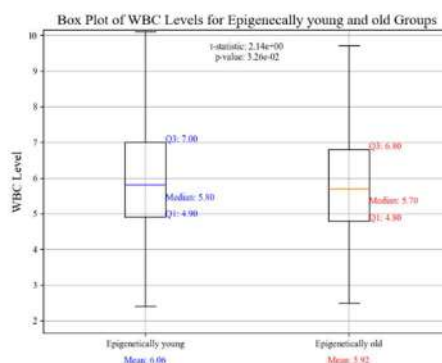


圖 29、表觀遺傳年輕組與年老組
白血球數值分布盒狀圖
(本研究作者自行繪製)

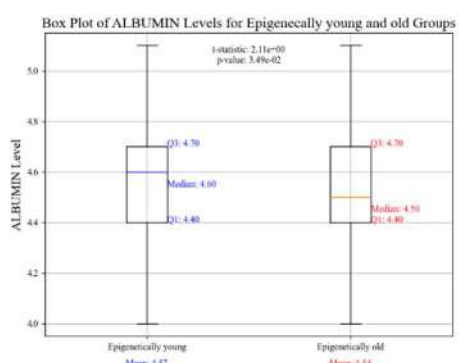


圖 30、表觀遺傳年輕組與年老組
白蛋白數值分布盒狀圖
(本研究作者自行繪製)

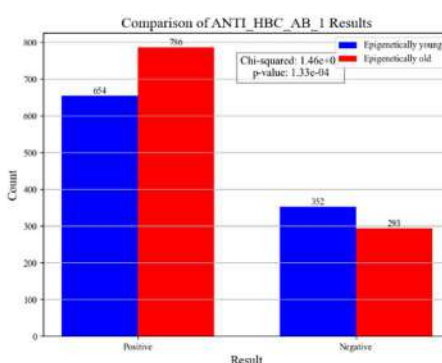


圖 31、表觀遺傳年輕組與年老組
抗 B 型肝炎核心抗原抗體結果之人數比較
(本研究作者自行繪製)

由 Model_Age 的結果可知表觀遺傳年輕與表觀遺傳年老兩資料集之生化指數能反映出明顯差異，可知老化和受環境、生活方式與遺傳等多種條件影響的表觀遺傳年齡具有關聯性，支持此表觀遺傳年輕與表觀遺傳年老的分類。

此外，本研究發現天門冬胺轉胺酶、抗 B 型肝炎核心抗原抗體和紅血球雖然並非高甲基化年齡相關，卻是高年齡相關，顯示這些生化指數較不受表觀遺傳中的甲基化影響。

而肌酸酐和血色素雖然並非高年齡正相關，卻是高甲基化年齡正相關，因此可用甲基化年齡推測肌酸酐和血色素的生化指數數值，進而評估其潛在相關疾病風險。

肆、 結論

- 一、染色體 8 和 19 以及轉錄起始區 1500 中呈現較多年齡相關性高的甲基化位點，支持前 1000 個年齡相關性高的甲基化位點與已知特徵一致或具有潛在生物學意義。
- 二、不同參數的支持向量機、隨機森林和線性模型中，樹數為 200 的隨機森林模型表現較佳，因此選擇作為甲基化年齡預測模型。
- 三、本研究之甲基化資料可反映與年齡的相關性，符合文獻探討所敘甲基化資料可用於建構表觀遺傳時鐘。
- 四、表觀遺傳年輕與年老兩組間的生化指數會反映出顯著性差異，支持甲基化年齡與潛在健康指標具有關聯性，並且驗證此分類。
- 五、本研究建構的甲基化年齡模型可應用於早期預測心血管、糖尿病及癌症風險，同時推動精準醫療及個人化健康管理，有助於老化相關疾病的早期識別。
- 六、目前尚待解決的問題還有評估表觀遺傳年輕與年老者之間的問卷調查疾病結果有否顯著性差異，以驗證表觀遺傳時鐘和此甲基化模型的在地適用性，預計能在後續研究階段完成。

伍、 未來展望

希望能擴大樣本量，並結合更多表觀遺傳標記的應用，以提升多樣性和精確性。除了驗證總體一致性外，希望可以找出台灣數據和國際資料之間有所差別之處，並由生化指數和疾病等相關資訊，推測地區間差異所代表的地理人文或生物學意涵，深入探討不同地域影響衰老機制的不同因子，助力健康預測模型精進，推動科學進展。

陸、 參考文獻

1. Moosavi, A., & Motevalizadeh Ardekani, A. (2016) . Role of epigenetics in biology and human diseases. *Iranian Biomedical Journal*, 20 (5) , 246–258.
2. Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T.-P., Beyan, H., Whittaker, P., McCann, O. T., Finer, S., Valdes, A. M., Leslie, R. D., Deloukas, P., & Spector, T. D. (2010) . Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Journal Name*, Volume (Issue) , page range.
3. Mazzone, R., Zwergel, C., Artico, M., Taurone, S., Ralli, M., Greco, A., & Mai, A. (2019) . The emerging role of epigenetics in human autoimmune disorders. *Clinical Epigenetics*, 11, Article 34.
4. Moore, L. D., Le, T., & Fan, G. (2012) . DNA methylation and its basic function. *Neuropsychopharmacology*, 38 (1) , 23–38.
5. Jung, M., & Pfeifer, G. P. (2015) . Aging and DNA methylation. *BMC Biology*, 13, Article 7.
6. Jones, M. J., Goodman, S. J., & Kobor, M. S. (2015) . DNA methylation and healthy human aging. *Aging Cell*, 14 (6) , 924–932.
7. Higham, J., Kerr, L., Zhang, Q., Walker, R. M., Harris, S. E., Howard, D. M., Hawkins, E. L., Sandu, A.-L., Steele, J. D., Waiter, G. D., Murray, A. D., Evans, K. L., McIntosh, A. M., Visscher, P. M., Deary, I. J., Cox, S. R., & Sproul, D. (2022) . Local CpG density affects the trajectory and variance of age-associated DNA methylation changes. *Genome Biology*, 23 (1) , 216.
8. Horvath, S. (2013) . DNA methylation age of human tissues and cell types. *Genome Biology*, 14 (10) , R115.
9. Field, A. E., Robertson, N. A., Wang, T., Havas, A., Ideker, T., & Adams, P. D. (2018) . DNA methylation clocks in aging: Categories, causes, and consequences. *Molecular Cell*, 71 (6) , 882–895.
10. Momeni, Z., & Saniee Abadeh, M. (2019) . MapReduce-based parallel genetic algorithm for CpG-site selection in age prediction. *Genes*, 10 (12) , 969.
11. Ying, K., Liu, H., Tarkhov, A. E., Sadler, M. C., Lu, A. T., Moqri, M., Horvath, S., Kutalik, Z., Shen, X., & Gladyshev, V. N. (2024) . Causality-enriched epigenetic age uncouples damage and adaptation. *Nature Aging*, 4 (2) , 231–246.
12. Bocklandt, S., Lin, W., Sehl, M. E., Sánchez, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011) . Epigenetic predictor of age. *PLOS ONE*, 6 (6) , Article e14821.
13. Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag,

- D. O., Jöckel, K.-H., Erbel, R., Mühleisen, T. W., Zenke, M., Brümmendorf, T. H., & Wagner, W. (2014). The aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, 15, Article R24.
14. Vetter, V. M., Kalies, C. H., Sommerer, Y., Bertram, L., & Demuth, I. (2022). Seven-CpG DNA methylation age determined by single nucleotide primer extension and Illumina's Infinium MethylationEPIC array provide highly comparable results. *Frontiers in Genetics*, 12, 759357.
 15. Jain, N., Li, J. L., Tong, L., Jasmine, F., Kibriya, M. G., Demanelis, K., Oliva, M., Chen, L. S., & Pierce, B. L. (2024). DNA methylation correlates of chronological age in diverse human tissue types. *Epigenetics & Chromatin*, 17 (1), 25.
 16. Garagnani, P., Bacalini, M. G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A. M., Gentilini, D., Vitale, G., Collino, S., & Rezzi, S. (2012). Methylation of the ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*, 11 (6), 1132–1134.
 17. Kondo, M., Aboshi, H., Yoshikawa, M., Ogata, A., Murayama, R., Takei, M., & Aizawa, S. (2020). A newly developed age estimation method based on CpG methylation of teeth-derived DNA using real-time methylation-specific PCR. *Journal of Oral Science*, 63 (1), 54–58.
 18. Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., Izeta, A., de Haan, G., Geiger, H., & Wagner, W. (2018). Epigenetic age-predictor for mice based on three CpG sites. *eLife*, 7, e37462.
 19. Iyer, R. (2024, February 18). Demystifying support vector machines: Kernel machines.
 20. 檢驗醫學部 (2013) 檢驗參考值|台大醫院|健康教育資訊網

柒、 附錄

附錄 A、Model_Age 甲基化年齡預測模型測試成本一覽表

此附錄展示了 Model_Age 甲基化年齡預測模型在不同測試階段的實際成本數據，這些數據來源於模型在真實測試過程中的成本記錄。本表所提供的成本數據旨在支持報告內「研究結果與討論」中第三點的分析，進一步比較模型的成本效益表現。

(一) 支持向量機模型 (SVM)

表 1 呈現使用不同核函數的支持向量機模型的測試成本，其中表現較佳的為徑向基函數，線性函數和多項式函數位居其後。

表 1、使用不同核函數（線性、多項式和徑向基）建構支持向量機模型之測試成本
（本研究作者自行製作）

Testing Cost	Linear	Polynomial	RBF
Testing Cost 1	6.95E+01	8.14E+01	5.85E+01
Testing Cost 2	7.23E+01	8.62E+01	5.13E+01
Testing Cost 3	8.74E+01	9.32E+01	5.94E+01
Testing Cost 4	8.23E+01	8.35E+01	5.86E+01
Testing Cost 5	7.85E+01	8.01E+01	5.68E+01
Testing Cost 6	6.89E+01	8.86E+01	5.68E+01
Testing Cost 7	7.73E+01	8.66E+01	5.90E+01
Testing Cost 8	7.53E+01	8.14E+01	6.55E+01
Testing Cost 9	6.68E+01	7.77E+01	5.61E+01
Testing Cost 10	8.32E+01	7.46E+01	5.35E+01
Average Testing Cost	7.62E+01	8.33E+01	5.76E+01

(二) 隨機森林模型 (Random Forest)

表 2 呈現使用不同樹數的隨機森林模型的測試成本，其中表現較佳的為樹數為 200 時，取更大或更小的樹數成本皆會增加。

表 2、使用不同樹數建構隨機森林模型之測試成本（本研究作者自行製作）

Testing Cost	100	200	400	600	800
Testing Cost 1	3.65E+01	3.29E+01	3.45E+01	3.63E+01	3.74E+01
Testing Cost 2	4.19E+01	3.24E+01	3.94E+01	4.24E+01	4.15E+01
Testing Cost 3	4.23E+01	3.24E+01	3.53E+01	3.74E+01	3.30E+01
Testing Cost 4	3.74E+01	3.92E+01	3.56E+01	3.59E+01	3.86E+01
Testing Cost 5	3.38E+01	3.66E+01	3.55E+01	3.52E+01	3.49E+01
Testing Cost 6	3.97E+01	3.42E+01	2.95E+01	3.57E+01	3.59E+01
Testing Cost 7	3.74E+01	3.61E+01	3.38E+01	3.66E+01	3.57E+01
Testing Cost 8	3.67E+01	3.48E+01	3.61E+01	3.85E+01	3.58E+01
Testing Cost 9	3.64E+01	3.82E+01	3.94E+01	3.40E+01	3.91E+01
Testing Cost 10	3.52E+01	3.88E+01	4.05E+01	3.60E+01	3.52E+01
Average Testing Cost	3.77E+01	3.56E+01	3.60E+01	3.68E+01	3.67E+01

(三) 線性模型 (Linear Regression)

表 3 呈現線性模型的測試成本，發現其成本明顯較大，考慮運作效率，本研究選擇不採用此模型。

表 3、使用不同學習率建構線性模型之測試成本（本研究作者自行製作）

Testing Cost	0.001	0.005	0.1
Testing Cost 1	3.81E+06	6.44E+05	
Testing Cost 2	3.85E+06	6.50E+05	
Testing Cost 3	3.66E+06	5.13E+05	
Testing Cost 4	3.26E+06	5.18E+05	
Testing Cost 5	2.87E+06	4.48E+05	
Testing Cost 6	3.10E+06	4.95E+05	NAN
Testing Cost 7	3.78E+06	4.83E+05	
Testing Cost 8	3.63E+06	4.97E+05	
Testing Cost 9	3.82E+06	5.07E+05	
Testing Cost 10	4.20E+06	5.30E+05	
Average Testing Cost	3.60E+06	5.29E+05	

附錄 B、年齡相關性高的甲基化位點與基因區域

為探討年齡相關性甲基化位點分布的基因區域，我們發現文獻中有出現 Primal、Hsf4、Kcns1、FHL2、PENK 和 ELOVL2 等基因區域，其中會影響脂肪酸延長的 ELOVL2 多次出現在各個文獻中，被多項研究證實和老化有關。

CpG 位點部分，文獻中記錄到 115 次年齡相關性高位點，其中 6 個沒有出現在台灣人體生物資料庫故不採用，6 個重複出現，因此總共取 103 個位點。表 4 顯示此 103 個位點的染色體、CpG 位點、所在基因、基因區域和 CpG 區域。

表 4、103 個已有文獻證實和年齡具顯著性相關的甲基化位點資料（本研究作者自行整理）

chr	cpgID	gene	gene region	CpG region
CHR5	cg00059225	GLRA1;GLRA1	1stExon;5'UTR	Island
CHR14	cg00107187	TMEM179	1stExon	Island
CHR3	cg00201234	FBLN2	5'UTR	Island
CHR18	cg00399483	DCC	Body	Island
CHR5	cg00911351	PCDHGA3;PCDHGA7; PCDHGB4;PCDHGA4; PCDHGA1;PCDHGA5; PCDHGB1;PCDHGA2; PCDHGB2;PCDHGB3;P CDHGA6	Body;Body;TSS200;Bo dy;Body;Body;Body;B ody;Body;Body;Body	Island
CHR20	cg01293143	TCEA2	TSS200	N_Shore
CHR12	cg01820374	LAG3	Body	N_Shore
CHR7	cg02008154	TBX20;TBX20	1stExon;5'UTR	Island
CHR8	cg02154186	PNMA2;PNMA2	5'UTR;1stExon	Island
CHR17	cg02228185	ASPA;ASPA	1stExon;Body	
CHR6	cg02844545	GCM2;GCM2	1stExon;5'UTR	Island
CHR22	cg02994956	NEFH	1stExon	Island
CHR20	cg03440846	ACSS2	TSS1500	N_Shore
CHR14	cg03734874	TMEM179	TSS1500	Island
CHR19	cg03975694	ZNF540;ZNF540	5'UTR;1stExon	S_Shelf
CHR7	cg04084157	VGF	TSS200	Island
CHR7	cg04528819	KLF14	1stExon	Island
CHR1	cg04977528	HEYL	3'UTR	

chr	cpGID	gene	gene region	CpG region
CHR19	cg05508084	ZNF667;ZNF667	5'UTR;Body	Island
CHR7	cg05822532	ELN	1stExon	
CHR2	cg06092815	SPHKAP	TSS200	Island
CHR3	cg06156376	SHOX2;SHOX2	1stExon;5'UTR	Island
CHR10	cg06291867	HTR7	1stExon	Island
CHR19	cg06572160	KCNC3	1stExon	Island
CHR10	cg06908778	SPAG6	Body	Island
CHR15	cg07306253	DNAJC17;C15orf62;C15orf62	Body;1stExon;5'UTR	
CHR19	cg07408456	PGLYRP2	TSS1500	
CHR1	cg07533148	TRIM58	1stExon	Island
CHR10	cg07621046	C10orf82	TSS200	Island
CHR17	cg08468689	GHDC;GHDC	TSS200;TSS1500	
CHR19	cg08668790	ZNF154	TSS200	S_Shore
CHR9	cg08686553	CDKN2A;CDKN2A	3'UTR;Body	S_Shore
CHR16	cg08872742	CDH5	TSS1500	
CHR9	cg08909157	DOCK8;C9orf66;C9orf66	Body;1stExon;5'UTR	S_Shore
CHR1	cg09809672	EDARADD;EDARADD;EDARADD	TSS1500;5'UTR;1stExon	N_Shore
CHR3	cg10031651	LRRC2;TDGF1	TSS200;Body	S_Shelf
CHR4	cg10235817	ADRA2C	1stExon	Island
CHR16	cg10917602	HSD3B7	5'UTR	
CHR5	cg11136562	ARAP3	5'UTR	
CHR8	cg11981599	GATA4	Body	Island
CHR13	cg12111714	ATP8A2	Body	Island
CHR6	cg12457773	NRSN1;NRSN1	5'UTR;1stExon	Island
CHR7	cg12782180	LEP	TSS1500	Island
CHR14	cg13282837	TCL1A	TSS200	Island
CHR8	cg13434842	GATA4	Body	S_Shore
CHR11	cg13547237	C11orf68;DRAP1	TSS1500;Body	S_Shore
CHR6	cg13603171	MOXD1	TSS1500	S_Shore
CHR13	cg13614181	C13orf15	TSS1500	Island
CHR3	cg13921352	FAM19A4	TSS200	Island
CHR3	cg14456683	ZIC1	TSS200	Island
CHR10	cg14614211	MKX	5'UTR	Island
CHR10	cg14826456	ADRB1	1stExon	Island
CHR16	cg15201635	SMPD3	TSS1500	Island
CHR4	cg15425280	GRIA2	TSS1500	Island
CHR8	cg15747595	TSPYL5	1stExon	Island
CHR12	cg15784615	LTBR	Body	S_Shore
CHR2	cg16232126	SLC5A7;SLC5A7	1stExon;5'UTR	Island
CHR9	cg16386080	CDK20	Body	N_Shore
CHR19	cg16464322	HNRNPL	TSS1500	S_Shore
CHR6	cg16867657	ELOVL2	TSS1500	Island
CHR15	cg17113968	DNAJC17;C15orf62	Body;TSS1500	
CHR18	cg17227156	BRUNOL4	TSS200	Island
CHR1	cg17241310	BARHL2	TSS200	Island
CHR19	cg17471102	FUT3	5'UTR	
CHR18	cg17589341	SLC14A1	TSS200	
CHR19	cg17861230	PDE4C	Body	Island
CHR13	cg18236477	ATP8A2	Body	Island
CHR2	cg18327056	KCNK3	TSS1500	N_Shore
CHR3	cg18328933	ABHD14B;ABHD14B;ABHD14B;ABHD14A	1stExon;Body;5'UTR;TSS1500	N_Shore
CHR15	cg19095568	DNAJC17;C15orf62	Body;TSS200	
CHR19	cg19246110	ZNF671;ZNF671	1stExon;5'UTR	Island
CHR7	cg19594666	LEP	TSS200	Island
CHR17	cg19761273	CSNK1D	TSS1500	S_Shore

CHR11	2.91%
CHR12	2.91%
CHR14	2.91%
CHR16	2.91%
CHR18	2.91%
CHR20	1.94%
CHR22	0.97%
CHRX	0.97%
CHR21	0.00%

表 6、103 個位點的基因區域分布比例（本研究作者自行整理）

CpG region	percentage
Island	58.25%
other	17.48%
S Shore	12.62%
N Shore	8.74%
S Shelf	2.91%

表 7、103 個位點的 CpG 區域分布比例（本研究作者自行整理）

gene region	percentage
TSS1500	19.42%
1stExon	15.53%
Body	13.59%
TSS200	12.62%
1stExon;5'UTR	7.77%
5'UTR	7.77%
5'UTR;1stExon	4.85%
1stExon;Body	1.94%
3'UTR	1.94%
Body;1stExon;5'UTR	1.94%
Body;TSS200	1.94%
1stExon;Body;5'UTR;TSS1500	0.97%
1stExon;Body;Body;Body;Body;Body;Body;Body;	0.97%
5'UTR;Body;Body;Body;Body;Body;Body;Body;	
Body;Body;Body;Body	
3'UTR;Body	0.97%
5'UTR;Body	0.97%
1stExon;Body;Body;Body;Body;Body;Body;Body;5	0.97%
'UTR;Body;Body;Body;Body;Body;Body;Body;Bo	
dy;Body;Body;Body	
Body;TSS1500	0.97%
TSS1500;1stExon	0.97%
TSS1500;5'UTR;1stExon	0.97%
TSS1500;Body	0.97%
TSS200;Body	0.97%
TSS200;TSS1500	0.97%

附錄 C、抽血檢驗生物指數

本研究之生化指數來自周邊血檢驗結果，資料參考台大醫學院檢驗醫學部[20]。表 8 為各生化指數之中文及英文名稱、本研究所用代碼、正常值範圍和生物學意義說明。

表 8、26 種生化指數的詳細資訊（本研究作者自行整理）

中英名稱	代碼和正常值	說明
糖化血色素 HbA1c	HBA1C 4.0-6.0%	葡萄糖與血紅素結合而成，數值較低可能為血紅病變、紅血球代謝速率改變或腎功能不佳等。
空腹血糖 Glu-AC	FASTING_ GLUCOSE 70-100mg/dL	空腹血液中葡萄糖的含量，為糖尿病指標。
血中尿素氮 BUN	BUN 7-25mg/dL	由腎臟過濾排出的含氮化合物，數值較高可能為腎臟功能不佳所致。
天門冬胺酸轉胺酶 AST (GOT)	SGOT 8-31U/L	主要分布於肝臟的酵素，數值較高可能是肝臟損毀，或與心肌梗塞、溶血及嚴重骨骼肌問題有關。
丙胺酸轉胺酶 ALT (GPT)	SGPT 0-41U/L	肝臟酵素，數值較高可能為肝病或肝損害所致，較低則較罕見。
紅血球 RBC	RBC 男性 4.21-5.9 女性 3.78-5.25 10 ⁶ /uL	負責運輸氧氣，數值較高可能為缺氧、肺部疾病或脫水，較低可能為貧血、出血或慢性疾病。
血小板 PLT	PLATELET 150-378 10 ³ /uL	負責傷口血液凝固和癒合的血球。數值較高可能為感染、發炎或骨髓增生疾病，較低可能為凝血障礙、骨髓疾病或自體免疫疾病。
白蛋白 Alb	ALBUMIN 3.5-5.7g/dL	主要由肝臟製造，維持血液滲透壓和平衡，數值較高可能為脫水，較低可能為營養不良或肝腎疾病。
抗 B 型肝炎核心抗原抗體 Anti-HBc	ANTI_HBC_AB Negative 小於 1 Positive 大於等於 1 S/CO	陽性 (≥1S/CO) 表示曾感染 B 型肝炎。
抗 B 型肝炎表面抗原抗體 Anti-HBs	ANTI_HBS_AB Negative 小於 10.00 mIU/mL Positive 大於等於 10.00 mIU/mL	陽性表示有感染 B 型肝炎。
抗 C 型肝炎抗體 Anti-HCV	ANTI_HCV_AB Negative 小於 1 Positive 大於等於 1 S/CO	陽性表示有感染 C 型肝炎。
B 型肝炎 e 抗原 HBeAg	HBEAG Negative 小於 1 Positive 大於等於 1 S/CO	陽性表示 B 型肝炎病毒活性高。

中英名稱	代碼和正常值	說明
B 型肝炎表面抗原 定量分析 HBsAg	HBSAG Negative 小於 1 Positive 大於等於 1 S/CO	陽性表示無感染 B 型肝炎。
尿酸 U.A	URIC_ACID 男性 4.4-7.6mg/dL 女性 2.3-6.6mg/dL	腎臟排出的代謝產物，數值較高可能為痛風或腎臟疾病，較低可能為藥物影響或腎臟疾病。
肌酸酐 CRE	CREATININE 0.6-1.3mg/dL	肌肉代謝產物，數值較高可能為腎功能受損所致，較低可能為肌肉質量低或營養不良所致。
總膽紅素 TBil	T_BILIRUBIN 0.3-1.0mg/dL	為紅血球分解產物，數值較高可能為肝病、膽道阻塞所致，較低則較罕見。
低密度脂蛋白 (壞膽固醇) LDL-C	LDL_C 小於 130mg/dL	動脈硬化相關膽固醇，數值較高可能為心血管疾病，較低則較罕見。
高密度脂蛋白 (好膽固醇) HDL-C	HDL_C 大於 40mg/dL	一種好膽固醇，數值較低可能為心血管疾病。
總膽固醇 T-CHO	T_CHO 小於 200mg/dL	血液中膽固醇的總和，數值較高可能為心血管疾病，較低則較罕見。
血比容 Hct	HCT 男性 39.6-51.5% 女性 34.8-46.3%	血液中紅血球的百分比，數值較高可能為脫水、紅血球增多症，較低可能為貧血及失血。
血色素 Hb	HB 男性 13.1-17.2g/dL 女性 11.0-15.2g/dL	紅血球中運輸氧氣的蛋白質，數值較高可能為脫水、紅血球增多症，較低可能為貧血及失血。
白血球 WBC	WBC 3.25-9.16 10 ³ /uL	免疫系統的一部分，數值較高可能為感染、發炎或白血病，較低可能為免疫功能低下或藥物影響。
微蛋白尿 Micro-albumin	microALB 小於 30mg/L	腎功能早期指標，數值較高可能為腎病。
麩胺酸轉胺酵素 γ-GT	GAMMA_GT 9-64U/L	肝臟酵素，數值較高可能為肝病、酗酒或藥物影響，較低則較罕見。
甲型胎兒蛋白 AFP	AFP 小於 20ng/mL	測肝癌和胎兒疾病，數值較高可能為肝癌或睪丸癌，較低則較罕見。
三酸甘油酯 TG	TG 小於 150mg/dL	脂肪代謝物，數值較高可能為心血管疾病所致，較低則較罕見，可能營養不良。

【評語】 090001

1. 創新性與重要性：

本研究在老化研究領域展現了顯著的創新性和重要性。首先，研究使用了台灣大量 CpG 位點的甲基化數據，結合血液生化指數和疾病問卷資料，提供了一個全面的老化研究數據集。藉由分析甲基化位點與年齡的關係，研究將表觀遺傳學與老化研究相結合，開闢了理解老化機制的新途徑。此外，研究建立了基於隨機森林算法的年齡預測模型（Model_Age），為老化研究提供了一個有力的工具。最重要的是，研究結果有望為老化相關血液疾病提供個人化診斷與治療策略，推動精準醫療的發展。

2. 研究優點：

該研究具有幾個突出的優點。首先，研究通過多維度數據整合，結合了甲基化數據、血液生化指數和疾病問卷資料，提供了全面的研究視角。其次，研究採用了創新的分類方法，通過比較實際年齡與預測的甲基化年齡，將數據分為「表觀遺傳年輕」與「表觀遺傳年老」兩類，為後續研究提供了新的思路。研究還發現這

兩類在特定血液生化指數上存在顯著差異，為進一步探討疾病關聯性奠定了基礎。最後，研究計劃檢驗台灣受試者與國際結果的一致性，有助於了解老化過程的普遍性和特殊性。

3. 待改進之處：

該研究仍有幾個方面需要改進。首先，研究的方法學細節不足，未詳細描述模型參數選擇過程和評估指標，這不利於其他研究者進行驗證或重現。其次，研究在因果關係探討方面存在不足，雖然表明了甲基化年齡與健康指標的相關性，但未深入探討其因果關係。此外，研究應進一步分析差異性 CpG 位點的甲基化如何影響特定基因的表達，以及這些基因表達變化如何影響信息傳遞或代謝反應，最終導致老化。最後，研究需要更深入地解釋表觀遺傳年輕與表觀遺傳年老之間的分子機制差異，以及這些差異如何造成生化指標的變化。通過改進這些方面，該研究將能更全面地揭示老化過程中的表觀遺傳學機制，並為精準醫療提供更可靠的理論基礎。