**臺北醫學大學**

**高教深耕111年度「轉譯創新研究計畫」計畫書**

1. **基本資料：**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 領域別 | | | ◻癌症轉譯 ◻神經醫學 ◻胸腔醫學 ◻人工智慧醫療 ■其他創新 ◻免疫監測 | | | | |
| 精準健康類別 | | | ■精準預防 ■精準診斷 ◻精準治療 ◻精準照護 ◻其他 | | | | |
| 研究型別 | | | ■整合型 ◻個別型(限人工智慧醫療、免疫監測及其他) | | | | |
| 總計畫名稱 | 中文 | | 開發高危險口腔癌前病變的精準診斷治療策略:聚焦口腔疣狀增生 | | | | |
| 英文 | | The development of theranostic approaches for high-risk oral potentially malignant disorders (OPMDs): focusing on oral verrucous hyperplasia (OVH) | | | | |
| 總計畫主持人姓名 | | | 吳明恒 | 職稱 | 副教授/主任 | 單位 | 醫學科技學院/國際轉譯科學博士學位學程 |
| 子計畫名稱  (個別型計畫免填) | | 中文 | 以深度學習及風險預測模式探討口腔疣狀增生惡轉風險 | | | | |
| 英文 | Deep Learning Platform And Risk Prediction Models For Malignant Transformation of Oral Verrucous Hyperplasia | | | | |
| 子計畫主持人 | | | 吳家佑 | 職稱 | 醫師/助理教授 | 單位 | 北醫附醫口腔顎面外科 |
| 電話 |  | E-Mail | borgiawu@gmail.com |
| 全程執行期限 | | | 自民國111年01月01日起至民國111年12月31日 | | | | |
| 本計畫是否有進行下列實驗：（勾選下列任一項，須附相關實驗之同意文件）  ■人體實驗/人體檢體 ◻人類胚胎/人類胚胎幹細胞 ◻基因重組實驗 ◻基因轉殖田間試驗 ◻第二級以上感染性生物材料 ◻動物實驗 | | | | | | | |
| 計畫連絡人 | | | 姓名： 王紋璋 電話： E-Mail： | | | | |

總計畫主持人簽章： 日期：

子計畫主持人簽章： 日期：

子計畫主持人簽章： 日期：

子計畫主持人簽章： 日期：

子計畫主持人簽章： 日期：

1. **申請補助經費：**
2. 請將本計畫書之第四項(表A004)、第五項(表A005)及第六向(表A006)所列費用個別加總後，分別填入「人事費」、「業務費」欄內。

　　　　　 　　　金額單位：新台幣元

|  |  |
| --- | --- |
| 執行年次  補助項目 | 111 年  (111年01月01日起至111年12月31日) |
| **人事費**  **(專任、兼任研究助理)** | 0 |
| **業務費**  **(臨時工資、耗材及雜支)** | 500,000.- |
| 合 計 |  |

**備註：請注意！！！**

1. **因應教育部高教深耕計畫相關規定，人事費及業務費不得互相流用。**
2. **臨時工資屬業務費。**
3. **經費使用方式僅限於人事費及業務費，不得用於出國補助及資本門(購買儀器設備，定義：耐用年限二年以上且金額1萬元以上者皆屬資本門，如電腦、建置平台等)。**
4. **業務費需於11/30前核銷完畢。**
5. **主要研究人力：**
6. 請依照「主持人」、「共同主持人」、「協同研究人員」及「博士後研究」等類別之順序分別填寫。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 類 別 | 姓名 | 服務機構/系所 | 職稱 | 在本研究計畫內擔任之具體工作性質、項目及範圍 | ＊每週平均投入  工作時數比率(%) | ORCID ID |
| 主持人 | 吳家佑 | 北醫附醫牙科部 | 主任 | 方法開發、招募病患、結果統整 | 20 % |  |
| 共同主持人 | 王紋璋 | 轉譯醫學博士學位學程 | 副教授 | 方法開發、臨床樣品應用、數據分析、結果統整 | 50 % |  |
| 共同主持人 | 祁力行 | 萬芳醫院口腔顎面外科 | 主治醫師 | 招募病患、方法開發、臨床樣品應用、數據分析 | 30 % | 0000-0002-4476-2600 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

* 註：每週平均投入工作時數比率係填寫每人每週平均投入本計畫工作時數佔其每週全部工作時間之比率，以百分比表示（例如：50%即表示該研究人員每週投入本計畫研究工作之時數佔其每週全部工時之百分五十）。

1. **研究人力費(人事費)：**
2. 類別/級別欄請依專任助理(含碩士、學士)、兼任助理(含博士生、碩士生、大專學生)等填寫。
3. 專任助理及兼任助理之每月工作酬金標準，請參考本校專題研究計畫專任助理人員工作酬金暨博士後研究員教學研究費用表及科技部補助專題研究計畫兼任助理人員工作酬金支給標準表之規定。
4. 申請專任助理者，除依工作月數填列工作酬金及至多1.5個月年終工作獎金外，須另填列投保勞保及健保之「雇主應負擔之勞、健保費及勞退」；勞務型兼任助理及臨時工資須另填列投保勞保之「雇主應負擔之勞保費」。

金額單位：新台幣元

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| （一）專任助理、勞務型兼任助理 | | | | | | |
| 類別/級別 | 人數 | 姓 名 | 工 作  月 數 | 月支酬金  （含勞健保費） | 小計 | 請述明：1.最高學歷2.曾擔任專題研究計畫專任助理之經歷3.在本計畫內擔任之具體工作性質、項目及範圍 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| 合　　計（一） | | | | |  | |
| （二）博士班研究生、碩士班研究生及大專學生兼任助理(學術型) | | | | | | |
| 級別/姓名 | 人數  （1） | 月支酬金(2) | 獎助月數(3) | 小計 (4)＝(1)×(2)×(3) | | 在本研究計畫內擔任之具體工作性質、項目及範圍 |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
|  |  |  |  |  | |  |
| 合計（二） | | | | | |  |
| 總計（三）＝合計（一）＋合計（二） | | | | | |  |

1. **臨時人員(業務費)：**

臨時人員(工讀生)至多聘任至111.11.30，111.12不得聘用臨時人員。

金額單位：新台幣元

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| （一）**臨時工資** | | | | | | |
| 類別/級別 | 人數 | 姓 名 | 工 作  月 數 | 月支酬金  （含勞健保費） | 小計 | 請述明：1.最高學歷2.在本計畫內擔任之具體工作性質、項目及範圍 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| 合　　計（一） | | | | | |  |

1. **耗材、物品、圖書及雜項費用：**
2. 凡執行研究計畫所需之耗材、物品(非屬研究設備者)及雜項費用等，均可填入本表內。
3. 說明欄請依該項目之規格、用途等相關資料詳細填寫。

金額單位：新台幣元

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 項目名稱 | 說明 | 單位 | 單價 | 數量 | 金額 | 備註 |
| 耗材費 | 研究計畫人體試驗委員會審查費 | 次 | 12000 | 1 | 12000 | 審查10000  修正1000  結案1000 |
| 耗材費 | TWB2.0 DNAchip | 盤 | 96000 | 2 | 192000 | 一盤96樣本，施作兩盤 |
| 耗材費 | 抽DNA | 例 | 180 | 192 | 34560 | 以上述兩盤樣本數計 |
| 耗材費 | TWB2.0 DNAchip檢體送樣前QC | 例 | 50 | 192 | 9600 | 以上述兩盤樣本數計 |
| 耗材費 | 資料儲存硬碟 | 顆 | 9890 | 1 | 9890 |  |
| 耗材費 | 國網運算資源 |  |  |  | 3950 |  |
| 耗材費 | 生物資料庫使用費(切片費) | 5片 | 330 | 30 | 9900 |  |
| 耗材費 | 生物資料庫使用費(HE染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(FAT1染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(p53染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(Ki-67染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(CK13染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(CK17染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(EGFR染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(*α*-SMA染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(CAMK2N1染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(p16INK4a染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | 生物資料庫使用費(p14ARF染色費) | 片 | 40 | 150 | 6000 |  |
| 耗材費 | IHC 免疫染色抗體 | 支 | 13000 | 9 | 117000 |  |
| 儀器使用費 | Aperio ScanScope 數位病理切片掃描 | 小時 | 300 | 105 | 31500 |  |
| 雲端運算 | 國網中心高速電腦GPU人工智慧訓練 | 小時 | 689 | 19.7 | 13600 | cm.4xsuper |
| 合計 | | | | | 500,000 |  |

1. **整合型研究計畫項目及重點說明：**（總計畫及子計畫之主持人均需填寫此表）
2. 整合型研究計畫項目：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 計 畫 項 目 | 主持人 | 服務機構/系所 | 職稱 | 計 畫 名 稱 | 計畫經費  (新台幣元) |
| 總計畫 | 吳明恒 | 國際轉譯科學博士學位學程 | 副教授/主任 | The development of theranostic approaches for high-risk oral potentially malignant disorders (OPMDs): focusing on oral verrucous hyperplasia (OVH) | 1500000 |
| 子計畫一 | 吳明恒 | 國際轉譯科學博士學位學程 | 副教授/主任 | Translational research of FAT1 gene mutation in malignant transformation of oral verrucous hyperplasia | 500000 |
| 子計畫二 | 蔡伊琳 | 生物化學暨細胞分子生物學科 | 副教授 | Discover potential biomarkers from human saliva, plasma, and extracellular vesicles for prognosis evaluation of oral verrucous  hyperplasia to oral cancer | 500000 |
| 子計畫三 | 吳家佑 | 北醫附醫口腔顎面外科 | 醫師/助理教授 | Risk prediction models for malignant transformation of oral verrucous hyperplasia | 500000 |
| 合計 | | | | | 1500000 |

1. 整合型研究計畫重點說明：如為子計畫請說明與總計畫之關聯性  
   請就下列各點分項述明：
   1. 整合之必要性：包括總體目標、整體分工合作架構及各子計畫間之相關性與整合程度。
   2. 人力配合度：包括總計畫主持人協調領導能力、各子計畫主持人之專業能力及合作諧和性。
   3. 資源之整合：包括各子計畫研究經驗與成果交流情況。
   4. 申請單位或其他單位之配合度。
   5. 預期綜合效益。

OBJECTIVES

Tis project aims to develop a prediction model for the malignant potentials of OVH through integration of molecular cell biology, saliva secretome, genetic statistics, and AI images. In addition, chemo-preventive strategies will be developed for the high-risk population of OVH by dissecting the molecular features, particularly focusing on FAT1-regulated Hippo/YAP1 signaling. The results will be very important not only for providing informed selection for optimal management approaches such as intensive follow-up or surgery, but also for deciphering the molecular mechanisms of OVH transformation therefore to develop novel treatment strategy.

THE PROJECTS and INTERGRATION

The present proposal consists of four projects which brings together different groups of investigators with strong complementary expertise in molecular cell biology, saliva extracellular vesicles (EVs) and proteomics, genetic statistics, and artificial intelligence (AI) of clinical and pathological images. This team will work together to study the cellular and genetic signatures, saliva biomarkers, medical images of OVH, for the precision medicine of OVH theranostics.

**SP1** (Dr. MH Wu吳明恒Associate professor/Director of Ph.D. Program for Translational Medicine) aims to dissect the cell and molecular signatures of OVH and focus on the reciprocal interactions of FAT1 and p53 in the malignant transformation of OVH. Dr. Wu is a molecular and cell biologist. His recent publications were listed here (Oncogene 2022(revision), Oncogene 2018, Head&Neck 2018, Scientific Reports 2017, Nature communication 2017).

**SP2** (Dr. IL Tsai蔡伊琳Associate professor/Department of Biochemistry and Molecular Cell Biology) will set up a comprehensive workflow to monitor potential protein and RNA components in saliva and plasma, and those enriched in extracellular vesicles (EVs). Dr. Tsai’s research focus on EV- and immunoglobulin-(Ig) protemics. Her recent publications were listed (Microbiology Spectrum 2021, Journal of Personalized Medicine 2021, International Journal of Molecular Sciences 2021).

**SP3** (Dr. CY Wu吳家佑, Director of Oral and Maxillofacial Surgery/Taipei Medical University Hospital) and (Dr. WC Wang王紋璋, Associate professor/Ph.D. Program for Translational Medicine) plans to establish an OVH cohort and collect medical images of OVH patients. Dr. Wu is an oral surgeon and is constructing an AI image system for early-diagnosis of OSCC. His recent publications were listed here (International Journal of Oral and Maxillofacial Surgery 2021, Applied Sciences 2020, Cancers 2020) Dr. Wang aims to develop risk prediction models for malignant transformation among patients with OVH based on the risk factors obtained from all subprojects. Dr. Wang is specialized in genetic statistics and GWAS analysis. His recent publications were listed here (Frontiers in Immunology 2021, Scientific Reports 2021, Diabetologia 2021).

We also invite Dr. WF Chiang 蔣維凡 (Director of Department of Oral and Maxillofacial Surgery/Chi Mei Medical Center, Louying) and Dr. LH Chi祁力行 (Director of Oral and Maxillofacial Surgery/Taipei Municipal Wanfang Hospital) to be Project Co-PI for clinical advises and sample collections and invite Dr. JW Shih 施景文 (Assistant professor/Graduate Institute of Cancer Biology and Drug Discovery) to help the analysis of miRNA and lncRNA in EVs. The integration among each sub-project is listed as below.

All four PIs and Co-PIs already had joint-publications (Nature communications 2017, Oncogene 2018, and Head Neck 2018) and formed a closed working group in the past five years in the TMU Deep Route PPG 臺北醫學大學深耕創新轉譯整合型計畫 (2017-2021) and MOST grants.

1. **研究計畫中英文摘要：**請就本計畫要點作一概述，並依本計畫性質自訂關鍵詞。
2. 計畫中文摘要。（五百字以內）

以深度學習及風險預測模式探討口腔疣狀增生惡轉風險

頭頸部鱗狀細胞癌(head and neck squamous cell carcinoma (HNSCC)) 在全世界都是重要的問題。手術、放射治療和全身性治療(化療、免疫)仍然是這些患者的標準治療方式。如何減少死亡率，取決於有效的預後生物標記(創新療法)，並早期預警口腔黏膜病變的癌化(加強預防)。

自2008年至2018年，臺灣男性罹患頭頸部鱗狀細胞癌的一年存活率為79.56% 至81.62%（健康促進管理局癌症登記處的統計數據，網址https://cris.hpa.gov.tw/pagepub/Home.aspx，2022年5月查詢）。2009年到2014年，男性患者的的五年存活率一直沒有進步(從55.13%到56.03%)。

利用The Cancer Genome Atlas (TCGA)資料庫做存活分析，是找尋頭頸部鱗狀細胞癌生物標記最常用的方法。癌症的生物標記已被用於診斷、預測病情與疾病的預後。預測性標記有助於選擇有利的治療方式。例如，腫瘤切除手術後，若是手術切緣不足、頸部淋巴結轉移，則應進行術後的放射治療，以增加治療的成功率。

口腔疣狀增生(oral verrucous hyperplasia (OVH))是一種外形為疣狀(或乳頭狀)的癌前病變；這種病變日後有機會轉化為疣狀癌(verrucous carcinoma, VC)或鱗狀細胞癌(squamous cell carcinoma, SCC)。我們的目標是找出預測性生物標記，期望無需頻繁切片，就能早期預警正在癌化中的口腔疣狀增生。

我們使用R語言開發”pvalueTex”程式，可執行資料檢索、資料處理、特徵選擇、資料探勘、Kaplan–Meier存活分析、Cox比例風險模型，以用來大規模搜尋生物標記。我們接下來計劃開發深度學習(Deep Learning)資料探勘和驗證平台，利用數位病理影像(the Cancer Imaging Archive (TCIA), 與TCGA串連)，及臺北醫學大學(TMU)生物資料庫中，口腔疣狀增生的病理切片數位影像，以AI人工智慧演算法，來做存活分析，預測(癌)惡轉變化。深度學習的卷積神經網路(Convolutional Neural Network (CNN))可由數位病理影像中提取特徵。圓論卷積神經網路(Graph Convolutional Neural Network (GCNN))則處理電子病歷的”全人照護特徵”。這些患者的特徵應該包括身體、病理、心理狀態，社會關係，甚至全人照顧醫師所紀錄的靈性紀錄。

結論：通過TCGA分析，我們得到生物標記，並確認臨床腫瘤大小和手術切緣狀態，都會是頭頸部鱗狀細胞癌患者的重要預後因素。使用這些生物標誌物，我們將分析口腔疣狀增生檢體的基因變化，結合數位病理影像，以開發鱗狀細胞癌的早期預警模型。

關鍵詞: 頭頸部鱗狀細胞癌; 口腔鱗狀細胞癌; 口腔疣狀增生; 存活分析; 手術切緣; 手術切緣; 卷積神經網路; 圓論卷積神經網路; 全人照顧.

1. 計畫英文摘要。（五百字以內）

Background: Head and neck squamous cell carcinoma (HNSCC) represents a significant health concern worldwide. Surgery and systemic therapy are still the standard of care for HNSCC patients. Oral squamous cell carcinoma (OSCC) is the most common part of HNSCC. The break-through improvement of those interventions should depend on 1) the discovery of high impact prognostic biomarkers, and 2) early-detection of potentially malignant oral lesion.

During 2008 to 2018 in Taiwan, the one-year survival rate for this disease among males is from 79.56% to 81.62% (statistics from Cancer Registry of Health Promotion Administration, available at https://cris.hpa.gov.tw/pagepub/Home.aspx, accessed on May 2022). The five-year survival rate for this disease among males ranged from 55.13 percent to 56.03 percent between 2009 and 2014.

The survival analysis of the the Cancer Genome Atlas (TCGA) dataset is a well-known method to discover biomarkers of HNSCC. Biomarkers of cancer have been utilized for diagnostic, predictive or prognostic purposes. A predictive marker helps to select which specific treatment benefits a better survival. For example, adjuvant radiation therapy should be done after tumor-ablative surgery in positive surgical margin or metastatic lymph node from the neck dissection.

Oral verrucous hyperplasia (OVH) is a pre-malignant mucosal lesion with a predominantly verrucous or papillary surface; this lesion can subsequently transform into verrucous carcinoma (VC) or squamous cell carcinoma (SCC). A predictive biomarker for OVH could provide early warning before malignant transformation without the need for frequent biopsies.

Methods:

R script was used to develop a comprehensive workflow, named ”pvalueTex”, running on the Rstu- dio platform. It includes data retrieving and pre-processing, feature selection, cutoff mining engine, Kaplan-Meier survival analysis, Cox proportional hazard modeling to discover prognostic biomarkers. We further plan to develop a deep-learning based data mining as well as validation platform. It might incorporate whole-slide images of pathology from cohort of Taipei Medical University (TMU) and Chi- Mei Hospital. Convolutional neural network (CNN) will be used for feature extraction from whole-slide images. Graph convolutional neural network (GCNN) and explainable autoencoder will deal with ”sur- vival analysis”. These patient’s features should include physical, pathological, psychological data, and even more spiritual information investigated by physicians. During their follow-up period, a ”survival analysis” should also be used to build a predictive model of malignant transformation.

Conclusions: The candidate biomarkers, clinical tumor size, and surgical margin status are sug- gested prognosis factors in HNSCC through TCGA analysis. Using those biomarkers, we will analyze the transcriptomic/mutational representatives of OVH in our cohorts. We will conclude that combin- ing genomics data with pathology images will aid in the development of a predictive model for early detection of OSCC.

Keyword: Head and Neck Squamous Cell Carcinoma (HNSCC); Oral Squamous Cell Carcinoma (OSCC); Oral Verrucous Hyperplasia (OVH); the Cancer Genome Atlas (TCGA); the Cancer Imaging Archive (TCIA); RNA-sequencing; Survival Analysis; Surgical Margin; Deep Learning; Explainable; Graph Neural Network (GNN); Graph Convolutional Neural Network (GCNN); Holistic Cancer Care.

1. **研究計畫內容：**(中英文不限)
2. 研究計畫內容與主要研究成果說明。
3. 研究計畫之背景及目的。請詳述本研究計畫之背景、目的、重要性及國內外有關本計畫之研究情況、重要參考文獻之評述等。
4. 研究方法、進行步驟及執行進度。1.本計畫採用之研究方法與原因。2.預計可能遭遇之困難及解決途徑。請就以上各點分別說明子計畫之相關性。
5. 預期完成之工作項目、成果及績效。1.預期完成之工作項目。2.對於學術研究、國家發展及其他應用方面預期之貢獻。3.對於參與之工作人員，預期可獲之訓練。4.預期完成之研究成果及績效（如期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期績效）。5.本計畫如為整合型研究計畫之子計畫，請就以上各點分別說明與其他子計畫之相關性。

**(**一**)**研究計畫之背景及目的

## Background

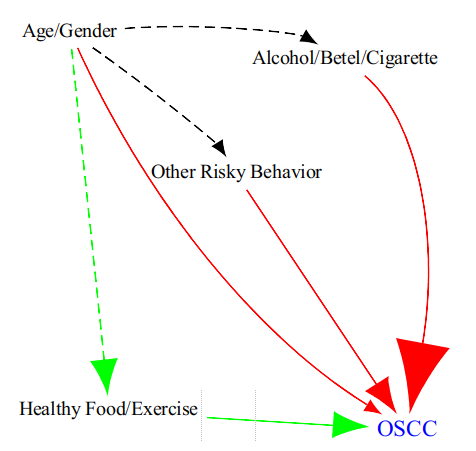
* 1. **Cancer Initiation**

The main risk factors for [OSCC](#_bookmark0) include carcinogens in tobacco and betel nut, alcohol, oncogenic viruses. The betel-nut chewing, alcohol [aldehyde dehydrogenase 2](#_bookmark0) ([ALDH2](#_bookmark0)) mutation should be the major concern as a variety of [OSCC](#_bookmark0) in Taiwan. In Asian population, [aldehyde dehydrogenase 2](#_bookmark0) ([ALDH2](#_bookmark0)) mutation plays a very important rule to enhance [OSCC](#_bookmark0) ([Chien et al. 2019](#_bookmark13)) by alcohol consumption. The oncogenic viruses such as [human papillomavirus](#_bookmark0) ([HPV](#_bookmark0)) and Epstein–Barr virus play a major carcinogenic role in tumors of the oropharynx and nasopharynx, respectively ([Ferrarotto et al. 2017](#_bookmark19)). Moreover, physical irritations, ionized radiation and host susceptibility may also contribute to the etiology of [OSCC](#_bookmark0) ([Ko et al. 1995](#_bookmark25); [Znaori et al.](#_bookmark45) [2003](#_bookmark45)) (Figure [1](#_bookmark1)). Primary prevention by minimizing exposure to known carcinogens is promising, but limited success is achieved due to addiction, economic, social or spiritual issues. The possible molecular mechanisms could be also related with the mind-brain-body axis ([Berens et al. 2017](#_bookmark7)).

## Molecular Oncology

Accumulative knowledge shows that some drivers could facilitate carcinogenesis in [OSCC](#_bookmark0). For example, p53 is a tumour suppressor protein encoded by the TP53 gene. The transcription factor, p53, that is acti- vated by many genotoxic insults and induces cellular apoptosis. The gene TP53 is frequently mutated or functionally inactivated in cancer cells. TP53 mutations are comprised of eight classes: missense, nonsense, frame-shift deletion, frame-shift insertion, in-frame deletion, in-frame insertion, silent and splice-site. Im- munohistochemical (IHC) staining for p53 is used as a surrogate for TP53 missense mutation in cancer cells. The mutations of TP53 are one of the most frequent abnormalities in [OSCC](#_bookmark0) found within up to 70% of surgical specimen ([Alsner et al. 2001](#_bookmark5); [Singh et al. 2016](#_bookmark38); [Wang & Sun 2017](#_bookmark42); [Ward & Helman 2018](#_bookmark43)). It is also found in severe dysplasia of [OVH](#_bookmark0). Ki-67, encoded by MKI67 gene, is a measurement of the proliferative rate of [OSCC](#_bookmark0) ([Silva et al. 2004](#_bookmark39)). Overexpression of Ki-67 impacts on poor overall survival of [OSCC](#_bookmark0) ([Perisanidis et al. 2012](#_bookmark34); [Szentku´ti et al. 2015](#_bookmark40)). The epidermal growth factor receptor (EGFR) is found in 70% of [OSCC](#_bookmark0) ([O-charoenrat et al. 2000](#_bookmark30); [Bentzen et al. 2005](#_bookmark8)). It is involved in cancer cells evading immunosurveillance ([Stefanidakis & Koivunen 2006](#_bookmark41)). Chronic mechanical and chemical irritation from be- tel nut chewing could increase myofibroblasts in terms of developing oral submucous fibrosis ([Angadi et al.](#_bookmark6) [2011](#_bookmark6)) and [OSCC](#_bookmark0). Myofibroblasts are contractile cells expressing [actin alpha 2, smooth muscle](#_bookmark0) ([ACTA2](#_bookmark0)) (encoding alpha-smooth muscle actin, *α*-SMA) when is stimulated by betel nut extract—arecoline—*in vitro* ([Chang et al. 2014](#_bookmark10)). A elevated [ACTA2](#_bookmark0) expression is also associated with poor overall survival in [OSCC](#_bookmark0) ([Marsh et al. 2011](#_bookmark27)). Moreover, overexpression of [ACTA2](#_bookmark0) in myofibroblastic cancer-associated fi- broblast (myCAF) could promote stemness of [OSCC](#_bookmark0) ([Patel et al. 2018](#_bookmark33); [Joshi et al. 2021](#_bookmark24)).

The possibility of malignant transformation is influenced by the severity of epithelial dysplasia on [OVH](#_bookmark0) samples. Diagnostic markers are those that distinguish between different categories of lesions. They are obtained by comparing cross-sectional samples taken from various patients. Diagnostic biomarkers, such as KRT13 (type I cytoskeletal 13, encoding Cytokeratin13—CK13), KRT17 (type I cytoskeletal 17, encod- ing Cytokeratin 17—CK17), TP53 (encoding p53), and MKI67 (Marker Of Proliferation Ki-67, encoding Ki-67), are applied in routine pathological examination for dysplasia grading (semi quantitative) , and a confirmation of [OSCC](#_bookmark0). The expressions for CK17 and CK13 are reciprocal in oral dysplasia lesions and that the CK17/p53 emergence is helpful to confirm a diagnosis of [OSCC](#_bookmark0) ([Mikami et al. 2011](#_bookmark28)).



**Figure 1: OSCC will be caused by carcinogen or other factors.** From the intrinsic factors (such as age and gender), the accumulated genetic alteration will be part of etiology of cancer. Carcinogen from alcohol, betel nut and/or cigarette will definitely induce [OSCC](#_bookmark0) by time passes. People usually learn ”risky” or ”healthy” behaviors by many reasons, which modify their possibility to get cancer.

(red line: increasing the chance of getting [OSCC](#_bookmark0); green line will reduce it; black dot: indicate unknown reasons for taking those habits)

## Preliminary Results

Using pvalueTex ([Chi et al. 2021](#_bookmark11)) on the [TCGA HNSCC](#_bookmark0) cohort, we scanned human protein-coding genes (20,500) programmatically. After adjustment with other confounders, the clinical tumor stage and the surgical margin involvement are independent risk factors in patient survival.

And overexpression of CAMK2N1 is significantly associated with the poor prognosis of [overall survival](#_bookmark0) ([OS](#_bookmark0)) (Table [1](#_bookmark2)).

**Table 1:** Clinical (T)umor size, positive surgical margin, and CAMK2N1 expression are independent prognostic factors in HNSCC (TCGA) cohort.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | |  | | | **Multivariate** | | |
| **HR** | **CI95%** | ***p* Value** | **HR** | **CI95%** | ***p* Value** |
| Gender | Female | 1 |  |  | 1 |  |  |
| Male | 1.157 | 0.843–1.587 | 0.367 | 1.076 | 0.767–1.510 | 0.671 |
| Age at diagnosis | *<*= 65*y* | 1 |  |  | 1 |  |  |
| *>* 65*y* | 1.329 | 0.990–1.784 | 0.058 | 1.391 | 1.025–1.888 | 0.034 |
| Clinical T Status | T1+T2 | 1 |  |  | 1 |  |  |
| T3+T4 | 1.409 | 1.028–1.931 | 0.033 | 1.982 | 1.048–3.745 | 0.035 |
| Clinical N Status | N0 | 1 |  |  | 1 |  |  |
| N1-3 | 1.185 | 0.890–1.577 | 0.246 | 1.145 | 0.801–1.636 | 0.457 |
| Clinical M Status | M0 | 1 |  |  | 1 |  |  |
| M1 | 4.097 | 1.009–16.644 | 0.049 | 7.314 | 1.590–33.631 | 0.011 |
| Clinical Stage | Stage I+II | 1 |  |  | 1 |  |  |
| Stage III+IV | 1.245 | 0.882–1.759 | 0.213 | 0.621 | 0.287–1.343 | 0.226 |
| Surgical Margin status | Negative | 1 |  |  | 1 |  |  |
| Positive | 1.591 | 1.155–2.191 | 0.004 | 1.631 | 1.182–2.250 | 0.003 |
| Tobacco Exposure | Low | 1 |  |  | 1 |  |  |
| High | 1.364 | 1.008–1.844 | 0.044 | 1.363 | 0.990–1.875 | 0.058 |
| CAMK2N1 Expression | Low | 1 |  |  | 1 |  |  |
| High | 2.101 | 1.572–2.809 | 5*.*324 *×* 10*−*7 | 2.007 | 1.490–2.704 | 4*.*565 *×* 10*−*6 |

(OS: overall survival; HR: hazard ratio; CI95%: 95% confidence interval; *p* value significant code is denoted: red *<* 0.05).

Wu and his colleagues (Wu et al., 2018, 2020, 2021) have serial studies for [OVH](#_bookmark0) through whole-exome sequencing to identify a biomarker. They found a candidate biomarker for malignant transformation: [FAT](#_bookmark0) [atypical cadherin 1](#_bookmark0) ([FAT1](#_bookmark0)). The [FAT1](#_bookmark0) missense mutation rate is higher in progressive [OVH](#_bookmark0) (83.3%) than in their non-progressive counter partner (22.2%), while 6.25% of TP53 is mutated in [OVH](#_bookmark0). Missense mutation of FAT1 is also significantly associated with poor recurrence-free survival (RFS) of [OSCC](#_bookmark0).

As a result, we believe that FAT1-regulated YAP1/Hippo pathway drives malignant transformation in [OVH](#_bookmark0). We assume those biomarkers, CK13, CK17, p53, Ki-67, EGFR, *α*-SMA, CAMK2N1, and FAT1, should be found within surgical margin (dysplasia zone) of [OSCC](#_bookmark0) specimen and [OVH](#_bookmark0). Further investigation will be conducted by using deep learning of whole-slide pathology images.

## Deep Learning in Bioinformatics

According to the recommendation of [the National Cancer Institute](#_bookmark0) ([NCI](#_bookmark0)) (accessed March 2022, at [https:](https://www.cancer.gov//research/areas/diagnosis/artificial-intelligence)

[//www.cancer.gov/\research/areas/diagnosis/artificial-intelligence](https://www.cancer.gov//research/areas/diagnosis/artificial-intelligence)),

* Artificial Intelligence (AI) is helpful to create opportunities in cancer research
* Image-based phenotype of cancer could be correlated with their underlying genomics data by deep learning methods

Deep learning (advanced machine learning algorithm in AI) can be used to identify specific gene mutations and mRNA expression from hyperplasia/tumor pathology images instead of using traditional genomic se- quencing. For instance, NCI-funded researchers ([Coudray et al. 2018](#_bookmark17)) at New York University used deep learning to analyze pathology images of lung tumors obtained from [the Cancer Imaging Archive](#_bookmark0) ([TCIA](#_bookmark0)). The [TCIA](#_bookmark0) archives pathology images of tumor specimens available as a quality control measure for re- searchers studying the genetic sequence data collected in the [TCGA](#_bookmark0) project. Coudray and his colleagues found that the deep learning method could accurately distinguish between lung adenocarcinoma and lung squamous cell carcinoma from the images ([Coudray et al. 2018](#_bookmark17)). Since a deep learning algorithm could potentially extract molecular phenotype from pathology images, mRNA expression could also be predicted from the images. A HE2RNA model, developed by Schmauch and his colleagues ([Schmauch et al. 2020](#_bookmark36)), could predict RNA-Seq profiles from whole-slide images of [TCIA](#_bookmark0). It has been validated by CD3- and CD20-[IHC](#_bookmark0) staining. The transcriptomic representation learned by HE2RNA can also be transferred on other datasets (i.e., transfer learning).

## Conclusion

Predictive markers are those that can suggest the risk of cancer without intervention. We will identify them by comparing pre-cancerous samples with follow-up results revealing malignant transformation.

High-throughput RNA-sequencing (RNA-Seq) is a powerful tool to characterize and quantify transcrip- tomes. Even the cost of RNA-Seq is continuously dropping, the budget of whole-genome transcriptomics study for survival analysis is still very high (USD350 to 500 per sample). By using the transcriptomic representation of RNA-Seq learned by NCI’s HE2RNA project ([Schmauch et al. 2020](#_bookmark36)), ST-Net (He et al., 2020) and HistoGene (Pang et al., 2021), we could find and validate predictive biomarkers from pathology images by deep learning platform under R/Pytorch/Keras framework.

**1 2 3 4 5 6 7 8 9 10 1112 1 2 3 4 5 6 7 8 9 10 1112**

10% complete

**2022**

**2023**

IRB application

## Aim 1

Retrospective cohorts (A) (B)

Prospective cohort (C)

## Aim 2

Implementing genome-wide association testing (GWAS)

## Aim 3

Construct a deep learning plat­form for pathology images

## Aim 4

Integration data & develop­ment of risk prediction models

**Figure 2:** 執行進度 (OVH and leukoplakia cohorts from (A)Chi-Mei Hospital; (B) and (C) Taipei Medical University Hospital)

**(**二**)**研究方法、進行步驟及執行進度

Specific Aims

* Aim 1: To construct [OVH](#_bookmark0) and leukoplakia datasets from Chi-Mei Hospital and [Taipei Medical Uni-](#_bookmark0) [versity Hospital](#_bookmark0) ([TMUH](#_bookmark0))
* Aim 2: To identify genetic variants associated with malignant transformation of [OVH](#_bookmark0) by implement- ing genome-wide association testing (GWAS)
* Aim 3: To construct a deep learning platform of genomics (mutation and expression) from pathology images
* Aim 4: To develop risk prediction model for malignant transformation of [OVH](#_bookmark0) by integrating risk factors identified from all sub-projects: sp1, sp2, and sp3

計畫主體內容

**Aim 1**: To construct [OVH](#_bookmark0) and leukoplakia datasets from Chi-Mei Hospital and [Taipei Medical University Hospital](#_bookmark0) ([TMUH](#_bookmark0)), with collection of clinical data and bio-specimen (i.g., pathological specimen, saliva, and serum).

**Aim 2**: To identify genetic variants associated with malignant transformation of [OVH](#_bookmark0) by implementing genome-wide association testing (GWAS).

**Aim 3**: To construct a deep learning platform of genomics (mutation and expression) from pathology images. From [OVH](#_bookmark0) and [OSCC](#_bookmark0), a deep-learning- based algorithm could extract clinical features and interpret pathology whole- slide images.

**Aim 4**: To develop risk prediction model for malignant transformation of [OVH](#_bookmark0) by integrating risk factors identified from all sub-projects: sp1, sp2, and sp3

## Aim 1:

Construct [OVH](#_bookmark0) and leukoplakia datasets

**Cohort collection** Construct [OVH](#_bookmark0) and leukoplakia datasets

**Study design for cohort (A) and (B)** First retrospective cohort will applied from biobank of Chi-Mei Hospital (CMH) as cohort (A). Data should include clinical features and bio-specimen (i.g., pathological specimen, saliva, and serum). The [OVH](#_bookmark0) and leukoplakia samples will be obtained from 150 consecutive participants at Chi-Mei Hospital. Written informed consent was obtained from all participants and was also reviewed by the Institutional Review Board Committee of CMH. Tissues, saliva and, serum were preserved in liquid nitrogen for DNA extraction. Formalin-fixed paraffin embedded tissues (FFPE) will be used for pathological imaging.

Second retrospective cohort (B) will retrieve the clinical data and FFPE from [TMUH](#_bookmark0). Data analysis will be waived review and consent by the TMU Joint Institutional Review Board (IRB), as all data will being analyzed after de-identification. The cohort (B) contains total 200 patients including xx men and xx women with their mean age at 53 year-old, range 30-88 years. The xx patients were diagnosed in non- or mild, and xx in moderate/sever oral epithelial dysplasia. Pathological imaging will be performed on FFPE samples.

**Study design for cohort (C)** According the STROKE statement ([von Elm et al. 2008](#_bookmark44)) for observational studies, one prospective cohort (C) study will be conducted.

Between 2023 and 2025, 200 patients will be chosen from cases of oral leukoplakia, verrucous hyperplasia, or lichen planus at [TMUH](#_bookmark0). Biopsy specimens are fixed in neutral buffered formalin, paraffin embedded,

and H&E stained. At least two senior pathologists should use criteria from the 2017 WHO Classification System to classify oral epithelial dysplasia of baseline lesions as mild, moderate, or severe [Gale et al.](#_bookmark20) ([2005](#_bookmark20)); [Ranganathan & Kavitha](#_bookmark35) ([2019](#_bookmark35)). The sample size was calculated according to a suggestion from the prospective study [Cao et al.](#_bookmark9) ([2009](#_bookmark9)). All of the participants have initial lesions that have never been treated with surgery, LASER, radiation therapy, or chemotherapy. Clinical features, lesion site, cigarette smoking, alcohol consumption, betel-nut chewing, and past medical history are all gathered. For at least three years, follow-up examinations will be conducted every three months in a double-blind fashion. If the clinician suspects a malignant event, a further examination and re-biopsy will be performed. Patients who do not require a re-biopsy during the follow-up examination period because of baseline lesion has clearly disappeared/regressed. TMU Joint Institutional Review Boards will approve the trial (TMU-IRB 111xxx), and all patients will provide written informed consent. This study will be also registered in the U.S. National Institutes of Health Clinical Trials Protocol Registration System in accordance with the criteria outlined by the International Committee of Medical Journal Editors (trial number NCTxxxxxxxx, available at [http://ClinicalTrials.gov](http://ClinicalTrials.gov/)).

Data should include clinical features and bio-specimen (i.g., pathological specimen in fresh state, saliva, and serum). All experimental protocols will be applied for approval under the IRB protocol with TMU Joint Institutional Review Board, and all experiments will be carried out in accordance with approved guidelines.

**Tissue sample preparation** The fresh frozen tissue samples (CMH cohort (A) and TMUH cohort (C)) are maintained with the desired orientation on the cutting block using Optimum Cutting Temperature (OCT) polymer. A thin frozen section is cut at -15 *◦*C by using a cryostat (Thermo Electron) and deposited on cold glass slides. The slide and tissue section are then quickly warmed, which let the tissue thaw-mounted onto the slide. Tissue sections are fixed by immersion in 70% then 95% ethanol bath each for 30 sec and allowed to dry for 40 min in a desiccator.

Formalin-fixed paraffin embedded tissues (FFPE), obtained from cohort (A)(B)(C), are organized into block for H&E staining as usual manner.

**Immunohistochemistry (IHC) staining** The blocks of FFPE are sectioned at 3-*µ*m slice thickness. Slides are heated for 1 hour at 60 *◦*C to deparaffinize in xylene, rehydrated in graded alcohol, and rinsed in distilled water. Antigen retrieval and immunohistochemistry staining are performed by using a Benchmark XT automated immunohistochemistry system (Ventana Medical Systems). Slides are stained with primary antibodies at 1:200 or 1:400 dilution. Primary antibodies are detected by using an indirect biotin strepta- vidin system (iVIEW DAB Detection Kit, Ventana Medical Systems, Inc.) and tissue is counterstained in haematoxylin and eosin. Images of stained slides are retrieved by digital scanner (ScanScope CS2 Systems Aperio, CA, USA) and diagnosed and graded manually by one pathologist, Dr. Ling-Cheng Mong.

We assume those biomarkers, CK13, CK17, p53, Ki-67, EGFR, *α*-SMA, CAMK2N1, FAT1, p16INK4a, and p14ARF, should be used in [IHC](#_bookmark0) staining.

**Whole-slide imaging** Construct a deep learning platform of Taiwan datasets

A quantitative tissue pathology (QTP) approach based on digital whole-slide images has enabled the profil- ing of a wide range of microscopic features at tissue, single cell and subcellular levels that are not visible to the naked eye. These phenotypes could be a reflection of its genetic carcinogenesis [Crawford et al.](#_bookmark18) ([2022](#_bookmark18)).

There is 3300 slides from 300 patients with [OVH](#_bookmark0) or leukoplakia in datasets (A)(B)(C). The 300 H&E stained pathology slides will be digitized via a whole-slide scanner at 40x magnification for this study, with

0.25 *µ*m/pixel resolution (Aperio Scanscope CS2). The other 3000 slides with stained (300 per antibody) will be scanned as the same manner. The SVS file will be obtained.

NCI’s [Genomic Data Commons](#_bookmark0) ([GDC](#_bookmark0)) stores SVS image of TCGA’s HNSCC pathology slide (we trans- ferred and saved them at

[https://github.com/texchi2/HE2RNA\_code/blob/master/gdc\_](https://github.com/texchi2/HE2RNA_code/blob/master/gdc_manifests/gdc_manifest.2018-06-26_diagnostic_TCGA-HNSC.txt) [manifests/gdc\_manifest.2018-06-26\_diagnostic\_TCGA-HNSC.txt](https://github.com/texchi2/HE2RNA_code/blob/master/gdc_manifests/gdc_manifest.2018-06-26_diagnostic_TCGA-HNSC.txt)).

The whole-slide image of pathology slide has high dimensionality (up to 100,000 x 100,000 pixels for a single image). It could be analyzed by deep learning algorithm.

## Aim 2:

## Aim 3:

Construct a deep learning platform to predict genotype alternations

**Transfer learning in Artificial Intelligent (AI)** A NCI’s HE2RNA workflow could divide the whole-slide images into squares of 112 x 112 *µ*m (224 x 224 pixels) called ”tiles”, and use the Otsu algo- rithm ([Otsu 1979](#_bookmark31)) (as implemented in python package scikit-image, available at [https://scikit-](https://scikit-image.org/docs/dev/api/skimage.html)image. [org/docs/dev/api/skimage.html](https://scikit-image.org/docs/dev/api/skimage.html)) to select only those containing tissue, excluding the white background. A maximum of 8000 such tiles from each slide are sampled then extracted 2048 features with a 50-layer ResNet model ([He et al. 2016](#_bookmark21)) pretrained on the ImageNet dataset (available at [https:](https://www.image-net.org/challenges/LSVRC/index.php)

[//www.image-net.org/challenges/LSVRC/index.php](https://www.image-net.org/challenges/LSVRC/index.php)). The ResNet and HE2RNA model are developed by using the TensorFlow/Keras implementation with a architecture of [Convolutional Neural](#_bookmark0)

[Network](#_bookmark0) ([CNN](#_bookmark0)). This AI-driven computer program analyzes tissue by creating a map of thousands of tiles. A pathology slide could be represented as a 8000 x 2048 matrix. THe k-means algorithm (as implemented in python package libkmcuda) could be used to create 100 clusters (supertiles) of tiles based on their loca- tion on the slide, and HE2RNA averages the features of the tiles within each cluster. Use of supertiles could reduce the dimension of a slide to 100 x 2048 pixels. Transcriptome prediction model will be first trained on this reduced dataset, with all the TCGA data. Then, fine-tuning of ”hyperparameters” will be achieved with full-scale data from [HNSCC](#_bookmark0) dataset by Bayesian optimization technique (i.g., Sequential Model-Based Op- timization algorithms, SMBO) ([Hutter et al. 2011](#_bookmark23)). The similar procedure of image process will be applied on datasets (A)(B)(C).

TCGA/TCIA modeling (with RNA-Seq data):

*Ymrna* = *β*0 + *β*1*X*1 + *β*2*X*2 + *β*3*X*3 + *...* + *βnXn* + *ϵ*

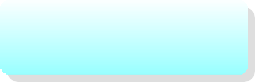
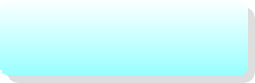
While *Ymrna* is known RNA-Seq data, and *X*1*, X*2*, ...Xn* are the features extracted by HE2RNA model from TCIA whole-slide images of pathology.

We transfer HE2RNA model on dataset (A)(B)(C) to predict their RNA-Seq representatives:

*Y*ˆ*mrna* = *β*0 + *β*1*X*1 + *β*2*X*2 + *β*3*X*3 + *...* + *βnXn* + *ϵ*

While *Y*ˆ*mrna* is predicted RNA-Seq data, and *X*1*, X*2*, ...Xn* are the features extracted from cohorts (A)(B)(C) whole-slide images of pathology.

The deep learning model of HE2RNA, which was trained under TCGA/TCIA [HNSCC](#_bookmark0) dataset, could be applied for RNA-Seq predictive of TMUH/CMH datasets. The representatives will be used for subsequent gene expression analysis (Figure [3](#_bookmark3)). A deep learning model, HE2Displasia, will also be trained for dig- ital grading system for oral epithelial dysplasia. A deep learning model, HE2DNA, will also be trained for a representative of DNA mutation. Those DNA/RNA representatives will be used for subsequent gene mutation/expression analysis.



Model (transfer from TCIA)

TMUH/CMH

datasets (WSI)

HE2RNA

RNA-Seq

(Representatives)

TMUH/CMH

datasets (WSI)

IHC

Validation

**Figure 3: Workflow of RNA-Seq representatives of TMUH/CMH datasets.** The deep learning model of HE2RNA, which was trained under TCGA/TCIA [HNSCC](#_bookmark0) dataset, could be applied for RNA-Seq predictive of TMUH/CMH datasets. The repre- sentatives will be used for subsequent gene mutation/expression analysis.

( TCGA: [the Cancer Genome Atlas](#_bookmark0); TCIA: [the Cancer Imaging Archive](#_bookmark0); WSI: whole-slide images of pathology; RNA-Seq: [RNA sequencing](#_bookmark0); [RNN](#_bookmark0): [Recurrent Neural Network](#_bookmark0) )

**Mutation/expression profile of FAT1** We will develop a deep-learning-based model (HE2DNA) as well as a validation platform. It might incorporate whole-slide images of pathology of TMUH/CMH cohorts with the mutation status of FAT1, TP53. It will also be validated in progressive and non-progressive [OVH](#_bookmark0) samples using DNA-Seq on targeted gene.

The HE2RNA model will also be used to compare the transcriptome of FAT1 in wild-type versus mutant [OVH](#_bookmark0) tissues.

## Aim 4:

**Feature engineering by explainable deep learning** HE2Displasia and [Graph Convolutional Neural Net-](#_bookmark0) [work](#_bookmark0) ([GCNN](#_bookmark0)) models will be conducted for survival analysis to predict malignant transformation of [OVH](#_bookmark0) and even leukoplakia. A survival time is defined as a progressive [OVH](#_bookmark0) becoming cancer during the follow- up period.

Forty-two features will be collected and analyzed as potential independent risk factors for malignant trans- formation. The features will be selected using the ’least absolute shrinkage and selection operator’ (LASSO) method for dimensional reduction.

The predictive model with coefficients will be established from cohorts (A)(B). The important input *X*1*...Xn* should be patients’ features: age, candidate gene mutations/expressions, and surgical margin so far. When we consider the survival prediction model, ground truth Y (i.e., malignant transformation):

*Y* = *β*0 + *β*1*X*1 + *β*2*X*2 + *β*3*X*3 + *...* + *βnXn* + *ϵ*

Actually, FAT1 gene expression is not an independent *Xmrna*. It could be influenced with other factors (for example):

*XCK*13*, XCK*17*, Xp*53*, XKi−*67*, Xp*53*, XCAMK*2*N* 1*, ...*

That is because there is much interaction in gene regulatory network.

Thus, deep learning can derive patient representations and gene network that offer improved clinical pre- dictions ([Miotto et al. 2016](#_bookmark29)). Once clinical features and models for pathology images are available, a graph convolutional neural network (GCNN) is mandatory for survival analysis ([Ching et al. 2018](#_bookmark15)) in our study.

Whole-slide images are the surrogate of expression data of RNA-Seq. From the result of Aim 3, the tran- scriptomic representatives of TMUH/CMH datasets could be applied to validate those candidates discovered by our previous study: FAT1, CAMK2N1. Moreover, the result of Aim 4 (features extracted from pathology images) will be used for survival analysis by [GCNN](#_bookmark0) (Figure [4](#_bookmark4)).

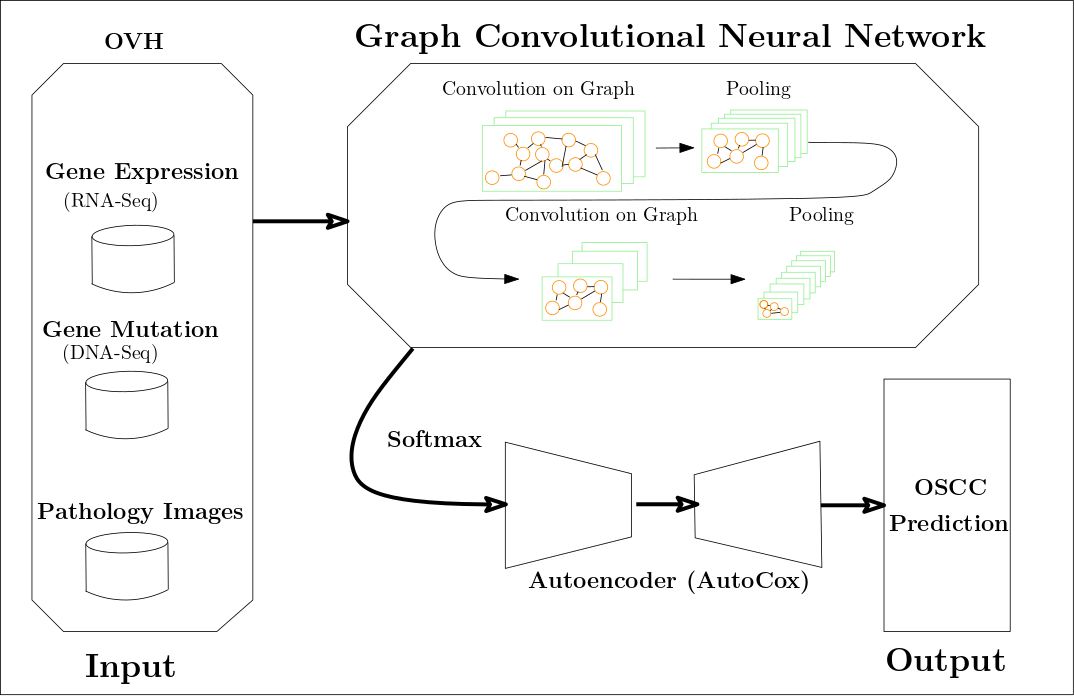
HE2Displasia model could be applied on an autoencoder for explainable machine learning. An autoencoder is a type of unsupervised learning for unlabeled data. Autoencoder with cox regression (AutoCox) (**?**) com- bines an explainable deep learning and Cox Proportional Hazards Modeling. Code is available at GitHub:

[https://github.com/Mostafa-Samy-Atlam/Autoencoder-with-survival-analysis-meth](https://github.com/Mostafa-Samy-Atlam/Autoencoder-with-survival-analysis-methods)

accessed on May 2022. As a result, AutoCox will inform us on how to estimate the malignant transforma- tion of an [oral verrucous hyperplasia](#_bookmark0) based on ”visual subjectively features” learned from pathology images.

**Modeling by Dr Wang** The candidate biomarkers, dysplasia grading, surgical margin status and saliva extract will be suggested as risk factors in malignant transformation of [OVH](#_bookmark0) through analysis of Chi-Mei and TMUH datasets.

* Statistical tools for genetic association testing
* Cox regression analyses
* Kaplan–Meier analyses
* Statistical tools for risk prediction models
* Cox regression analyses with backward stepwise elimination approach
* Nomogram prediction model
* The receiver operating characteristic (ROC) curve analysis



**Figure 4: Overview of the proposed workflow for explainable survival analysis**

With whole-slide-images pathology, the [Graph Convolutional Neural Network](#_bookmark0) ([GCNN](#_bookmark0)) should extract high-dimensional infor- mation for [OVH](#_bookmark0) survival prediction. Predictors should be ”seen” by AutoCox algorithm.

**(**三**)**預期完成之工作項目、成果及績效

* Validation of candidate biomarkers of [OVH](#_bookmark0). It will be submitted for further diagnostic method for personalized medicine of [OVH](#_bookmark0).
* We will improve our programming skill of R, PyTorch(Python) and C ++.
* We will finish manuscripts for journal publication.

**Acknowledgments**: We thank the staff of the Clinical Data Center, Office of Data Science, Taipei Medical University, Taiwan, for statistical consultation, analysis/interpretation of data, and technical support. The results shown here are in whole based upon data generated by the [TCGA](#_bookmark0) Research Network: [https://www](http://www.cancer.gov/tcga).cancer[.gov/tcga.](http://www.cancer.gov/tcga) The tables were generated by latex code with the help from [https://www.tablesgenerator.com/late](http://www.tablesgenerator.com/latex)x tables#. We sincerely want to express our thanks to the [OVH](#_bookmark0) patients who donated their data, and help from TMU Joint Institutional review board.

## REFERENCES

Alsner, J., Sørensen, S. B., & Overgaard, J. 2001, Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology, 59, 179

Angadi, P. V., Kale, A. D., & Hallikerimath, S. 2011, Journal of Oral Pathology & Medicine, 40, 208 Bentzen, S. M., Atasoy, B. M., Daley, F. M., et al. 2005, Journal of Clinical Oncology, 23, 5560

Berens, A. E., Jensen, S. K., & Nelson, C. A. 2017, Biological embedding of childhood adversity: From physiological mechanisms to clinical implications, doi:10.1186/s12916-017-0895-4

Cao, J., Zhou, J., Gao, Y., et al. 2009, Clinical Cancer Research, 15, 5178

Chang, Y.-C., Tsai, C.-H., Lai, Y.-L., et al. 2014, Journal of cellular and molecular medicine, 18, 698 Chi, L.-H. 2022, Ph.d. dissertation, Taipei Medical University, Taipei

Chi, L.-H., Wu, A. T., Hsiao, M., & Li, Y.-C. C. J. 2021, Journal of Personalized Medicine, 11, 782 Chi, L.-H., Chang, W.-M., Chang, Y.-C., et al. 2017, Scientific Reports, 7, 9031

Chien, H.-T., Young, C.-K., Chen, T.-P., et al. 2019, Cancer Prevention Research, 12, 247 LP

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., et al. 2018, Journal of The Royal Society Interface, 15, doi:10.1098/rsif.2017.0387

Chu, Y., You, M., Zhang, J., et al. 2019, Stem cells international, 2019, 9037197 Coudray, N., Ocampo, P. S., Sakellaropoulos, T., et al. 2018, Nature medicine, 24, 1559

Crawford, M., Johnson, E. H., Liu, K. Y. P., Poh, C., & Tsai, R. Y. L. 2022, Cells, 11, doi:10.3390/cells11061033

Ferrarotto, R., Kies, M. S., Garden, A. S., & Kupferman, M. E. 2017, Neoplasms of the Head and Neck, doi:doi:10.1002/9781119000822.hfcm084

Gale, N., Westra, W., & Pilch, B. Z. 2005, in (IARC), 177–179

He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778

Huang, Z., Johnson, T. S., Han, Z., et al. 2020, BMC Medical Genomics, 13, 41

Hutter, F., Hoos, H., & Leyton-Brown, K. 2011, in Learning and Intelligent Optimization, ed. C. C.A.C. (Berlin, Heidelberg: Springer-Verlag), 507–523

Joshi, R. S., Kanugula, S. S., Sudhir, S., et al. 2021, Cancers, 13, 1

Ko, Y.-C., Huang, Y., Lee, C.-H., et al. 1995, Journal of Oral Pathology and Medicine, 24, 450 Makowiecka, A., Malek, N., Mazurkiewicz, E., et al. 2019, Frontiers in cell and developmental biology, 7,

304

Marsh, D., Suchak, K., Moutasim, K. A., et al. 2011, The Journal of pathology, 223, 470 Mikami, T., Cheng, J., Maruyama, S., et al. 2011, Oral oncology, 47, 497

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. 2016, Scientific Reports, 6, 26094

O-charoenrat, P., Modjtahedi, H., Rhys-Evans, P., et al. 2000, Cancer Research, 60, 1121 Otsu, N. 1979, IEEE Transactions on Systems, Man, and Cybernetics, 9, 62

Padmanabhan, K., Grobe, H., Cohen, J., et al. 2020, Development (Cambridge, England), 147, doi:10.1242/dev.193425

Patel, A. K., Vipparthi, K., Thatikonda, V., et al. 2018, Oncogenesis, 7, doi:10.1038/s41389-018-0087-x Perisanidis, C., Perisanidis, B., Wrba, F., et al. 2012, Journal of oral pathology & medicine : official

publication of the International Association of Oral Pathologists and the American Academy of Oral

Pathology, 41, 40

Ranganathan, K., & Kavitha, L. 2019, Journal of oral and maxillofacial pathology : JOMFP, 23, 19 Rosˇka, J., Wachsmannova´, L., Hurbanova´, L., et al. 2020, Oncotarget, 11, 4735

Schmauch, B., Romagnoni, A., Pronier, E., et al. 2020, Nature Communications, 11, 3877 Silva, S. D., Agostini, M., Nishimoto, I. N., et al. 2004, Oral oncology, 40, 688

Singh, R. D., Patel, K. R., & Patel, P. S. 2016, Mutation research, 783, 15 Stefanidakis, M., & Koivunen, E. 2006, Blood, 108, 1441

Szentku´ti, G., Da´nos, K., Brauswetter, D., et al. 2015, Pathology oncology research : POR, 21, 643 von Elm, E., Altman, D. G., Egger, M., et al. 2008, Journal of clinical epidemiology, 61, 344 Wang, X., & Sun, Q. 2017, Oncotarget, 8, 624

Ward, B. B., & Helman, J. I. 2018, Squamous Cell Carcinoma of the Oral and Maxillofacial Region, third edit edn. No. chapter 28 (Elsevier), 670–689, doi:10.1016/B978-0-323-41499-9.00071-6

Znaori, A., Brennan, P., Gajalakshmi, V., et al. 2003, International Journal of Cancer, 105, 681

轉譯創新研究計畫之子計畫書

子計畫編號：( 3 )

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 子 計畫名稱 | Deep learning platform and risk prediction models for malignant transformation of oral verrucous hyperplasia  以深度學習及風險預測模式探討口腔疣狀增生惡轉風險 | | | | |
| 子計畫主持人 | 吳家佑 | 職稱 | 醫師/助理教授 | 學院/科系所 | 北醫附醫口腔顎面外科 |
| 連絡電話 | (O) 02-  (M) | | | E-mail | borgiawu@gmail.com |

1. **近三年內執行之研究計畫**

（請填寫近三年研究計畫）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 計畫名稱 | 計畫內擔  任之工作 | 起迄年月 | 補助或委託機構 | 執行情形 | 經費總額 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**國際合作對象資料表**

一、基本資料：

1. 國際合作對象指與外國研究者進行合作研究，且與外國合作研究者有共同發表成果或申請專利潛力。

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 |  | 職稱 |  |
| 任職機構 |  | 單位 |  |
| 聯絡方式 | 電話：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_傳真（Fax）：\_\_\_\_\_\_\_\_\_\_\_\_\_  E-mail：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | |
| 合作國家 | □與單一國家合作，國家名稱：  □與多國合作，主要國家名稱：  其他參與國家：1. 2. 3. 4. | | |

備註：如有經費需求，請依據校內相關規定辦理。

**個人資料表使用說明**

1. 資料項目：

(一)、基本資料：包括「身分證號碼」、「中、英文姓名」等資料。外籍人士若無身分證號碼可填寫居留證號，若無居留證號時，請將西元出生年月日八碼再加英文姓氏(Last Name)前二碼，組成十碼後填入身分證號碼欄位，例如「YYYYMMDD」。

(二)、主要學歷：以獲有學位之學歷為主，或個人之最高學歷。

(三)、現職及經歷：限與研究相關之編制內專任職務。

(四)、專長：請自行填寫與研究方向有關之專長學科。

(五)、著作目錄：指個人申請截止日前五年內(此段期間曾生產或請育嬰假者，得延長至七年內，曾服國民義務役者，得依實際服役時間予以延長，但應檢附相關證明文件)發表之學術性論文或著作，包括：期刊論文、專書及專書論文、研討會論文、技術報告及其他等。

(六)、研發成果智慧財產權及其應用績效：指個人研發成果所產生之智慧財產權及其應用績效，分為：(1)專利 (2)技術移轉 (3)著作授權 (4)其他等類別。

個人資料表

以下各項資料均收錄於科技部研究人才資料庫。

一、基本資料：　　　　　 簽名：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 身分證號碼 |  |  |  |  |  |  |  |  |  |  | 填表日期： | | | | | 20 / / |
| 中文姓名 |  | | | | | | 英文姓名 | | | |  | | | | | |
| (Last Name) (First Name) (Middle Name) | | | | | |
| 國籍 |  | | | | | | 性 別 | | | | □男 □女 | | 出生日期 | | 19\_\_\_\_年\_\_\_\_月\_\_\_\_日 | |
| 聯絡地址 | □□□□□ | | | | | | | | | | | | | | | |
| 聯絡電話 | (公) (宅 /手機) | | | | | | | | | | | | | | | |
| 傳真號碼 |  | | | | | | | | | | | E-mail | |  | | |

二、主要學歷 由最高學歷依次填寫，若仍在學者，請在學位欄填「肄業」。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 學校名稱 | 國別 | 主修學門系所 | 學位 | 起訖年月(西元年/月) |
|  |  |  |  | 自 / 至 / |
|  |  |  |  | 自 / 至 / |
|  |  |  |  | 自 / 至 / |
|  |  |  |  | 自 / 至 / |

三、現職及與專長相關之經歷 指與研究相關之專任職務，請依任職之時間先後順序由最近者往前追溯。

|  |  |  |  |
| --- | --- | --- | --- |
| 服務機構 | 服務部門／系所 | 職稱 | 起訖年月(西元年/月) |
| 現職： |  |  | 自 / 至 / |
| 經歷： |  |  | 自 / 至 / |
|  |  |  | 自 / 至 / |
|  |  |  | 自 / 至 / |
|  |  |  | 自 / 至 / |
|  |  |  | 自 / 至 / |

四、專長 請填寫與研究方向有關之學術專長名稱。

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | 4 |

表C301 共 頁 第 頁

五、著作目錄：

* + 1. 請詳列個人申請截止日前五年內(此段期間曾生產或請育嬰假者，得延長至七年內，曾服國民義務役者，得依實際服役時間予以延長，但應檢附相關證明文件)發表之學術性著作，包括：期刊論文、專書及專書論文、研討會論文、技術報告及其他等，並請依各類著作之重要性自行排列先後順序。
    2. 各類著作請按發表時間先後順序填寫。各項著作請務必依作者姓名(按原出版之次序**，通訊作者請加註\***。)、出版年、月份、題目、期刊名稱（專書出版社）、起迄頁數之順序填寫，被接受刊登尚未正式出版者請附被接受函。
    3. 若期刊是屬國內或國際期刊資料庫(如SCI、EI、SSCI、A&HCI、Scopus、TSSCI、THCI Core…等)所收錄者，請於該著作書目後註明資料庫名稱；若著作係經由國科會補助之研究計畫所產生，請於最後填入相關之國科會計畫編號。

表C302 共 頁 第 頁

六、研發成果智慧財產權及其應用績效：

1. 請將個人研發成果所產生之智慧財產權及其應用績效分為1.專利2.技術移轉3.著作授權  
   4.其他等類別，分別填入下列表中。如欄位不足，請自行加印填寫。
2. 填寫順序請依專利期間起始日排列，或技術移轉及著作授權之簽約日期排列。

1.專利：

請填入目前仍有效之專利。「類別」請填入代碼：(A)發明專利(B)新型專利(C)新式樣專利。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 類別 | 專利名稱 | 國別 | 專利號碼 | 發明人 | 專利權人 | 專利核准日 期 | 國科會計畫編號 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

2.技術移轉：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 技術名稱 | 專利名稱 | 授權單位 | 被授權單位 | 簽約日期 | 國科會計畫編號 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| 產生績效：(可另紙繕寫) | | | | | |

3.著作授權「類別」分(1)語文著作(2)電腦程式著作(3)視聽著作(4)錄音著作(5)其他，請擇一代碼填入。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 著作名稱 | 類別 | 著作人 | 著作財產權人 | 被授權人 | 國科會計畫編號 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| 產生績效：(可另紙繕寫) | | | | | |

4.其他協助產業技術發展之具體績效

|  |
| --- |
|  |
|  |
|  |

表C303 共 頁 第 頁